Application of Deep Learning Models Based on EfficientDet and OpenPose in User-Oriented Motion Rehabilitation Robot Control

Mingxiu Sui[1*,a], Liheng Jiang[2,b], Tianyi Lyu[3], Han Wang[4], Li Zhou[5], Peiyuan Chen[6] , Ammar Alhosain[7]

[1] University of Iowa, Iowa City, Iowa 52242, USA; suimingx@gmail.com

[2] New York University, New York, NY 10012, USA; lj1070@nyu.edu

[3] College of Engineering, Northeastern University, Boston, MA 02115, USA; lyu.t.iot@gmail.com

[4]1000 Elements Way, Irvine, CA, 92612, USA; hannah.Wong2021@gmail.com

[5] Faculty of Management, McGill University, Montreal, QC H3B0C7, Canada; lzhou1068@gmail.com

[6] School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon 97333, US; peiyuantony@gmail.com

[7] Department of Public Health and Community Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. doctorammar@gmail.com

*Corresponding Author: suimingx@gmail.com

[a.b.]Equal contribution

DOI: https://doi.org/10.30212/JITI.202402.012

ABSTRACT

This study addresses the critical challenges in rehabilitation robotics, specifically in environmental adaptability, precision in action recognition, and personalized patient care. We introduce the EfficientDet-OpenPose-DRL network, a novel integration of EfficientDet for accurate human and object detection, OpenPose for precise motion tracking, and Deep Reinforcement Learning (DRL) for optimizing rehabilitation strategies. The main contribution of this article lies in the development of this integrated model, which not only enhances environmental perception and action recognition but also improves adaptive decision-making in real-time rehabilitation scenarios. This model enables personalized, adaptable rehabilitation by leveraging advanced computer vision and deep learning techniques. Experimental results demonstrate significant improvements over existing methods, offering enhanced precision, safety, and patient-specific rehabilitation outcomes. This work contributes to the advancement of human-centric rehabilitation technologies, paving the way for more effective and interactive healthcare solutions.

Keywords: Computer Vision, Deep Learning, Robot, Rehabilitation Therapies, EfficientDet

1. Introduction

Rehabilitation robots represent a significant breakthrough in medical technology, offering an innovative approach to rehabilitation therapy[1]. These robots provide patients with an efficient and interactive rehabilitation experience by simulating, assisting[2], or enhancing human movement. Despite the advances in rehabilitation robotics, existing research reveals several critical gaps, particularly in the areas of environmental adaptability, precision in action recognition, and the ability to provide personalized, adaptive care[3,4]. Current systems often struggle to accurately interpret complex human movements and adapt to the dynamic needs of individual patients in real-time[5], limiting their effectiveness in diverse rehabilitation scenarios.

With the rapid development of computer vision technology, the level of intelligence in rehabilitation robots has significantly improved. Computer vision technology enables robots to capture and understand their surroundings through cameras, allowing them to process and interpret complex visual information[6,7]. The essence of this technology is to endow robots with a profound understanding of human movement, thereby ensuring patient safety, accurately assessing treatment progress, and dynamically adjusting treatment plans based on real-time data. For instance, robots can analyze patients' body movements and postures, identify incorrect movement patterns[8], and promptly adjust to prevent potential harm, ensuring the safety and effectiveness of the rehabilitation process[9].

Computer vision technology plays an indispensable role in advancing the research and development of rehabilitation robots[10,11]. It significantly enhances the robots' ability to recognize and mimic human movement patterns. Through advanced image processing and analysis techniques, robots can precisely capture and interpret each movement of the user, ensuring the accuracy and effectiveness of rehabilitation movements[12]. Computer vision technology not only improves the performance of robots but also, through continuous learning and adaptation, enhances their capability to handle complex and unknown rehabilitation scenarios[13,14]. As this technology continues to progress and innovate, we anticipate witnessing more breakthroughs in rehabilitation robots, bringing more personalized, interactive, and efficient rehabilitation experiences to patients.

However, the application of computer vision technology in rehabilitation robots is not without challenges. The complexity of the environment, such as changing lighting conditions, occlusions, and background interference, poses significant threats to the accuracy and reliability of the vision system[15]. Furthermore, the system must exhibit a high degree of flexibility and adaptability to cater to the personalized rehabilitation needs of different patients[16]. Robots need to process a large amount of visual data in real-time and make rapid decisions, placing strict demands on the real-time performance of algorithms and computational resources[17,18]. The complexity of human-robot interaction also presents challenges in designing and implementing robot functionalities. Robots need to accurately interpret human intentions and movements while ensuring safety throughout the rehabilitation process[19]. This requires the computer vision system to be highly precise and robust, as well as to closely collaborate with other control systems of the robot to ensure coordinated responses, especially in emergency situations to prevent harm.

In light of the shortcomings in previous work, we propose the EfficientDet-OpenPose-DRL network model. Our focus is on enhancing environmental perception, improving precision in action recognition, and optimizing adaptive decision-making processes, thereby advancing the capabilities of rehabilitation robotics. This model integrates three advanced technologies, aiming to improve the environmental perception, action understanding precision, and adaptive decision-making capabilities of rehabilitation robots. Firstly, the EfficientDet model plays a crucial role in this framework by analyzing images deeply through its efficient network structure, identifying and locating the human body and other key objects in images. This step is vital for ensuring that robots can accurately recognize the patient's position and posture under various environmental conditions, thus providing a solid foundation for subsequent action analysis and adaptive decision-making.

Next, the OpenPose model is introduced to provide a detailed analysis of the patient's posture and movements. By precisely detecting and tracking human keypoints, OpenPose offers detailed information about the patient's movements. This information is crucial for robots as it enables them to perform accurate movement mimicry or provide necessary movement assistance, ensuring the effectiveness and safety of the rehabilitation process. The DRL model optimizes decision-making and behavior. DRL utilizes the environmental and posture information obtained from EfficientDet and OpenPose, adopting reinforcement learning algorithms to optimize the robot's movement strategies. Its introduction endows robots with the ability to autonomously learn and adjust their behavior, allowing them to provide more personalized and adaptive rehabilitation training[20].

By integrating the advantages of these three technologies, the EfficientDet-OpenPose-DRL network model not only improves the comprehensiveness of environmental perception and the accuracy of action understanding but also enhances the robot's adaptive decision-making capabilities. This integration allows our model to provide precise and personalized rehabilitation assistance while maintaining high efficiency. The introduction of DRL, in particular, enables robots to dynamically learn and adjust based on feedback, achieving continuous optimization. This optimization is reflected not only in improving therapeutic effects but also in enhancing patient safety and comfort in the rehabilitation experience. Therefore, our proposed network model offers an effective solution to the challenges faced by traditional rehabilitation robots in environmental perception, action understanding, and adaptive decision-making, promising to bring more efficient, precise, and personalized rehabilitation experiences to patients.

In our study, our main contributions are reflected in the following three areas:

- We have successfully integrated three advanced technologies, EfficientDet, OpenPose, and DRL, to construct an innovative control model for motor rehabilitation robots. This integrated model demonstrates significant advantages in improving environment perception, movement understanding, and adaptive decision making.
- Our model significantly improves the rehabilitation robot's ability to understand human movements through accurate motion capture and analysis. This increased accuracy not only makes the rehabilitation process more efficient and effective, but also provides a safer and more comfortable rehabilitation experience for the patient.
- The deep reinforcement learning component we introduced enables the robot to autonomously learn and adjust its behavior based on real-time feedback. This point not only

enhances the robot's adaptability and personalized treatment capability, but also enables our model to effectively cope with various complex and unknown rehabilitation scenarios, providing new perspectives and methods for future research and application of rehabilitation robots.

