INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 4, PP. 1-10
Manuscript received April 20, 2025; revised September 2025. doi: 10.24425/ijet.2025.155473

An improved human pose estimation using Deep Neural Network for the optimization of human-robot interactions

Ravi Raj, and Andrzej Kos

***Abstract*—Research shows that mobile support robots are becoming increasingly valuable in various situations, such as monitoring daily activities, providing medical services, and supporting elderly people. For interpreting human conduct and intention, these robots largely depend on human activity recognition (HAR). However, previous awareness of human appearance (human recognition) and recognition of humans for monitoring (human surveillance) are necessary to enable HAR to work with assistance robots. Al-so However, multimodal human behavior recognition is constrained by costly hardware and a rigorous setting, making it challenging to effectively balance inference accuracy and system expense. Naturally, a key problem in human pose or behavior detection is the ability to extract additional purposeful interpretations from easily accessible live videos. In this paper, we employ human pose detection to address the problem and provide well-crafted assessment measures to show demonstrate the effectiveness of our approach, which utilizes deep neural networks (DNNs) This article proposes a human intention detection system that anticipates human intentions in human- and robot-centered scenarios by utilizing the incorporation of visual information as well as input features, including human positions, head orientations, and critical skeletal key points. Our goal is to aid human-robot interactions by helping mobile robots through real-time human pose prediction using the recognition of 18 distinct key points in the body's structure. The effectiveness of this strategy is demonstrated by the suggested study using Python, and the results of simulations verify the reliability and accuracy of this method.**

***Keywords*—Deep Learning (DL); Deep Neural Network (DNN); Human-Robot Interaction (HRI); Human Pose Estimation; Key Points; Machine Learning (ML)**

I. INTRODUCTION

TRENDS in society are making more and more surveillance systems necessary. Particularly, applications, including the identification of questionable human behavior and efficient merchandise display arrangements in stores, are anticipated to result from the incorporation of cognitive processing capabilities to the surveillance of photographs. Methods for identifying people in photos, their faces, and their body postures are needed for these activities. The term "human pose estimation methods" refers to all of these. This research investigates an information-based human pose estimation method for individual-specific human-robot interaction (HRI). The multidisciplinary area of HRI studies the creation, use, and analysis of robots that communicate with people various of settings [1]. Fundamentally, HRI wants to build robots that can interact, cooperate, and live alongside people in shared spaces. These encounters might range from basic task-oriented exchanges to intricate social and emotional exchanges. As humans become more accustomed to working with mobile robots, their incorporation into daily life is becoming more widespread [2]. Providing a mobile robot with the ability to do standard manipulating tasks is an essential precondition to achieving this [3], [4].

There is a tremendous opportunity for autonomous assistive robots to benefit caregivers and relatives who are responsible for patients attending medical facilities. Additionally, a benefit of ambient supported living is automated caregivers who monitor the well-being of elderly or handicapped individuals in a range of settings, including homes, eldercare centers, and hospitals. Techniques for recognizing human actions are necessary to allow robotics to comprehend human conduct and respond accordingly [5]. Robots can deduce information about a person's goals, actions, and situations by employing human activity recognition (HAR) systems. This information can be utilized anywhere from understanding everyday chores to aiding in an emergency. But before aid robots are able to have such features built, certain criteria need to be achieved.

The subject of human pose estimation, described as the identification of human body joints, has received a lot of interest in the image processing field [6]. Predicting human pose is significant because it allows robots to interact with human beings by recognizing the human being's poses, actions, and behaviors [7]. Real-time human pose estimation is a crucial endeavor in computer vision, focused on swiftly determining the spatiotemporal configuration of human key points, including the head, arms, shoulders, and limbs, from videos or image frames and afterwards inferring their poses, including rotating, stretching, or bending [8].

In this work, we adopt a comprehensive approach to human pose estimation. We leverage the latest DL breakthroughs and

This work was supported financially by the AGH University of Science and Technology, Krakow, Poland, under subvention no. 16.16.230.434.

Ravi Raj is with Department of Mining, Industrial,and ICT Engineering (EMIT), Manresa School of Engineering (EPSEM), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain (e-mail: ravi.raj@upc.edu; ORCID: https://orcid.org/0000-0001-8073-1812; corresponding Author).

Andrzej Kos is with the Faculty of Computer Science, Electronics, and Telecommunications, AGH University of Krakow, Krakow, Poland (e-mail: kos@agh.edu.pl).

offer a unique DNN-based approach. DNNs have demonstrated remarkable success in the recognition of images [9] and, recently, in detecting objects [10]. In this study, we aim to shed some insight on this subject by presenting a simple but effective approach to comprehensive human pose estimation using DNN. We model the estimation of human pose via a joint regression issue and demonstrate how to implement the issue properly into DNN scenarios. Figure 1 shows the basic structure of a DNN that contains an input layer, a hidden layer, and an output layer.

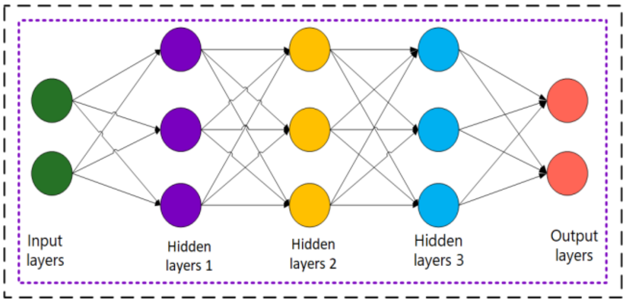

Fig. 1. Basic structure of deep neural network.

Robots can now more easily interpret and grasp human intentions and actions due to a significant technological advancement: the capacity to determine posture in video clips. This increases the possibility of human-robot interaction. Robots can recognize complex movements like walking, sitting, standing up straight, waving, and handling objects by analyzing individual frames of a film [11]. Some crucial steps need to be executed for the purpose of making it easier for people and robots to work collaboratively. These consist of gathering data, preprocessing it, extracting pertinent characteristics, using machine learning (ML) methods, categorizing actions, and establishing efficient human-robot interaction. Pre-processing methods are used to reduce noise, normalize frames, and optimize lighting settings, while data collection involves a sizable dataset made up of labeled videos [12]. Figure 2 illustrates the process of human activity or behavior recognition. Finding and extracting relevant characteristics from each frame or segment is the method of feature extraction. These features are then used as data for ML algorithms. To allow robots to modify their actions and conduct accordingly, activity segmentation is the procedure of classifying actions based on real-time footage that generates other novel information. When it comes to HRI, recognizing behaviors has several uses in public places, medical facilities, manufacturing facilities, and personal assistance.