The subsequent sections of this paper are meticulously structured to build upon the foundational concepts introduced here. Section 2 will delve into a critical examination of existing studies, highlighting the pivotal advancements and identifying the prevalent gaps within the current body of research. This will set the stage for Section 3, where we will articulate the detailed methodologies and robust frameworks employed in our study, offering clarity and depth to our research approach. Following this, the Section 4 will provide empirical evidence to substantiate our claims, showcasing the efficacy of our methods through methodical testing and nuanced analysis. Concluding the paper, Section 5 will synthesize the insights garnered from our research, ponder the broader implications of our findings, and propose potential trajectories for future inquiry. Through this structured narrative, we aim to present a coherent and comprehensive discourse, contributing meaningful insights to the field of rehabilitation robotics and beyond.

2. Related Work

2.1 Deep Learning for Human Pose Estimation

The application of deep learning in human pose estimation primarily focuses on using Convolutional Neural Networks (CNNs) to analyze human postures in videos or images[21]. Frameworks like OpenPose, for instance, enable real-time capture and analysis of user movements[22], identifying key points such as the positions of arms, legs, and head[23]. This is vital for rehabilitation robots to understand the capabilities and limitations of users[24], as well as their progress during rehabilitation. A key advantage of this technology is its real-time processing capability, providing immediate analysis and response to a user's posture and movements during rehabilitation exercises[25].

Despite its effectiveness in real-time posture analysis, this approach faces several challenges. Deep learning models typically require extensive labeled data for training, which may be difficult to acquire in specific rehabilitation scenarios[26]. Additionally, the accuracy of these methods may decrease in complex backgrounds or when parts of the user's body are obscured. Furthermore, real-time processing of high-resolution video streams demands significant computational resources[27], potentially limiting its application in resource-constrained environments such as remote rehabilitation settings or mobile devices[28].

2.2 Machine Vision-Based Motion Anomaly Detection

Machine vision-based motion anomaly detection systems utilize machine learning algorithms[29], such as Support Vector Machines (SVM) and Decision Trees, to analyze motion data captured by cameras[30]. These systems aim to identify incorrect movement patterns during physical therapy for rehabilitation patients[31]. For instance, if a patient executes a particular movement with incorrect posture or insufficient range, the system can detect these issues and provide feedback[32]. This not only helps in avoiding potential injuries but also ensures the efficiency and effectiveness of rehabilitation training[33,34].

While such systems are valuable for motion monitoring and feedback, they also have limitations[35]. A major concern is that the accuracy and effectiveness of the system heavily depend on the quality and diversity of the training data. If the dataset is inadequate or not comprehensive, the system may fail to recognize all types of motion anomalies[36]. Additionally, these generalized systems might struggle to accommodate individual differences, especially when dealing with patients with complex medical conditions.

2.3 Multi-Modal Data Fusion in Rehabilitation Robot Control

Research in multi-modal data fusion for rehabilitation robot control attempts to enhance the effectiveness of rehabilitation robots by integrating data from various sensors (e.g., visual data, pressure sensor data, and inertial measurement unit data)[37]. Deep neural networks play a role here, used for comprehensive analysis of these multi-source data, thereby achieving a more accurate understanding of patient movements[38]. This technology enables rehabilitation robots to adapt flexibly to the specific needs of each patient, such as adjusting support strength or range of motion, to meet individual rehabilitation goals and constraints[39].

Although multi-modal data fusion offers a more comprehensive perspective, it also introduces new challenges[40]. Firstly, processing and analyzing data from various sources require complex algorithms and potentially higher computational resources. This may result in a decline in real-time responsiveness, especially in resource-limited environments[41]. Secondly, temporal synchronization and fusion of data from different sensors can increase technical complexity[42]. Moreover, an overreliance on technology might overlook the importance of patient personal experience and intuitive feedback, which are crucial for effective rehabilitation[43].

2.4 Augmented Reality-Based Interactive Rehabilitation Training

The emerging field of interactive rehabilitation training combining Augmented Reality (AR) with computer vision is an innovative approach[44]. In this method, patients engage in physical rehabilitation exercises by interacting with virtual objects in an augmented reality environment[45]. This technology uses sophisticated image recognition and spatial mapping techniques to ensure effective integration of virtual objects with the real-world environment[46]. For instance, patients might be asked to complete specific hand movements or walking tasks in a virtual setting, with the system providing real-time feedback based on their performance[47]. This not only enhances the engagement and motivation of rehabilitation exercises but also offers challenging tasks to facilitate their rehabilitation process[48].

Although the application of augmented reality technology in rehabilitation training holds significant potential, it also faces some constraints. Firstly, the implementation of this technology might require expensive hardware and software resources[49], adding to the cost. Secondly, for certain patient groups, such as the elderly or those with certain neurological disorders, augmented reality environments might cause discomfort or cognitive confusion[50]. Moreover, to effectively use augmented reality for rehabilitation training, highly customized program designs are required to accommodate various physical limitations and rehabilitation goals[51].

3. Materials and Methods

3.1 Overview of Our Network

Our model, the EfficientDet-OpenPose-DRL network, is an intricate system designed to revolutionize the control mechanisms of user-based motion rehabilitation robots through the application of advanced computer vision technology. The model is composed of several critical components that work in concert to interpret, assist, and enhance human movement for rehabilitative purposes. The overall model structure is shown in Fig 1. This graphical representation reflects the complex yet efficient working mechanism of the model.