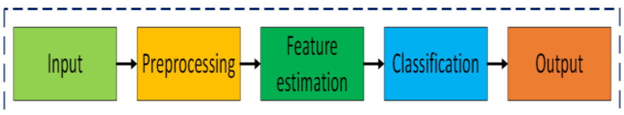

Fig. 2. Process of human activity or behavior estimation.

The focus of robotics society has consistently been on HRI. The robotics society has expanded thanks to the Industry 4.0 project, allowing for more flexible interactions between robots and their surroundings [13]. The application of robotics was greatly constrained in the past by the fact that most of them could only do basic repetitive activities inside a limited working space. The concept of HRI advances the concept that robots can exist alongside people in environments including houses, restaurants, and healthcare facilities, helping the elderly, blind, and physically disabled with a variety of work [14]. Human requirements for both work and life might be met by robots with successful HRI, releasing humans from risky, monotonous duties and enabling individuals to focus on more complex tasks. Moreover, the global pattern of aging populations has made the demand for assistance robots imperative. Despite this, robotic assistants currently under development have not yet reached the level of functionality required to operate effectively in our homes and workplaces. HRI is defined in various ways, from collaborative physical tasks [15], [16] to cognitive functions [17]. Physical cooperation for older assistance Robots concentrates on giving robotic devices the tools required to fulfill different senior citizen demands in the real world. However, cognitive features focus more on factors like intention recognition, pose estimation, HAR, interpersonal interaction, and HRI that affect how robots and older people interact [18].

HRI procedures are essential to robots, and they are rapidly becoming more multimodal and genuine in response. It is thought that visual-based interactions provide a more logical and instinctive kind of communication. An essential component of efficient HRI is human pose estimation [19]. Robotic platforms intended to provide aid and rehabilitation must possess accurate sensory systems facilitating HRI. Thus, robots are required to be capable of recognizing human postures or expressions for the purpose of enhancing the effectiveness and security of human-robot cooperation [20]. The skeleton-based technique is one of the pose recognition systems that has drawn the most attention because of its resilience against background variability and compact design [21]. This article presents an efficient technique for human pose estimation using a skeleton-based approach for optimizing HRI. Robots are better able to interpret and react to human movements, intentions, and gestures whenever they are capable of accurately recognizing human postures. Techniques that identify a human pose from images captured with a webcam have several advantages for pose estimation, which is very useful for HRI and below are some other methods of major types of human movement techniques for HRI:

A. Human Activity Recognition

Robots for assistive work are able to comprehend human behavior by regularly observing human actions. Artificial intelligence (AI) allows assistive robots to anticipate human needs and provide appropriate assistance [22]. Perhaps when someone makes a demand, an automated system might be able to determine that they are searching for something and provide it to them. HAR is essential for the creation of cognitive and adaptive automated systems that can recognize, interact with, and assist humans in a range of scenarios [23]. The standard of user communication and its uses in HRI is going to improve with HAR, or human pose estimation technologies. HAR is necessary for optimal HRI because it allows intelligent machines to understand human behavior and respond to it instantly.

B. Human Gesture Recognition

The pose assessment technique enables intelligent machines to recognize and analyze human actions. Human gesture recognition is very useful when speaking with someone who could have speech challenges or in circumstances where verbal engagement is prohibited or unwanted, including noisy environments, dumbness and deafness, and many more [24]. Precise human pose estimation techniques help humans and robotic devices to interact more intelligently and quickly. The capacity of an autonomous system to mimic or respond to human gestures enhances collaboration and engagement in the form of HRI [25]. Pose estimation is often used to create dynamic patterns that adapt based on human movement and the posture of their bodies. Human pose estimation is used to optimize the intelligent robot's capacity to better awareness of human behavior and HRI.

C. Robotics Assistance

Robotics can follow and assist humans with accurate motions or exercises in settings including hospitals or psychological rehabilitation centers due to human pose estimation. Human pose estimation also helps to ensure secure relationships among humans and robotic devices in cooperative environments, including hospitals, rehabilitation facilities, offices, companies, and many more [26]. By continuously observing human poses, automated systems can assist in mitigating incidents by sensing and reacting to possibly dangerous movements or collisions. By enabling intelligent machines to assist people more strategically and receptively across a range of fields, the estimation of human poses enhances safety, performance, and interpersonal interaction in HRI scenarios.

The Further advances in HRI systems are essential for robots, which will focus on making them natural, fast-response, and multifunctional. However, the shortcomings of the interaction approaches used now are as follows: delayed reaction of generalized pose-recognizing algorithms, particularly in the early stages; inadequate ability to extract and fuse features from spatial and temporal data; and inadequate human pose recognition framework in HRI. We suggest a fast-responding deep neural network (DNN) in this paper, which is used for human pose detection, to be able to get over such challenges. In this research, we study how to estimate the human 2D location and orientation by measuring the 2D location of a human body's joints using a wide-field-of-view RGB camera. Our suggested approach solves all the problems that previously had not been solved since they use RGB pictures.

The human pose estimation in this paper has been performed by OpenPose [27], which relies upon a DNN and can identify human poses in a picture, video, or live webcam. It enables a mobile robot to identify a human pose whenever it is integrated with the mobile robot when it is in motion and to identify its stance in relation to its surroundings. The robot can create a social connection with its users and an appropriate navigation route towards them using this approach. This study can encourage multidisciplinary cooperation, develop human pose estimation approaches, and create more secure, diverse, and natural-feeling HRIs.

The proposed approach provides an important advancement in the field of human pose estimation, specifically focusing on its application for improving human-robot interactions. The key contributions are as follows:

- The proposed model is designed to improve the accuracy and real-time performance of human pose detection, even in diverse environments. It incorporates advanced techniques in computer vision and machine learning, making it capable of more accurately detecting the positions and movements of human bodies in various poses.
- The research demonstrates how improved pose estimation can be directly applied to the optimization of human-robot interactions. By accurately understanding human gestures, movements, and postures, robots can better interpret and respond to human actions. This results in more intuitive and efficient collaboration between humans and robots in tasks such as assistance, guidance, and shared operations.
- The paper highlights the model's robustness in dealing with different lighting conditions, occlusions, and varying human postures. This ensures that the approach can be widely applied in practical, real-world scenarios where human-robot interaction is crucial, such as healthcare, manufacturing, and service industries.
- Through the use of deep learning methods, the authors emphasize that their system achieves high accuracy in pose estimation and operates efficiently, which is essential for real-time applications. This is particularly beneficial in dynamic environments where robots must react promptly to human actions.
- This paper lays the foundation for further improvements in human pose estimation techniques, which are crucial for optimizing the effectiveness and safety of human-robot interactions, particularly in complex and real-world environments.

We further explain this paper in five sections, as follows: Section 2 provides a literature survey related to this work; Section 3 explains background information about the methodology of human pose estimation; Section 4 discusses the simulation and experimental analysis, and section 5 describes conclusion and future re-search perspectives.

II. Literature Survey

Significant advancements have been achieved in predicting human body positions, especially those related to estimating how people move, by using various deep learning (DL) methods, such as graph convolutional networks (GCNs) and recurrent neural networks (RNNs). Estimating human poses is the primary objective, and it is specified in predetermined time frames that simulate a preset camera recording speed. Recently, many researchers have focused on developing an effective human pose estimation technique for efficient HRI.

Cao et al. [28] present 6IMPOSE, a unique framework for 6D pose prediction and the creation of sim-to-real data. There are four sections in 6IMPOSE: Initially, a process for generating data is used, which uses the 3D software package Blender to generate data from synthesized RGBD images with 6D pose labels. Secondly, we used the recommended process to create an augmented RGBD dataset featuring five typical household products. Thirdly, a simplified, real-time variant of the 6D posture prediction method PVN3D that is tailored for robotics uses a limited time frame and an object detector, YOLO-V4. Lastly, a codebase was created to make it easier to include the visual system in a task involving robotic grasping. This method achieves a total success rate of about 87% in capturing five distinct everyday items from congested backgrounds in varied conditions of light. It also shows how to efficiently generate many RGBD pictures and successfully transfer the trained inferred designs to robot grasping studies.

Huo et al. [29] present a graph convolutional network (GCN) and transformer-transformers that are frequently utilized in natural language processing to create a human pose estimation system for videos using a 2D lift to 3D method. More precise 3D pose coordinates might be obtained by using transformers in the proposed method for obtaining sequential features and convolution of graphs to collect information between local joints. The suggested 3D pose estimation system is utilized to create animated figure movements, follow robot motion, and develop HRI application systems. The Human3.6M dataset is used for testing the presented 3D human pose estimation system, which operates better than the most advanced systems.

Bhuiyan et al. [30] provide a knowledge-driven strategy for HRI using a visual-based pose estimation method. The system begins with a facial identification and pose identification strategy utilizing subdomain PCA-based pattern-matching algorithms. This relies on a visual representation of the face via connected element assessment of human skin color division of photos in the HSV color system. The subdomain technique performs better for facial pose segmentation than the usual PCA technique, according to experimental data. The method's application to communicate with AIBO robots in HRI has served as a demonstration of the technology.

Amorim et al. [31] suggested a combination system that combines a collection of inertial measurement units (IMUs) installed in human limbs to perform pose assessment with the human monitoring abilities of a 3D vision sensor. The IMUs maintain monitoring consistency by making up for the gaps in obscured regions. An ongoing live computation of the offset value is done in this study for the purpose of lessening the remaining impact on the IMU offset. The outcomes illustrate that this technique can accurately determine a person's location, such as their forearm, within millimeters and is resistant to occlusions.

Lombardi et al. [32] provide a learning-based system that autonomously recognizes instances of eye contact when interacting with human partners virtually. This paper implements a system for evaluating gaze orientation, paying special attention to simultaneous gaze, which is a crucial behavioral signal in interactions with one another. The suggested approach achieved excellent efficiency both in silico and in real-world situations. It is anticipated that this study will serve as a foundation for attentive architecture that can support situations where robots are viewed as social companions.

Saadatnejad et al. [33] create an open-source human pose prediction library that supports numerous datasets, incorporates various scenarios, and uses defined evaluation standards with the goal of advancing studies and the development of a single, consistent assessment system. To improve performance and develop deeper confidence, two different forms of uncertainty are explored in the problem. First, it provides a way to incorporate information regarding the unknown distribution into an algorithmic, unknown framework by using unknown assumptions. Second, it provides a unique method for assessing the complexity of a model's responsibilities and grouping them to measure its cognitive uncertainty.

Fan et al. [34] seek to address the problem of joint hand-object posture prediction in a cooperative human-robot disassembling situation. This approach has applications in practice in several other close-range HRI scenarios. This research presents a method that can determine the hand's and object's 3D posture simultaneously in a unified model. The comparison trials indicate that the suggested method works better than many current hand-object estimation methods.

Yang et al. [35] provide an in-hand tactile-based perception of objects method that is reinforced with a sim-to-real strategy for a data-efficient learning procedure. Two vision-based sensory devices captured high-fidelity sensory data, which was interpreted as a single-point cloud tinterpreting itclassification and pose estimation. The framework was initially trained using a dependable simulation approach using tactile input, and it was then refined using actual tactile data. A re-grasping technique is presented in this work using the conditional gathering of category distributions of probabilities, drawing reference from human actions. By using the suggested method, robots might be able to perceive irregular surroundings with haptic exploration abilities like those of humans.

Salimi et al. [36] present a novel approach to human fall detection that depends upon the Rapid Pose Estimation technique. Human falls are a problem that raises significant issues, particularly for elderly people. Fall events might be detected with the finest precision using machine learning and computer vision techniques. These imaging-based technologies serve as a valuable substitute for body-worn ones. The method classifies the data retrieved from frames of photos using the One-Dimensional Convolutional Neural Network (1D-CNN) and Time-Distributed Convolutional Long Short-Term Memory (TD-CNN-LSTM) models, achieving substantial accuracy rates.

III. Background Information

The continuous digitalization of science, technology, and humanity is changing the methods of all areas of research. The prospect of automated tasks has frequently inspired human interest. During automated activities, human pose recognition represents one of the most important components [37]. A branch of computer vision called "pose predictions" looks at past poses to estimate the potential position of the joints of the human body [38]. It covers not only the movement and direction of the human body but also the estimation of joint locations. Human pose estimation is crucial in numerous fields, including navigation, sports activity correction, HRI, healthcare assistance, and several more. Wellness and athlete trainers might use such tools to create more individualized plans for training and make better changes to workouts and athletics by gaining greater knowledge of the athletic actions and routines that drive their trainees [39]. With the goal of improving the security and precision of HRI, robot navigation, athletic instruction, health care, monitoring, and several other uses, this proposed research intends to build an improved AI approach to human pose estimation.

The human pose estimation method enhances the transformation process in many important ways and could help predict patterns in situations involving people. Powerful dynamic models are particularly essential when there are many people around since they allow for narrower search zones, which drastically lowers the difficulty of data processing. Our goal is to quickly and reliably identify human behaviors using the robot's integrated sensors, allowing for more seamless, safe, inherent, and anticipatory navigation. We provide human pose estimation that leverages various data sources, including the previous position of every individual and vision-based detail, such as the head's alignment or the key points of the skeleton when available. Additionally, the simulation is not dependent on the number of individuals within every frame, allowing for a fully attention-based approach. This sug-gests that the system might dynamically accept several types of human populations at different time intervals during estimation.