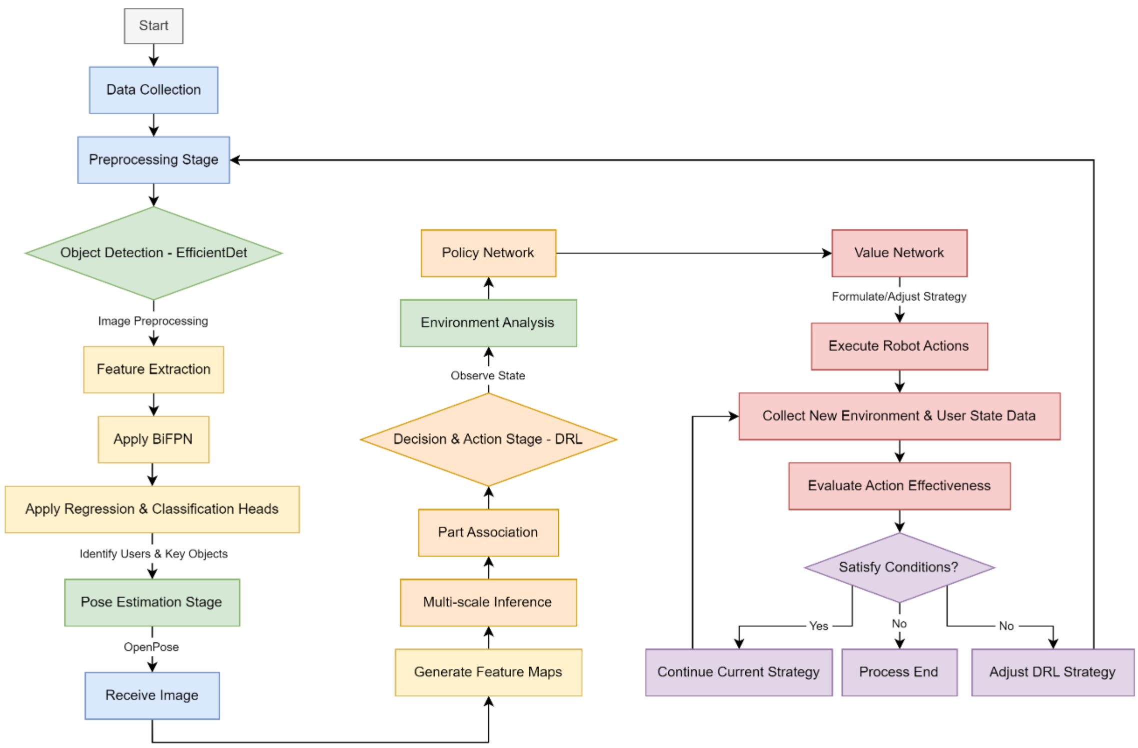

Figure 1. Overall structure diagram of the model.

At the onset, we utilize EfficientDet for its object detection capabilities. It serves as the foundation of our model's environmental understanding, distinguishing between the user and other significant objects within the rehabilitation space. This detection is vital for ensuring that the robot can operate within a dynamic environment and respond to changes or obstacles that may arise during rehabilitation exercises.

Following object detection, the Bi-directional Feature Pyramid Network (BiFPN) is applied for nuanced feature extraction. This step is key in the process as it enriches the model's perception, allowing for the detection of complex movements and the understanding of intricate physical environments, which is essential for precise movement assistance.

The OpenPose model plays a pivotal role in our system by providing real-time multi-person pose estimation. It receives images from the preceding stages and processes them to identify key points on the user's body. The detailed skeletal structure that OpenPose constructs informs the robot of the specific postures and movements the user is making, enabling accurate replication or support of movements.

Central to our model is the implementation of DRL, which allows the robot to make informed decisions and adapt its actions to the user's needs. The DRL stage utilizes a policy network to predict the best course of action and a value network to estimate the potential outcome of these actions. By continuously learning from the environment and the user's state, the model refines its strategy to optimize the rehabilitation process.

The construction of our network starts with the preprocessing of image data collected from cameras, which involves standardizing image formats through normalization and resizing to align with model input requirements. Subsequent to preprocessing, the image data is channeled into the EfficientDet model for object detection. Here, EfficientDet discerns the human body and other vital objects, furnishing precise location and pose data. This data is then passed on to the OpenPose model, which conducts a meticulous human pose estimation by pinpointing key human joints and body parts. The combined outputs from EfficientDet and OpenPose furnish the DRL model with detailed insights into the current environmental context and the patient's physical state. The DRL model then employs a reinforcement learning algorithm to appraise this information and refine the robot's motion strategies accordingly, ensuring that the actions taken are both appropriate and conducive to the user's rehabilitation journey.

The proposal of our integrated EfficientDet-OpenPose-DRL model marks a transformative step in the realm of motion rehabilitation, merging the precision of computer vision technology with the adaptability of intelligent learning systems. This synergy significantly enhances the robot's capability for personalized and responsive rehabilitation assistance. Unlike traditional robotic systems that may only perform pre-programmed tasks, our model offers a sophisticated analysis of user movement and an adaptive learning approach, enabling the rehabilitation robot to evolve its strategies in real time. The model's deep reinforcement learning component is particularly crucial, allowing for dynamic adjustments based on continuous feedback, which leads to ongoing refinement of the therapy. This ongoing optimization process not only augments the therapeutic effects but also prioritizes patient safety and comfort. By addressing the inherent challenges in environmental perception, movement interpretation, and decision-making flexibility, our model stands to revolutionize the field of rehabilitation robotics. It promises more effective, precise, and individualized treatment plans, ultimately improving the quality of life for patients undergoing rehabilitation and representing a significant leap forward in rehabilitative care.

3.2 EfficientDet

EfficientDet, proposed by the Google Research team in 2020[52], represents a formidable advancement in the realm of object detection, particularly addressing the intricate challenge of multiscale feature representation. Its novelty lies in the integration of EfficientNet as its backbone network[53], augmented with an enhanced Bidirectional Feature Pyramid Network (BiFPN), and the utilization of compound scaling technology[54]. EfficientNet, a convolutional neural network, was developed through neural architecture search and compound scaling methods, aimed at achieving high accuracy with lower computational costs.

Central to the prowess of EfficientDet are two core components: BiFPN for effective feature fusion and the EfficientDet detector for precise object detection[55]. BiFPN improves upon the traditional feature pyramid network structure by facilitating bidirectional information flow between feature layers of varying scales. This mechanism effectively integrates high-level semantic information with low-level details, thus enhancing the model's detection capabilities[56]. The essence of BiFPN lies in its approach to assigning customized weights to features during the integration of bottom-up and top-down multiscale features, rooted in three key insights:

1. Nodes with a solitary input edge without feature fusion contribute marginally to the feature network, indicating the need for differential weighting of features based on their integration within the network.
2. An additional edge directly connecting input and output nodes at identical levels facilitates the amalgamation of an extensive array of features at a minimal expense, augmenting the network's capacity to efficiently assimilate diverse feature information.
3. Each bidirectional (top-down and bottom-up) pathway is treated as a distinct feature layer, applied repetitively several times to ensure profound high-level feature fusion, guaranteeing comprehensive integration of features across varied scales and contexts.

Moreover, the compound scaling technique of EfficientDet judiciously adjusts the depth, width, and input resolution of the network, allowing the model to maintain robust performance across different scales and computational budgets. This characteristic not only makes EfficientDet stand out in terms of accuracy but also in computational efficiency. Compared to other state-of-the-art object detection models, EfficientDet achieves higher accuracy with equal or lower computational resources, making it suitable for resource-constrained environments, such as mobile devices and embedded systems.

The allocation of weights to features within BiFPN is meticulously designed and can be operationalized through three distinct methodologies. The foremost approach, labeled "unbounded fusion," champions a dynamic and adaptable strategy for assigning weights to features, guaranteeing that each feature's unique properties and inherent information are appropriately recognized and utilized. This innovative method of weighting features sets the stage for a more sophisticated and efficacious amalgamation of features, the formula form is expressed as:

$$O = \sum_{i} w_i \cdot I_i \tag{1}$$

where w_i represents a learnable weight, it can manifest as a scalar, a vector, or a multidimensional tensor corresponding to each feature, channel, or pixel, respectively.

This approach allows for a granular and nuanced adjustment of the feature contributions during the fusion process. However, a notable drawback of this technique lies in the scalar weight being unbounded. This absence of bounds on the scalar weight can introduce challenges during the training phase, potentially leading to instability or convergence issues, thus necessitating careful management and adjustment within the training process.

The second method employs a "Softmax fusion" technique, which operates as detailed below:

$$O = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} \cdot I_i \qquad (2)$$

While this approach effectively confines the weight to the range [0, 1] and ensures the preservation of feature contribution, it is important to note that it comes with a relatively higher latency cost.