A. *Input data*

The robot's recognized latest T + 1 intervals can be processed as operating attributes and image data. An image frame at that exact phase or an initial cloud of points can make up visual information, which is made up of information given by every agent in proximity. Agent properties include things like each agent's median position and vision-based features like head position and skeletal key points. Using both external and internal camera evaluations, patches of pictures representing each agent's identified 2D boundary lines are first formed across the 360-degree vision to extract critical vision-based features from the raw data. To acquire skeletal key points by using such patches, one can choose from a variety of publicly accessible skeleton key point extraction tools from photographs. Conversely, data extraction often produces critical points in a 2D image reference system. We produce 2D key points leveraging the methods derived by Grishchenko *et al.* [41] where utilizing an existing trained system for identifying 2D crucial points using images. After an informative human-shaped framework is fitted to the provided 2D key points, the optimized equation (1) [42] might be solved, yielding the 2D labeling necessary for pre-training under supervision.

$$\arg min_k \left(\left\| r(k) - \hat{k}_2 \right\|_2 + \lambda H(k) \right) \qquad (1)$$

Where key points of a 2D skeleton are abbreviated as k, the function known as re-projection, which employs camera assessments to display 2D key points as 2D images, is indicated by, $r: \mathbb{R}^{33\times3} \to \mathbb{R}^{33\times2}$, and the arrangement for a human pose can be determined by $H(k)$.

B. *Framework details*

Many The present research uses OpenPose, a popular system for estimating human pose. It recognizes and locates important features of the human skeleton, including joints and other components, in pictures or videos using deep learning methods [43]. Using a multiple-phase convolutional neural network (CNN), OpenPose can concurrently identify key points on the human body, hand, and face. It operates by first employing a sequence of convolutional layers to identify the human body components, then utilizing an improved network to fine-tune the key point's placements. A transformation level is the primary architectural component inside a structure. It is composed of a multi-head consciousness level and many thick and normalization layers [44]. The transformation level receives three vectors: key point (*K*), variable (*V*), and query (*Q*). Although this, each tensor can handle many inputs; hence, we characterize the self-aware activity via a transformation level having resources *K*, *Q*, and *V* being a unique tensor: the tensor transfers information across different ways while acting on its own.

The independently transferred agent attributes are combined in a trained attention search. For an entire self-awareness function, every human timestep sign is expected to have accessibility to every additional human timestep sign, time, and human assessment. When an agent's feature is absent for a given timestep, we mask those timesteps using *0*. This provides a quick way for data to be shared. This method requires that every agent or robot at every timestep possess the ability to communicate with each other at every timestep as well as with the extra agents during that duration. One of the key findings of the study is that, based on how the agent is now implemented utilizing the same previous qualities, its prospects might be predicted substantially. Before using a dense layer to the project per modalities features, the learned modalities recognition is adjusted by using agent-timestep signs, that transformation levels, once again using total self-awareness. To naturally combine many information flows, the query pays attention to extra data from a separate tensor. Figure 3 shows an illustration of human pose estimation using deep neural network by detecting 18 distinct key points.

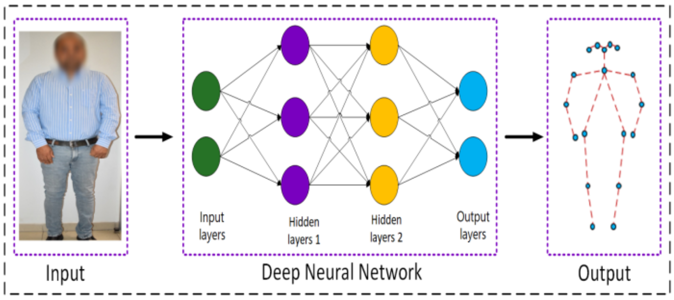

Fig. 3. Framework of human pose estimation using deep neural network.

C. *Multi-model Pattern Distributions*

Our methodology can forecast many plausible possibilities for a given situation. This is accomplished by multiplying the agent-time step indicators with the total quantity of possible modes (*M*) within the multi-model patterns propagation, which results in a pattern expression [*A, H, M, h*]. Where the quantity of agents over *T+1* current and previous time steps is denoted by *A*, *H= T+1+F*, and the token's size is *h*. *F* represents each agent's subsequent steps. To facilitate mode differentiation, it is combined with an acquired mode-identifier vector [*1, 1, A, h*].

Every potential outcome P_m: $m \in 1,2,\dots M$. The human pose estimation systems for the exemptions of the median location of the ith robot for every phase t using the 2D Gaussian hybrid system in equation (2) [42] are used when combining variables (α, β) for forecasting per mode based on the probable mode P_m obtained using the multi-modal induction.

$$Q_\theta^i\big(Y_t|0(t),..,0(t-T)\big) = \sum_{m=1}^{M} w_m\, A\big(Y, \beta_{m,i,t}, \alpha_{m,i,t}\big) \quad (2)$$

Where m is the mth future mode. The position of an agent during a particular period is expressed above using a Gaussian hybrid system with combined weights w comparable to the variance of probability for projected patterns.

IV. Simulation And Experimental Analysis

In everyday uses including self-driving cars, social distance supervision, medical assistance to elderly people, HRI, military surveillance, and sports performance tracking, human pose estimation is more vital for gesture recognition. Our research methodology is designed to augment our contributions: first, we demonstrate analytically and statistically that our approach anticipates human pose estimation with high accuracy. We demonstrate how a pose estimation might be applied to constant HRI simulation in several possible contexts. Finally, we demonstrate how vision-based features might be leveraged by human pose estimation to improve forecasting precision in human-centered environments, especially in situations where errors in prediction are large, and history is limited. Predictive systems relying completely on previous position information persist in scenarios when the framework has no or little prior information regarding human orientation. To prevent people from colliding with robots when they are exploring their surroundings, all robots need to be equipped with a pose estimation system built for situations like these [45]. The set of hyperparameters and the descriptions used in this research are shown in Table I.

Table I

List of Hyperparameters

System Configuration	Descriptions
Python version	2.8.1
NumPy version	1.21.5
TensorFlow version	2.9.1
Optimizer	Adam
Learning rate	0.001
RAM	8GB
Processor	Intel(R) core (TM) i3-4005U

Practical applications of human pose estimation are a difficult issue. We assessed how well the proposed pose estimation systems performed when it came to recognizing both static and dynamic activity patterns done by humans in realistic observation situations. The elderly living independently within the home might want a mobile robot that tracks and identifies their positions autonomously since they run the danger of slipping and hurting themselves [46]. Even though deep learning techniques are still in their infancy, they are not yet capable of accurately estimating poses which are uncommon or nonexistent in training datasets. Globally, the number of elderly people is constantly increasing because of advancements in health care and healthy eating.