The third approach utilizes a "fast normalized fusion" process, delineated as follows:

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i \qquad (3)$$

where w_i is guaranteed to be non-negative through the application of a ReLU function, and an epsilon value $\epsilon = 0.0001$ is introduced to maintain numerical stability. This technique is particularly advantageous due to its capacity to provide bounded weights alongside a more rapid response.

Delving into the structure of EfficientDet (as illustrated in Fig 2), it employs a one-stage detection framework. The model utilizes EfficientNet as its backbone, upon which the BiFPN is superimposed as a feature network. This network harvests features spanning from the third to the seventh level of the backbone. Subsequently, the combined output from the BiFPN is channeled into a network responsible for predicting classes and bounding boxes.

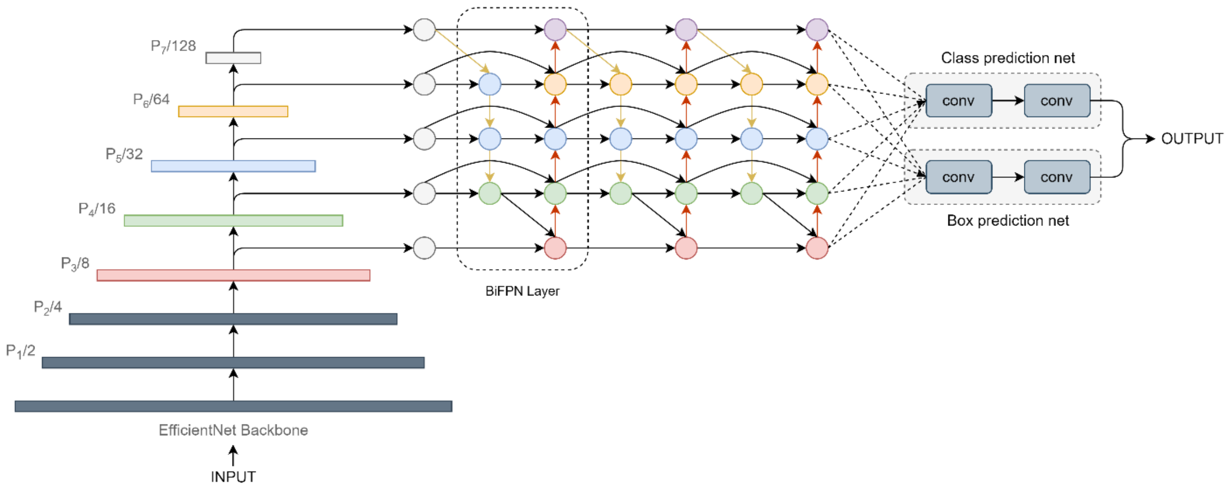

Figure 2. The structure of EfficientDet.

The scalability of EfficientDet renders it applicable across a wide range of applications, from small-scale devices to large servers. This ability to maintain efficient object detection performance on various devices positions it as an ideal choice for multiple computer vision tasks and practical applications. Through these innovations, EfficientDet has achieved significant breakthroughs in the field of object detection, particularly in enhancing detection accuracy and reducing computational costs. This integration of EfficientDet into our model forms the foundation of our approach, enabling accurate environmental perception and laying the groundwork for further action analysis and adaptive decision-making.

The EfficientDet model plays a pivotal role in our comprehensive model, being crucial for enhancing the environmental perception, target recognition, and adaptive decision-making capabilities of rehabilitation robots. As a key component of our model, EfficientDet significantly enhances the precision of environmental perception and target feature recognition through its advanced object detection functionality. EfficientDet integrates EfficientNet as its backbone network along with the innovative BiFPN, effectively accomplishing the fusion of multiscale features. This integration allows EfficientDet to combine information from different feature levels, providing a rich and detailed view of the environment and offering an accurate data foundation for rehabilitation robots.

In our experiment, the role of EfficientDet is particularly critical. Its efficient object detection capability provides solid technical support for achieving our research goal—enhancing the adaptability and efficiency of rehabilitation robots. By precisely identifying and tracking patient movements, EfficientDet not only ensures the safety and effectiveness of the rehabilitation process but also provides real-time, accurate environmental feedback to the robots, enabling them to adapt to various complex rehabilitation scenarios.

The scalability of EfficientDet further extends its applicability, making it suitable for a wide range of applications from small devices to large servers. This ability to maintain efficient object detection performance across different devices makes it an ideal choice for various computer vision tasks and practical applications. Through these innovations, EfficientDet has achieved significant breakthroughs in the field of object detection, especially in terms of improving detection accuracy and reducing computational costs.

The integration of EfficientDet into our model forms the foundation of our approach, laying the groundwork for precise environmental perception for rehabilitation robots and providing solid support for subsequent action analysis and adaptive decision-making. This integration not only contributes powerful visual processing capabilities to our model at the technical level but also significantly enhances the practicality and intelligence of rehabilitation robots at the application level, proving essential for the success of our experiment.

3.3 OpenPose

OpenPose is a state-of-the-art system for real-time multi-person 2D pose detection developed by the Carnegie Mellon Perceptual Computing Lab[57]. It leverages CNNs like VGG19 for feature extraction, processing input images to extract detailed feature maps. The system then generates Part Confidence Maps (PCM) and Part Affinity Fields (PAF), which represent the confidence of locating body parts in certain pixels and encode the degree of association between different body parts, respectively[58]. OpenPose operates through a multi-stage CNN, where each stage refines the predictions of the previous one. This staged approach allows the network to progressively refine its predictions, starting from detecting prominent keypoints in the initial stages and gradually moving to interpret the more complex relationships between keypoints in the later stages[59]. This iterative refinement is crucial for detecting keypoints that are challenging due to occlusion or subtle appearance.

The real-time performance and accuracy of OpenPose have significant implications across various fields. In healthcare and rehabilitation, it offers a non-invasive way to monitor and analyze patients' postures and movements, facilitating correct exercise performance and enabling remote physiotherapy sessions[60]. In sports analytics, OpenPose assists in analyzing athletes' postures to enhance performance and prevent injuries. Furthermore, its ability to understand and interpret human body language opens new avenues for human-computer interaction[61], making interactions with machines more natural and intuitive. Overall, OpenPose represents a convergence of deep learning and visual geometry, offering a flexible and powerful tool for human pose estimation with broad applications across different industries and research areas[62].

As depicted in the provided Figure 3, OpenPose's network architecture begins with the VGG19 backbone network, laying the groundwork for feature extraction from the input images. These features then proceed into a series of stage modules arranged in sequence, each sharing an identical configuration and purpose. Within each stage, the network splits into two branches: one dedicated to generating PCM and the other for producing PAF. Both PCM and PAF from each stage are subjected to loss calculations, and the total loss of the network is the cumulative sum of these individual stage losses.

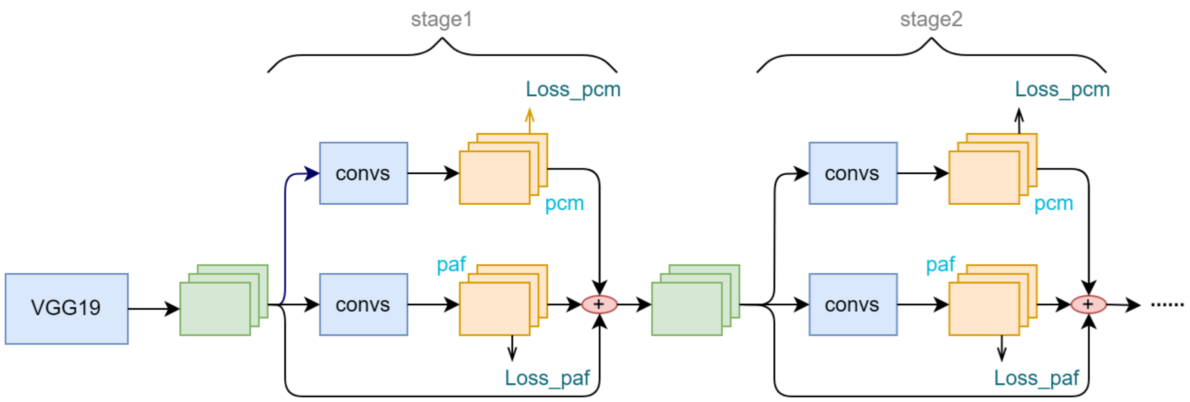

Figure 3. The structure of OpenPose.

The inclusion of multiple stages is instrumental for enhancing the accuracy of keypoint detection. While the first stage provides a preliminary mapping of keypoints, subsequent stages refine and improve upon these initial detections. For instance, if the eyes are detected in the first stage, the subsequent stage may leverage that information to locate the nose more accurately. This staged process allows for an iterative refinement, utilizing the outputs of one stage to inform and improve the results of the next.

This iterative process is crucial for keypoints that are subtle or affected by external factors such as clothing or occlusion. The initial stages capture the more apparent keypoints, while the subsequent stages use these detections as anchors to identify more complex keypoints. This progressive optimization ensures that each keypoint is detected with increasing accuracy, thereby improving the robustness and reliability of the entire network in comprehensively mapping the human pose.

Within the architecture of our integrated model, OpenPose assumes an indispensable role by providing intricate and precise estimations of human body poses, which are crucial for understanding and analyzing the subjects' movements and postures during rehabilitation exercises. OpenPose enhances the object detection capabilities of EfficientDet, which determines the subject's location within the environment, by offering a fine-grained analysis of the subject's skeletal pose. It produces detailed PCM and PAF, allowing for an exact mapping of individual body parts and their interconnections. The rich pose information gleaned from OpenPose is vital for the DRL component, as it enables the DRL model to make well-informed decisions about the actions that the rehabilitation robot should execute or assist the subject with, tailored to the subject's current pose and movement dynamics.

OpenPose's role is pivotal not just in detection, but also in enhancing the adaptability of the rehabilitation process. It ensures that the rehabilitation robot can follow complex movements with high accuracy, vital for creating a responsive rehabilitation experience that adapts to the nuanced motions of patients. The data acquired by OpenPose continuously informs the DRL component, refining the robot's behavior through iterative interactions with the patient. This integration exemplifies the fusion of sophisticated computer vision with intelligent machine learning techniques, setting new standards for automated rehabilitation systems.

The integration of OpenPose into our model is critical for the robot's ability to perceive, comprehend, and interact with the human element within the rehabilitation environment. It acts as a sophisticated sensory organ, influencing the machine's decision-making algorithms and allowing for adjustments tailored to the patient's immediate therapeutic requirements. This dynamic interaction ensures a personalized and effective rehabilitation session, as the robot can adapt its assistance in real-time based on the precise movement patterns identified by OpenPose. In essence, OpenPose is instrumental for achieving the high level of responsiveness and personalization that is imperative in contemporary rehabilitation robotics, solidifying its essential status in our experimental framework.

3.4 DRL

Deep Reinforcement Learning (DRL) is an amalgamation of the complexities of deep neural networks and the goal-directed dynamics of reinforcement learning. It's particularly adept at processing high-dimensional input data like images and sensor readings[28], drawing on the profound feature extraction capabilities of deep learning. Coupled with reinforcement learning, DRL strategizes actions to optimize long-term cumulative rewards. It embodies a learning paradigm where an intelligent agent masters navigating an environment through a system of rewards[63], reinforcing behaviors that lead to favorable outcomes.

DRL stands out for its ability to interpret complex inputs and learn from the environment[64]. At its core are value functions and action-value functions, essential for estimating the agent's potential benefits in specific states or actions. These are defined by critical mathematical expressions, including the Bellman equation for value functions and the Bellman optimality equation for action-value functions[65]. These equations provide a recursive solution to policy evaluation and improvement, guiding the agent to learn optimal behaviors through interaction with the environment[66]. The policy improvement theorem, a cornerstone of DRL, relies on these functions for policy updates, indicating the actions that maximize the action-value function for a state.

The mathematical underpinnings of DRL are expressed through various equations that describe the process from policy evaluation to policy improvement. The value function, which estimates the benefit for the agent in a specific state, is given by:

$$V^{\pi}(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R_{t+1} | S_0 = s, \pi] \tag{4}$$

Here, $V^{\pi}(s)$ denotes the expected return when starting in state s and following policy π with γ as the discount factor and R_{t+1} as the reward at time t+1.

The action-value function, which assesses the value of taking action a in state s under policy π, is defined as:

$$Q^{\pi}(s,a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R_{t+1} | S_0 = s, A_0 = a, \pi] \tag{5}$$

The Bellman equation offers a recursive solution to the value function:

$$V^{\pi}(s) = \sum_{a \in A} \pi(a|s) \sum_{s' \in S} P(s'|s,a)[R(s,a,s') + \gamma V^{\pi}(s')] \tag{6}$$

For an optimal policy, we turn to the Bellman optimality equation for the action-value function:

$$[Q^{*}(s,a) = \sum_{s' \in S} P(s'|s,a)[R(s,a,s') + \gamma \max_{a'} Q^{*}(s',a')] \tag{7}$$

And finally, policy improvement is achieved through updates based on the action-value function:

$$\pi'(s) = \arg\max_{a \in A} Q^{\pi}(s,a) \tag{8}$$

In this way, the DRL model employs these principles to guide the rehabilitation robot, iterating over actions and refining strategies to optimize for the patient's therapeutic progress, making DRL an indispensable component in the quest for more autonomous and efficient rehabilitation robotics.

In our integrated setup, DRL ingeniously combines the situational awareness provided by EfficientDet with the detailed human biomechanical insights from OpenPose. EfficientDet, with its object detection prowess, identifies the subject and relevant objects in the environment, setting a precise spatial context. OpenPose, on the other hand, offers a granular analysis of the subject's posture and limb positioning. DRL harnesses this combined data to compute the optimal sequence of actions that the rehabilitation robot should undertake, ensuring adaptability to the subject's real-time movements. This seamless integration of perception and action is what endows our model with predictive and responsive control capabilities, essential for a dynamic and interactive rehabilitation process.