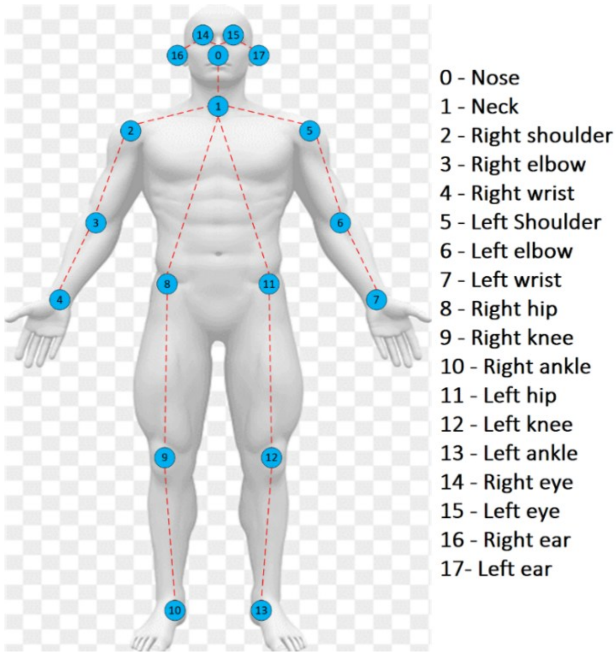

Fig. 4. Illustration of different key points for human pose estimation.

A mobile robot that walks throughout the household snaps images of senior citizens in suitable poses, and then autonomously assesses their present activities or stance to notify the right parties when a potentially hazardous scenario or issue emerges might be highly helpful [47]. This study improves the quality of HRI by implementing accurate human pose estimation. The main domains of interest for this study are detecting movement and recognizing image systems to estimate human poses. Using a camera lens, real-time photos are taken for the purpose of recognizing and distinguishing between body gestures. Our focus is on identifying human poses, where each action that is recorded signifies a directive in a human-centered environment. Figure 4 shows the human image with 18-different key points, which are used to identify human poses. In this study, the essential points of significance for a human's pose have been identified by considering 18 distinct human body parts.

We processed and recognized images and videos using the OpenCV functions in Python for the algorithm's execution. We provide pre-action evaluation using machine learning (ML) and image identification for whole-body motions, collecting user movement behaviors by utilizing the Open Pose modeling technique. These recorded gestures are then included in the creation of dynamic recognition applications, including HRI and medical assistance for physically disabled people. A multiple-threaded strategy was used to provide a smooth connection among the simulated executions in HRI, leading to the construction of two different processing frames on the interfaces. The first is devoted to displaying real-time findings of human movement detection, whereas the last one determines when computer simulations respond properly and synchronously to human behavior orders inside the HRI environment. The video capture tool using the OpenCV unit is used in the framework to record user-initiated activities while real-time footage is captured from the perspective of the lens. To accurately assess if the user's completed activities correspond with the prescribed in-HRI, like providing medical aid to a physically handicapped individual, this method involves human pose estimation. Fig. 5 presents the general flowchart for the proposed human pose estimation. Since the model-based pose estimation essentially depends on the subject's bone length details, our method verifies the input data.

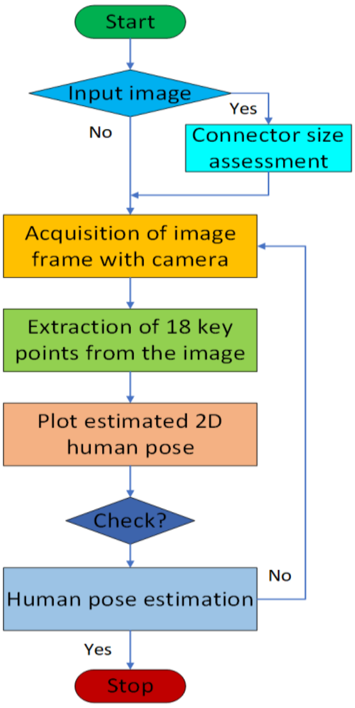

Fig. 5. The framework of the proposed human pose estimation technique.

The procedure for assessing connection size operates as follows: the subject presents with their arms extended, pictures are taken for at least ten frames, and the usual distance among the position of the bone's end connections at every picture is used to determine the size of every bone connection if the present setup does not contain connection size information for the present subject. Images are acquired from an RGB camera using the laptop webcam's image grabber module. The next step involves extracting 18 key points from the captured image. Thereafter, plot the approximate poses in two dimensions based on the context of applicability. Lastly, the estimation of poses must be terminated if an end-user requirement is satisfied or if this is the final picture frame. Anticipating human actions in dynamic environments such as homes and offices is crucial for reliable and efficient assistance robot navigation. The recommended model is trained to perform pose estimation tasks using a deep neural network (DNN). We specifically focus on demonstrating the use of probabilistic human skeletal data obtained from on-the-ground human pose estimation. We present an estimating system that integrates and evaluates, in an adaptive manner, exact vision-based behavioral characteristics, including head position and main skeletal spots.

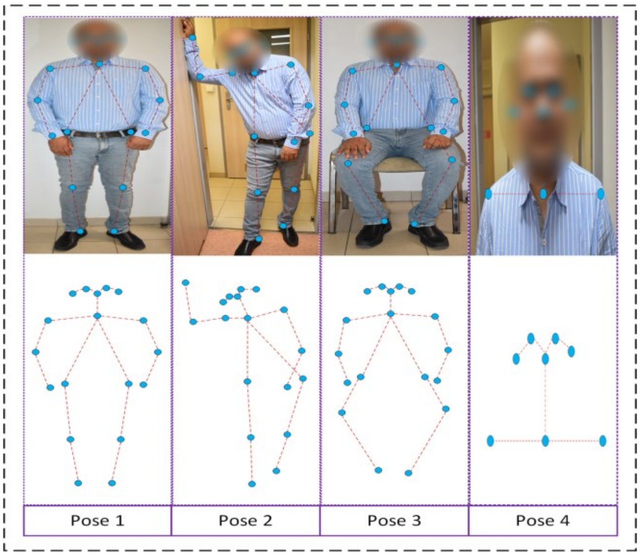

Fig. 6. Simulation output for static human poses with skeleton key points.

Anticipating human actions in dynamic environments such as homes, businesses, hospitals, and senior living facilities is crucial to ensuring safe and effective robot movement and improved HRI. These locations remain challenging since individuals tend not to abide by the regulations when navigating them, as well as because there are sometimes multiple doors and other hidden paths of access that increase the risk of unintentional collisions. The main reason is that visual footage frequently makes it difficult to recognize human behavior. It is possible to completely remove deceptive backdrops by eliminating poses from their surroundings. To illustrate the efficacy of the proposed approach, this section of research looks at how well our methodologies perform in various input datasets, including videos, images, and real-time live-streaming inspection. The outline of human skeleton key points generated for a collection of static activities by humans is shown in Fig. 6.