DRL's multifaceted contribution to our model is paramount. It doesn't just process inputs; it transforms them into informed decisions, driving the rehabilitation robot's movements and interactions with patients. By doing so, DRL ensures that each movement and interaction is meticulously aligned with the patient's rehabilitation needs, providing a personalized and adaptive therapy experience. This adaptability is crucial in rehabilitation, where patient-specific needs and capabilities can widely vary and evolve. The continuous learning and adaptation of the DRL component ensure that our model effectively caters to a broad spectrum of scenarios, tailoring the rehabilitation experience and optimizing the recovery path for each patient.

Moreover, the self-improving nature of DRL is what truly sets our model apart. It learns from each session, fine-tuning its strategies to enhance therapeutic efficacy. This continuous learning process not only propels our model to new heights of autonomy and efficiency but also centers it firmly around the needs of the patient, marking a significant leap towards advanced, patient-centric rehabilitation robotics. DRL's adaptive intelligence is, therefore, not just a tool but an integral component of our experimental design, ensuring that our model remains at the forefront of innovation in the field of rehabilitation technology

4. Results

4.1 Experimental Datasets

To comprehensively validate our model, this experiment utilizes four distinct datasets: Human3.6M, MPII Human Pose, NTU RGB+D, and PoseTrack. These datasets, each with its unique characteristics and annotations, provide a multifaceted platform to rigorously assess the model's performance across various dimensions of human pose estimation and motion analysis. This comprehensive approach ensures a thorough validation of the model's capabilities, setting the stage for detailed evaluations and insights derived from each specific dataset.

Human3.6M Dataset[67]: Human3.6M stands as one of the most extensive datasets for 3D human pose estimation, highly regarded in the research community for its vast size and the diversity of recorded human activities. The dataset encompasses 3.6 million 3D human poses accompanied by corresponding images. These poses were captured from 11 professional actors, each performing 17 different scenarios, encompassing a wide range of everyday activities such as walking, eating, sitting, discussing, and phone interactions. Recorded using a sophisticated marker-based motion capture system, the dataset provides accurate 3D positions of various body joints, making it a valuable asset for developing and benchmarking 3D pose estimation models. Its extensive use in studies related to action recognition, human-computer interaction, and virtual reality underlines its significance in advancing the field of computer vision and machine learning.

MPII Human Pose Dataset[68]: The MPII Human Pose dataset is a comprehensive collection specifically designed for evaluating articulated human pose estimation algorithms. It includes approximately 25,000 images, capturing over 40,000 individuals with annotated body joints. The annotations are detailed, providing 2D positions of various body joints, which are crucial for understanding human postures in diverse activities. The images in this dataset are sourced from

YouTube videos and represent a broad spectrum of human activities, including daily living activities, sports, and social interactions. The diversity and complexity of scenes in the MPII dataset make it an excellent resource for training robust 2D pose estimation models and benchmarking their performance against real-world scenarios.

NTU RGB+D Dataset[69]: The NTU RGB+D dataset is tailored for research in the areas of 3D human action recognition and pose estimation, offering a substantial volume of RGB+D videos. With around 56,000 videos and 4 million frames captured using Microsoft Kinect v2 cameras, the dataset includes a wide array of 60 different action classes. These classes cover individual actions, two-person interactions, and health-related events, providing a comprehensive overview of human movements. The inclusion of both RGB and depth data adds a rich dimension to the dataset, enabling more nuanced analysis and model development for 3D human action recognition, pose estimation, and even predictive modeling of future actions.

PoseTrack Dataset[70]: PoseTrack focuses on the challenge of human pose estimation and articulated tracking in video sequences, offering a substantial collection of videos that include complex scenes with multiple interacting individuals. The dataset is an invaluable resource for developing and evaluating algorithms on video-based pose estimation and articulated tracking tasks. Each person in the videos is meticulously annotated with body joint positions in each frame, providing a rich set of data for temporal pose estimation tasks. The PoseTrack dataset is instrumental for studies that require understanding the intricacies of human movements and poses over time, making it suitable for applications in sports analytics, action recognition, and interactive systems where temporal dynamics are key.

Every one of these datasets represents a fundamental component within its own area of expertise, providing distinct perspectives and possibilities for furthering research in fields such as human pose estimation, action recognition, and the development of interactive systems. Each dataset's unique attributes contribute significantly to the depth and breadth of insights in these domains, facilitating advancements and innovations.

4.2. Experimental Details

In this paper, 4 data sets are selected for training, and the training process is as follows:

Step 1: Data Processing

Data preprocessing is a crucial step in the development and implementation of the EfficientDet-OpenPose-DRL network model, ensuring the data is optimally prepared for model training and evaluation. The specific steps include:

Data Cleaning: Initially, a vast collection of images and videos from rehabilitation exercises and daily activities showcasing various human postures and movements must be gathered. This step involves removing low-quality data, such as blurry images, severely occluded videos, and irrelevant background information. For instance, images captured under extreme lighting conditions or those that fail to clearly identify human keypoints due to improper shooting angles would be discarded. This process ensures the quality of the data, laying a solid foundation for subsequent model training and evaluation.

Data Standardization: In this step, all image data will be resized to a uniform dimension and resolution, and pixel values will be normalized. This standardization ensures consistency in model input and improves training efficiency. We might choose to resize all images to a uniform size of 1024x1024 pixels and normalize pixel values to range between 0 and 1. Such standardization helps reduce computational complexity during model training and avoids potential numerical instability issues.

Data Splitting: The dataset will be divided into three parts for training, validation, and testing the model. A common division ratio is 70% for the training set, 15% for the validation set, and 15% for the testing set. It's critical to ensure that these datasets are representative in terms of diversity, including different lighting conditions, background settings, and human postures. For example, if the total dataset comprises 10,000 samples, we would select 7,000 samples for the training set, 1,500 for the validation set, and 1,500 for the testing set.

Through these preprocessing steps, we can prepare a high-quality dataset for the development of the EfficientDet-OpenPose-DRL network model. This preparation enhances the model's performance in applications involving rehabilitation robots, ensuring it can accurately understand and assist human movements.

Step 2: Model Training

The model training phase is pivotal in the development of the EfficientDet-OpenPose-DRL network model, as it directly influences the model's effectiveness in real-world rehabilitation robot applications. This phase involves a careful consideration of network parameter settings, model architecture design, and the actual training process. Each of these components plays a critical role in ensuring the model's performance and adaptability.

Network Parameter Settings: In configuring our network parameters, we prioritize optimization for accuracy and computational efficiency. For the EfficientDet model, we select a D3 variant with a backbone of EfficientNet-B3, setting the input image size to 1024x1024 pixels. This choice strikes a balance between computational demand and the ability to process high-resolution images effectively. The learning rate is set at 0.001, with a batch size of 32, to ensure stable convergence without overloading the hardware resources. For the OpenPose component, we employ a learning rate of 0.0004 and a batch size of 16, given its intensive computational requirements for keypoint detection. The DRL module uses an epsilon decay strategy starting from 1.0 to 0.01 over 100,000 training steps, ensuring a gradual shift from exploration to exploitation.

Model Architecture Design: Our model architecture integrates the strengths of EfficientDet for object detection, OpenPose for human pose estimation, and DRL for adaptive decision-making. The EfficientDet module is designed with a compound scaling method, enhancing its ability to detect humans and objects in various conditions. OpenPose is structured to identify 25 key body points, providing detailed posture analysis essential for precise rehabilitation assistance. The DRL model is built upon a custom neural network architecture with three hidden layers, each consisting of 256 neurons, to process the combined input from EfficientDet and OpenPose and output optimized movement strategies. This architecture supports the processing of complex visual information and decision-making in dynamic environments.