Here, we offer a Transformer-based methodology for estimating human patterns in human-oriented scenarios using input parameters including head orientations, person spots, and skeleton key points using incorporated inside-the-wild sensor information. Skeletal-based techniques have potential applications in real-time human behavior interpretation. We examined three different kinds of input datasets: live webcam, video, and picture. These input data are either type of dynamic or static in nature. The resulting system achieves optimal accuracy on widely used estimation standards and uses a human tracking dataset that was captured using a camera that was adjusted specifically for the estimation task. Additionally, it needs to take note of the inherent errors of later human pattern estimation. Human-focused assistance When robots are used for autonomous work situations, they might reach average precision in estimation by simply using humans for their location situation. To summarize, the instantaneous pose estimation duties for the fully autonomous navigation data transmission, analysis: and storing method are completed by the vision and DL-based concurrent processors. A unified computing structure and an HRI make it possible to provide results related to human activity monitoring. The outline of human skeleton key points generated for a collection of dynamic activities by humans is displayed in Fig. 7.

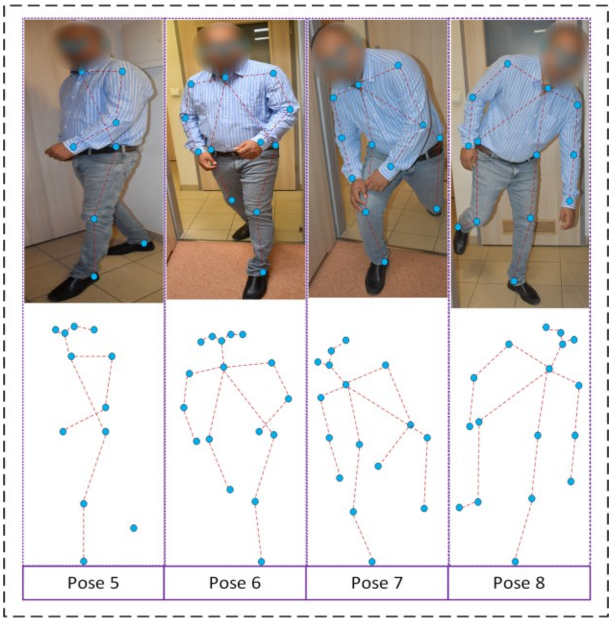

Fig. 7. Simulation output for dynamic human poses with skeleton key points.

To estimate the efficiency and effectiveness of the approach, we used metrics like accuracy score, F1 score, precision score, and recall score. These parameters are successively defined by the following equations: 3, 4, 5, and 6.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

$$Precision = \frac{TP}{TP + FP}, \quad (4)$$

$$Recall = \frac{TP}{TP + FN}, \quad (5)$$

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision}, \quad (6)$$

Where *TP, TN, FP*, and *FN* stand for true positive, true negative, false positive, and false negative, respectively.

A qualitative assessment was carried out using image samples exhibiting various forms of activities. The qualitative assessment is presented in Figs. 4 and 5, in which each first row represents a distinct scenario problem, including glowing and low light indoors, high and poor light outside, shadows, and static and dynamic stances. The key point identification outcome without the human image of the proposed pose estimator is depicted in the second row of Figures. 6 and 7. The primary focus of the work is to utilize advanced human pose estimation approaches for picture-based activity identification and identifying falls. The comparative experiment findings are displayed in Figures. 6 and 7 for different types of activities. The following analysis is done on the numerical outcomes: First, compared to the static-view technique, the dynamic-view technique's inference accuracy is lower. This is why the quality of pose detection is fully determined by the DNN model's capabilities in static scenes, while the dynamic-scene settings have missed various key points of the human skeleton because of occultation in input data. High background semantics seriously impair the estimation strategy's functionality. The optimum inference accuracy is obtained for the static pose process's use of an intention semantic inducer, which extracts the best possible number of key points for the human skeleton. Ultimately, the estimation strategy accuracy falls further when the dynamic-view and static-view approaches are applied in tandem with the concurrent rise in ambient and intentionality semantics.

$$Rate\ of\ accuracy = \frac{No.\ of\ detected\ key\ points}{Total\ key\ points} \times 100 \quad (7)$$

The rate of accuracy of output for each human pose can be calculated by using equations (7), which are depicted in Figure 8. The rate of accuracy for every visible key point that is exclusively gathered by the webcam is 100%; that means if the human body is completely in the frame of the webcam, then the rate of accuracy is highest. The rate of accuracy for pose 4 is just 44.44% because only half a human body is visible to the camera sensor in this scenario. Thus, it is clear that the overall rate of accuracy of the proposed approach is 100% in the case of no blind spots, and the rate of effectiveness of the proposed approach depends upon the human body's exposure in front of the camera sensor. OpenPose, utilizing OpenCV, is a prominent human pose estimate method that uniquely identifies numerous key points concurrently and associates them to construct skeletons for persons within an image. Techniques such as Mask R-CNN or AlphaPose, although effective, typically employ a two-step methodology: initially identifying persons and subsequently calculating their poses, which might be computationally complicated and less efficient in dense environments. Conversely, OpenPose employs an evolutionary methodology, directly forecasting body segments and their interrelations, hence enhancing its scalability for real-time applications. Although it may encounter difficulties with occlusions or intricate postures in comparison to top-down approaches such as HRNet, which enhance pose estimation using higher-resolution feature maps. OpenPose effectively balances accuracy and efficiency, especially in multi-person contexts; however, emerging techniques are advancing the limits of accuracy and resilience.

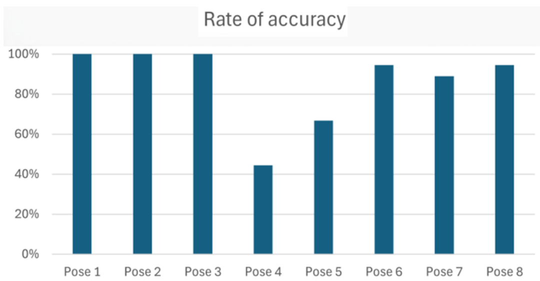

Fig. 8. Accuracy of different human poses in different scenarios.

V. Conclusion

In the realm of human-centered assistance robots, this work addressed the problem of human pose estimation instances with the goal of improving efficient HRI. This work demonstrated whether the relative closeness of individuals under these scenarios might be utilized to intentionally add vision-based human traits and improve prediction precision. The proposed study provides an outstanding basic methodology for human pose estimation using the OpenPose technique with OpenCV. Based on widely accepted prediction criteria and human observation information captured with a laptop webcam specially designed for work estimation, the resulting model achieves optimized accuracy. It also considers the inherent unpredictability of future human intentions. The proposed approach combines real-time flexibility and significant dependable properties; it can also realize notions of HRI, perform effectively, and self-adjust in response to input from users. Multiple whole-body activities, a risky falling workout, and sitting-to-standing activities were all recorded on camera for the suggested system test, and each image was sent into the system. The findings demonstrate that the lateral motions of the elbow, knee, hip, and shoulder joints vary rapidly and dramatically, offering a wealth of details for activity detection.

The suggested pose estimation technique might be used in subsequent years to track dementia and Parkinson's disease patients as well as evaluate building workers' movements with the aim of compiling a record of joint positions for human activities in specific locations. The goal of subsequent studies in human pose estimation for HRI is probably to improve the system's resilience and versatility in a range of settings. The task involves developing algorithms that can deal with occlusions, changing lighting, and a wide range of human body types. Additionally, there is increasing interest in enhancing precision and reliability by utilizing multifunctional monitoring approaches, including wearable gadgets and proximity sensors. To get a deeper understanding of the basic mechanisms of human motion for optimized HRI, combining DL with models based on physics can be an additional line of study in the future. Furthermore, research endeavors can prioritize optimizing real-time performance to facilitate smooth communication between humans and robotic devices in variable settings.

References

[1] T. B. Sheridan, “Human–robot interaction: status and challenges,” Human Factors: *The Journal of the Human Factors and Ergonomics Society*, vol. 58, issue 4, 2016, pp. 525-532. https://doi.org/10.1177/0018720816644

[2] R. Raj and A. Kos, “Discussion on different controllers used for the navigation of mobile robot,” *International Journal of Electronics and Telecommunications*, vol. 70, issue 1, 2024, pp. 229-239. https://doi.org/10.24425/ijet.2024.149535

[3] R. Raj and A. Kos, “Intelligent mobile robot navigation in unknown and complex environment using reinforcement learning technique,” *Scientific Reports*, vol. 14, pp. 22852. https://doi.org/10.1038/s41598-024-72857-3

[4] R. Raj and A. Kos, “An Optimized Energy and Time Constraints-Based Path Planning for the Navigation of Mobile Robots Using an Intelligent Particle Swarm Optimization Technique,” *Applied Sciences*, vol. 13, no. 17, p. 9667, Aug. 2023. https://doi.org/10.3390/app13179667

[5] R. Raj and A. Kos, “An Extensive Study of Convolutional Neural Networks: Applications in Computer Vision for Improved Robotics Perceptions,” *Sensors*, vol. 25, no. 4, pp. 1033, 2025. https://doi.org/10.3390/s25041033

[6] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA*, 2014, pp. 1653-1660. https://doi.org/10.1109/CVPR.2014.214

[7] K. Sun, B. Xiao, D. Liu and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA*, 2019, pp. 5686-5696. https://doi.org/10.1109/CVPR.2019.00584

[8] C. Dong and G. Du, "An enhanced real-time human pose estimation method based on modified YOLOv8 framework," *Scientific Reports*, vol. 14, 2024, pp. 8012. https://doi.org/10.1038/s41598-024-58146-z

[9] Y. -L. Hsu, H. -C. Chang and Y. -J. Chiu, "Wearable Sport Activity Classification Based on Deep Convolutional Neural Network," in *IEEE Access*, vol. 7, pp. 170199-170212, 2019. https://doi.org/10.1109/ACCESS.2019.2955545

[10] T. Sharma, B. Debaque, N. Duclos, A. Chehri, B. Kinder, and P. Fortier, "Deep Learning-Based Object Detection and Scene Perception under Bad Weather Conditions," Electronics, vol. 11, pp. 563, 2022. https://doi.org/10.3390/electronics11040563

[11] R. Raj and A. Kos, "Study of Human–Robot Interactions for Assistive Robots Using Machine Learning and Sensor Fusion Technologies," Electronics, vol. 13, 3285, 2024. https://doi.org/10.3390/electronics13163285

[12] Z. Liu, X. Lu, W. Liu, W. Qi and H. Su, "Human-Robot Collaboration Through a Multi-Scale Graph Convolution Neural Network With Temporal Attention," in IEEE Robotics and Automation Letters, vol. 9, no. 3, pp. 2248-2255, March 2024, https://doi.org/10.1109/LRA.2024.3355752

[13] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019. https://doi.org/10.1126/science.aat8414

[14] C. K. Lakde and P. S. Prasad, "Navigation system for visually impaired people," 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), Melmaruvathur, India, 2015, pp. 0093-0098, https://doi.org/10.1109/ICCPEIC.2015.7259447

[15] S. Kumar KN, R. Sathish, S. Vinayak and T. Parasad Pandit, "Braille Assistance System for Visually Impaired, Blind & Deaf-Mute people in Indoor & Outdoor Application," 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2019, pp. 1505-1509, https://doi.org/10.1109/RTEICT46194.2019.9016765

[16] A. Thomaz, G. Hoffman, and M. Cakmak, "Computational Human-Robot Interaction," *Foundations and Trends in Robotics*, vol. 4, no. 2-3, pp. 104–223, 2016. https://doi.org/10.1561/2300000049

[17] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," Robotics and Autonomous Systems, vol. 42, no. 3–4, pp. 143–166, 2003. https://doi.org/10.1016/S0921-8890(02)00372-X

[18] J. Fasola and M. J. Mataric, "Using Socially Assistive Human–Robot Interaction to Motivate Physical Exercise for Older Adults," in Proceedings of the IEEE, vol. 100, no. 8, pp. 2512-2526, Aug. 2012, https://doi.org/10.1109/JPROC.2012.2200539

[19] Y. Cheng, P. Yi, R. Liu, J. Dong, D. Zhou and Q. Zhang, "Human-robot Interaction Method Combining Human Pose Estimation and Motion Intention Recognition," 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 2021, pp. 958-963, https://doi.org/10.1109/CSCWD49262.2021.9437772

[20] K. Ashley, R. Alqasemi and R. Dubey, "Robotic assistance for performing vocational rehabilitation activities using BaxBot," 2017 International Conference on Rehabilitation Robotics (ICORR), London, UK, 2017, pp. 977-982, https://doi.org/10.1109/ICORR.2017.8009376

[21] R. Raj and A. Kos, "Learning the Dynamics of Human Patterns for Autonomous Navigation," 2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), Gdynia, Poland, 2024, pp. 1-6, https://doi.org/10.1109/CPE-POWERENG60842.2024.10604363

[22] C. Ionescu, D. Papava, V. Olaru and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 7, pp. 1325-1339, July 2014, https://doi.org/10.1109/TPAMI.2013.248

[23] N. A. Choudhury and B. Soni, "Enhanced Complex Human Activity Recognition System: A Proficient Deep Learning Framework Exploiting

Physiological Sensors and Feature Learning," in IEEE Sensors Letters, vol. 7, no. 11, pp. 1-4, Nov. 2023, Art no. 6008104, https://doi.org/10.1109/LSENS.2023.3326126