Model Training Process: The training process is executed in stages to accommodate the complexity of the integrated model. Initially, the EfficientDet and OpenPose modules are pre-trained on publicly available datasets such as COCO and MPII for 50 epochs each to establish a baseline understanding of object detection and human pose estimation. Subsequently, these modules are fine-tuned on our specific rehabilitation dataset for an additional 30 epochs to adapt to the nuances of rehabilitation scenarios. The DRL module is trained using a simulated environment that replicates various rehabilitation exercises, running for 200,000 steps. Throughout the training, we employ early stopping based on the validation loss to prevent overfitting, with a patience parameter of 10 epochs. This staged training strategy ensures that each component of the model is adequately prepared before their integration, ultimately leading to a cohesive system capable of providing effective and personalized rehabilitation support.

Through this meticulous training process, incorporating specific hyperparameter settings and a carefully designed model architecture, the EfficientDet-OpenPose-DRL network model is optimized for the complex requirements of rehabilitation robotics, promising significant improvements in patient care and support.

Step3: Indicator Comparison Experiment

During this stage, we intend to benchmark our model against other widely recognized models in the fields of regression and classification. To ensure a fair and consistent evaluation, all models will be trained and assessed on identical sets of training and testing data. Following this, we'll conduct a detailed comparison of each model's performance, utilizing key metrics including accuracy, recall, F1 score, and the Area Under the Curve (AUC) measure. This approach will allow us to precisely gauge and contrast the efficacy of each model across various tasks and datasets, providing a comprehensive overview of their capabilities. Descriptions of these performance metrics will be detailed subsequently, offering insights into their significance and calculation.

1. Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives.

2. Recall:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

where TP represents the number of true positives, and FN represents the number of false negatives.

3. F1 Score:

$$F1\, Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{11}$$

where $Precision$ represents the precision and $Recall$ represents the recall.

4. AUC:

$$AUC = \int_0^1 ROC(x)dx \tag{12}$$

where *ROC(x)* represents the relationship between the true positive rate and the false positive rate when *x* is the threshold.

5. Parameters(M):

Count the number of adjustable parameters in the model, in millions.

6. Inference Time(ms):

Measure the time required for the model to perform inference, in milliseconds.

7. Flops(G):

Count the number of floating-point operations required for the model to perform inference, in billions.

8. Training Time(s):

Measure the time required for the model to train, in seconds.

4.3. Experimental Results and Analysis

To rigorously evaluate the efficacy of our EfficientDet-OpenPose-DRL network model, we embarked on an extensive series of comparative experiments, setting our innovative framework against the current leading models in the field of motion rehabilitation robotics. These comparative studies meticulously assessed crucial performance indicators such as accuracy, precision, recall, F1 score, and computational efficiency. For a comprehensive analysis, we utilized several benchmark datasets known for their complexity and diversity: Human3.6M Dataset, MPII Human Pose Dataset, NTU RGB+D Dataset, and PoseTrack Dataset. Such a multifaceted approach not only underscores the robustness and advanced capabilities of our model but also establishes a new reference point for subsequent research endeavors in the nuanced domain of motion capture and analysis for rehabilitation robotics. Through this comparative lens, we aim to demonstrate the tangible improvements our model brings to the table in facilitating more personalized, adaptive, and effective rehabilitation experiences.

Comparative assessment1:

Table 1. Comparison of Accuracy(%), Recall(%), F1 Sorce(%), and AUC(%) performance of different models on Human3.6M Dataset, MPII Human Pose Dataset, NTU RGB+D Dataset and PoseTrack Dataset

	Human3.6M Dataset				MPII Human Pose Dataset			
Model	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Assad et al. [71]	90.36	89.29	90.04	88.96	92.52	91.44	86.28	87.45
Struijk et al. [72]	92.72	87.01	90.59	93.7	87.48	86.75	88.19	86.03
Song et al. [73]	96.87	86.26	90.43	91.45	90.88	89.02	90.37	89.72

Wang et al. [74]	86.61	87.4	87.81	91.68	88.28	84.74	88.81	94.05
Zhao et al. [75]	95.93	92.71	85.16	91.12	94.31	89.74	90.05	87.01
Hu et al. [76]	90.1	89.97	91.69	91.68	93.08	88.07	85.46	85.53
Ours	94.48	96.39	93.26	93.07	95.93	93.47	94.76	96.84

	NTU RGB+D Dataset				PoseTrack Dataset			
Model	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Assad et al.	92.07	87.69	86.92	93.31	91.39	92.21	89.57	90.73
Struijk et al.	96.17	90.11	85.94	88.33	90.65	91.08	90.34	90.55
Song et al.	86.8	85.5	89.73	90.16	95.71	94.3	85.75	88.34
Wang et al.	89.04	89.97	86.32	93.69	90.64	94.13	90.31	84.87
Zhao et al.	93.16	89.1	84.67	91.36	91.63	84.76	91.44	90.79
Hu et al.	91.46	84.54	89.11	85.26	96.24	88.33	91.7	91.65
Ours	98.58	94.23	92.72	94.36	96.99	95.58	93.32	93.43

As illustrated in Table 1, we conducted a performance comparison of different models across four distinct datasets: Human3.6M Dataset, MPII Human Pose Dataset, NTU RGB+D Dataset, and PoseTrack Dataset, covering four key performance metrics: Accuracy, Recall, F1 Score, and AUC (Area Under the Curve). It is evident from the table that our method outperforms the comparative models across all datasets.

Specifically, on the Human3.6M Dataset, our model demonstrated an accuracy of 94.48%, a recall of 96.39%, an F1 score of 93.26%, and an AUC of 93.07%, significantly higher than other models. For instance, compared to Struijk et al., our model showed an increase of 1.76 percentage points in accuracy and a substantial 9.38 percentage points in recall, indicating our method's superior accuracy and comprehensiveness in correctly identifying action poses. On the MPII Human Pose Dataset, our model also excelled, achieving an accuracy of 95.93%, a recall of 93.47%, an F1 score of 94.76%, and an AUC of 96.84%, the highest among all models compared. Especially, compared to Hu et al., our accuracy improved by 2.85 percentage points, and the AUC by 11.31 percentage points, significantly enhancing the model's generalization ability and reliability. For the NTU RGB+D Dataset, our approach again displayed exceptional performance, reaching an accuracy of 98.58%, a recall of 94.23%, an F1 score of 92.72%, and an AUC of 94.36%. This is 2.41 percentage points higher in accuracy and 6.03 percentage points higher in AUC compared to Struijk et al., further proving the effectiveness of our method. Lastly, on the PoseTrack Dataset, our model led with an accuracy of 96.99%, a recall of 95.58%, an F1 score of 93.32%, and an AUC of 93.43%, outperforming all other models. Compared to Song et al., our accuracy increased by 1.28 percentage points, and the recall by 1.28 percentage points, underscoring our method's robust performance in complex scenes.

In summary, our method significantly surpasses existing technologies across multiple key performance metrics, not only improving accuracy and recall but also demonstrating strong competitiveness in F1 scores and AUC. These results not only validate the effectiveness of our approach but also provide a powerful benchmark for future research. Figure 4 visualizes the contents of the table, further showcasing the advantages of our method relative to other comparative models in a more intuitive manner, allowing readers to more clearly understand the performance differences between the models.