[24] U. E. Ogenyi, J. Liu, C. Yang, Z. Ju and H. Liu, "Physical Human–Robot Collaboration: Robotic Systems, Learning Methods, Collaborative Strategies, Sensors, and Actuators," in IEEE Transactions on Cybernetics, vol. 51, no. 4, pp. 1888-1901, April 2021, https://doi.org/10.1109/TCYB.2019.2947532

[25] C. Zhu and W. Sheng, "Wearable Sensor-Based Hand Gesture and Daily Activity Recognition for Robot-Assisted Living," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 41, no. 3, pp. 569-573, May 2011, https://doi.org/10.1109/TSMCA.2010.2093883

[26] C. Xu, X. Yu, Z. Wang and L. Ou, "Multi-View Human Pose Estimation in Human-Robot Interaction," IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 2020, pp. 4769-4775, https://doi.org/10.1109/IECON43393.2020.9255211

[27] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, https://doi.org/10.1109/TPAMI.2019.2929257

[28] H. Cao, L. Dirnberger, D. Bernardini, C. Piazza, and M. Caccamo, "6IMPOSE: bridging the reality gap in 6D pose estimation for robotic grasping," Frontiers in Robotics and AI, vol. 10, 2023. https://doi.org/10.3389/frobt.2023.1176492

[29] R. Huo, Q. Gao, J. Qi, and Z. Ju, "3D Human Pose Estimation in Video for Human-Computer/Robot Interaction," In: H. Yang, et al., Intelligent Robotics and Applications. ICIRA 2023. Lecture Notes in Computer Science, Springer, Singapore, 14273, 2023. https://doi.org/10.1007/978-981-99-6498-7_16

[30] Md. A.-A. Bhuiyan, C. H. Liu, and H. Ueno, "On Pose Estimation for Human-Robot Symbiosis," International Journal of Advanced Robotic Systems, vol. 5, 2008. https://doi.org/10.5772/5663

[31] A. Amorim, D. Guimares, T. Mendona, P. Neto, P. Costa, and A. P. Moreira, "Robust human position estimation in cooperative robotic cells," Robotics and Computer-Integrated Manufacturing, vol. 67, no. 102035, 2021. https://doi.org/10.1016/j.rcim.2020.102035

[32] M. Lombardi, E. Maiettini, D. De Tommaso, A. Wykowska, and L. Natale, "Toward an Attentive Robotic Architecture: Learning-Based Mutual Gaze Estimation in Human–Robot Interaction," Frontiers in Robotics and AI, vol. 9, 2022. https://doi.org/10.3389/frobt.2022.770165

[33] S. Saadatnejad et al., "Toward Reliable Human Pose Forecasting With Uncertainty," in IEEE Robotics and Automation Letters, vol. 9, no. 5, pp. 4447-4454, May 2024, https://doi.org/10.1109/LRA.2024.3374188

[34] J. Fan, P. Zheng, S. Li and L. Wang, "An Integrated Hand-Object Dense Pose Estimation Approach With Explicit Occlusion Awareness for Human-Robot Collaborative Disassembly," in IEEE Transactions on Automation Science and Engineering, vol. 21, no. 1, pp. 147-156, Jan. 2024, https://doi.org/10.1109/TASE.2022.3215584

[35] S. Yang, W. D. Kim, H. Park, S. Min, H. Han and J. Kim, "In-Hand Object Classification and Pose Estimation With Sim-to-Real Tactile Transfer for Robotic Manipulation," in IEEE Robotics and Automation Letters, vol. 9, no. 1, pp. 659-666, Jan. 2024, https://doi.org/10.1109/LRA.2023.3334971

[36] M. Salimi, J. J. M. Machado, and J. M. R. S. Tavares, "Using Deep Neural Networks for Human Fall Detection Based on Pose Estimation," Sensors, vol. 22, no. 4544, 2022. https://doi.org/10.3390/s22124544

[37] G. Lan, Y. Wu, F. Hu and Q. Hao, "Vision-Based Human Pose Estimation via Deep Learning: A Survey," in IEEE Transactions on Human-Machine Systems, vol. 53, no. 1, pp. 253-268, Feb. 2023, https://doi.org/10.1109/THMS.2022.3219242

[38] A. Simoni, G. Borghi, L. Garattoni, G. Francesca and R. Vezzani, "D-SPDH: Improving 3D Robot Pose Estimation in Sim2Real Scenario via Depth Data," in IEEE Access, vol. 12, pp. 166660-166673, 2024, https://doi.org/10.1109/ACCESS.2024.3492812

[39] C. Zimmermann, T. Welschehold, C. Dornhege, W. Burgard and T. Brox, "3D Human Pose Estimation in RGBD Images for Robotic Task Learning," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 2018, pp. 1986-1992, https://doi.org/10.1109/ICRA.2018.8462833

[40] T. A. J. Schoonderwoerd, W. Jorritsma, M. A. Neerincx, and K. Van Den Bosch, "Human-Centered XAI: Developing Design Patterns for Explanations of Clinical Decision Support Systems," International Journal of Human-Computer Studies, vol. 154, no. 102684, 2021. https://doi.org/10.1016/j.ijhcs.2021.102684

[41] I. Grishchenko, et al., "BlazePose GHUM Holistic: Real-time 3D human landmarks and pose estimation," 2022. https://doi.org/10.48550/arXiv.2206.11678

[42] T. Salzmann, H. -T. L. Chiang, M. Ryll, D. Sadigh, C. Parada and A. Bewley, "Robots That Can See: Leveraging Human Pose for Trajectory Prediction," in IEEE Robotics and Automation Letters, vol. 8, no. 11, pp. 7090-7097, Nov. 2023, https://doi.org/10.1109/LRA.2023.3312035

[43] S. Mroz et al., "Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose," 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART), Paris / Créteil, France, 2021, pp. 1-4, https://doi.org/10.1109/BioSMART54244.2021.9677850

[44] A. Vaswani, "Attention Is All You Need," 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017. https://doi.org/10.48550/arXiv.1706.03762

[45] Y. Zhang, G. Tian and X. Shao, "Safe and Efficient Robot Manipulation: Task-Oriented Environment Modeling and Object Pose Estimation," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 7502412, https://doi.org/10.1109/TIM.2021.3071222

[46] S. Juraev, A. Ghimire, J. Alikhanov, V. Kakani and H. Kim, "Exploring Human Pose Estimation and the Usage of Synthetic Data for Elderly Fall Detection in Real-World Surveillance," in IEEE Access, vol. 10, pp. 94249-94261, 2022, https://doi.org/10.1109/ACCESS.2022.3203174

[47] S. Li, K. Milligan, P. Blythe, et al., "Exploring the role of human-following robots in supporting the mobility and wellbeing of older people," Scientific Reports, vol. 13, no. 6512, 2023. https://doi.org/10.1038/s41598-023-33837-1