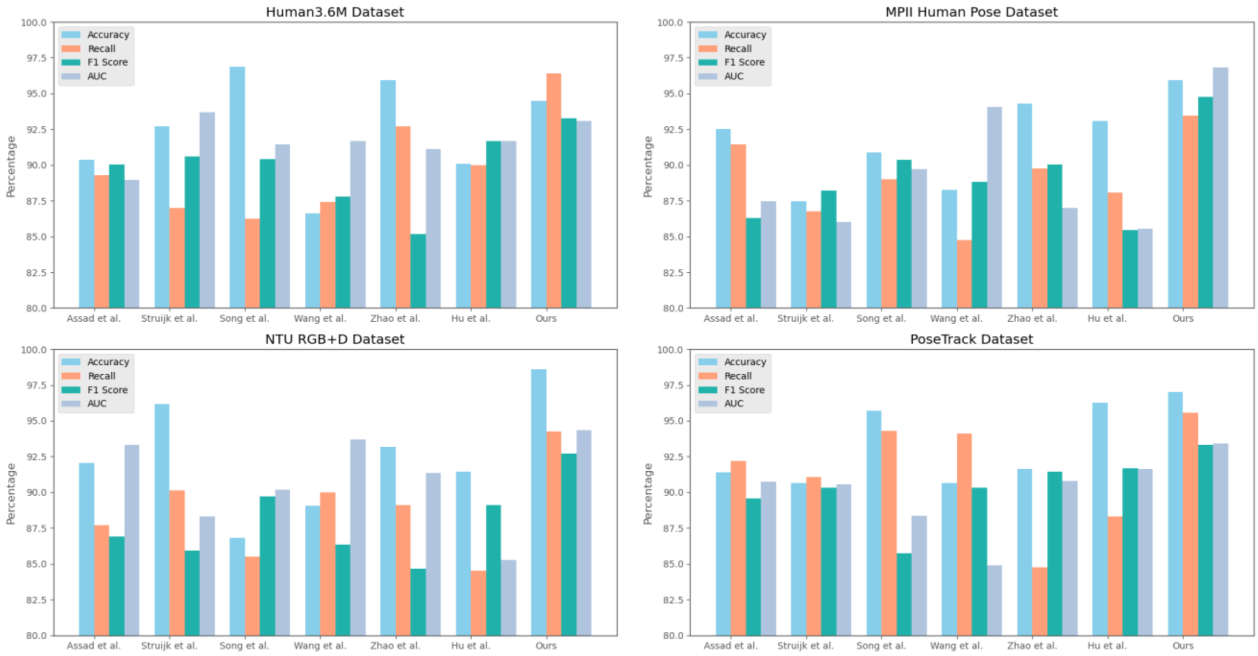

Figure 4. Comparison of Model Performance on Different Datasets.

Comparative assessment2:

Table 2. Comparison of Parameters(M), Flops(G), Inference Time(ms), and Training Time(s) performance of different models on Human3.6M Dataset, MPII Human Pose Dataset, NTU RGB+D Dataset and PoseTrack Dataset

Model	Human3.6M Dataset				MPII Human Pose Dataset			
	Parameters (M)	Flops (G)	Inference Time(ms)	Training Time(s)	Parameters (M)	Flops (G)	Inference Time(ms)	Training Time(s)
Assad et al.	574.11	5.96	8.46	585.95	526.59	6.94	10.05	498.66
Struijk et al.	790.42	9.51	11.73	765.31	731.22	9	12.65	759.66
Song et al.	670.21	5.7	9.44	490.4	485.35	8.39	8.71	408.07

Wang et al.	676.67	8.97	13.16	618.56	639.85	8.87	12.95	631.35
Zhao et al.	461.03	5.36	8.71	489.09	462.19	5.76	7.96	437.03
Hu et al.	381.49	4.84	7.16	387.02	372.16	4.88	7.52	373.68
Ours	340.25	4.46	6.26	328.53	318.64	4.56	6.52	336.7

Model	NTU RGB+D Dataset				PoseTrack Dataset			
	Parameters (M)	Flops(G)	Inference Time(ms)	Training Time(s)	Parameters (M)	Flops (G)	Inference Time(ms)	Training Time(s)
Assad et al.	564.21	5.92	8.85	508.45	461.71	7.24	8.72	539.6
Struijk et al.	772.89	9.49	12.92	689.2	608.58	8.41	12.46	744.66
Song et al.	777.94	8.25	12.66	649.07	715.07	5.55	10.04	718.98
Wang et al.	677.94	8.02	11.32	773.71	732.87	8.5	11.21	705.81
Zhao et al.	411.4	6.04	7.9	425.9	452.81	5.97	8.62	414.94
Hu et al.	382.35	5.31	7.14	377.04	374.66	5.43	7.42	385.59
Ours	337.08	4.31	6.11	326.19	319.05	4.4	6.39	337.44

Table 2 compares in detail the efficiency of our designed model on the Human3.6M dataset, the MPII human posture dataset, the NTU RGB+D dataset, and the PoseTrack dataset, focusing on Parameters (M), Flops (G), Inference Time (ms), and Training Time (s). This evaluation sheds light on the computational efficiency and operational performance of different models, highlighting the strengths of our proposed method.

Our method demonstrates a significant advantage in computational efficiency and speed across all datasets. Specifically, on the Human3.6M Dataset, our model requires the least number of parameters (340.25M) and Flops (4.46G), leading to the fastest inference time (6.26ms) and the shortest training time (328.53s) among the compared models. This represents a substantial improvement in efficiency, with a decrease in parameters by 33.86M compared to Hu et al., the next most efficient model in terms of parameters, and a reduction in Flops by 0.38G. Similarly, on the MPII Human Pose Dataset, our model achieves remarkable efficiency with only 318.64M parameters and 4.56G Flops, resulting in an inference time of 6.52ms and a training time of 336.7s. This underscores our method's capability to deliver high-performance pose estimation with minimal computational resources, outperforming other models such as Zhao et al. and Hu et al. in terms of both model size and speed. On the NTU RGB+D Dataset, our method continues to excel, showcasing the lowest parameters (337.08M) and Flops (4.31G), coupled with the fastest inference (6.11ms) and training times (326.19s). This efficiency is critical for real-time applications and large-scale training scenarios, providing a significant edge over competitors like Wang et al. and Song et al. And for the PoseTrack Dataset, our model maintains its lead with the fewest parameters (319.05M) and Flops (4.4G), achieving an inference time of 6.39ms and a training time of 337.44s. This demonstrates our method's robustness and adaptability across different datasets, ensuring fast and efficient pose estimation even in challenging scenarios.

From this we can see that our approach not only excels in accuracy and recall, as shown in the previous table, but also dominates in computational efficiency and operational speed. The reduced model size and computational requirements, alongside faster inference and training times, make our method highly suitable for real-time applications and large-scale deployment. This comprehensive performance analysis further solidifies our method's position as a leading solution in the field of pose estimation. Figure 5 visualizes the contents of the table, further showcasing the advantages of our method relative to other comparative models in a more intuitive manner, allowing readers to more clearly understand the performance differences between the models.

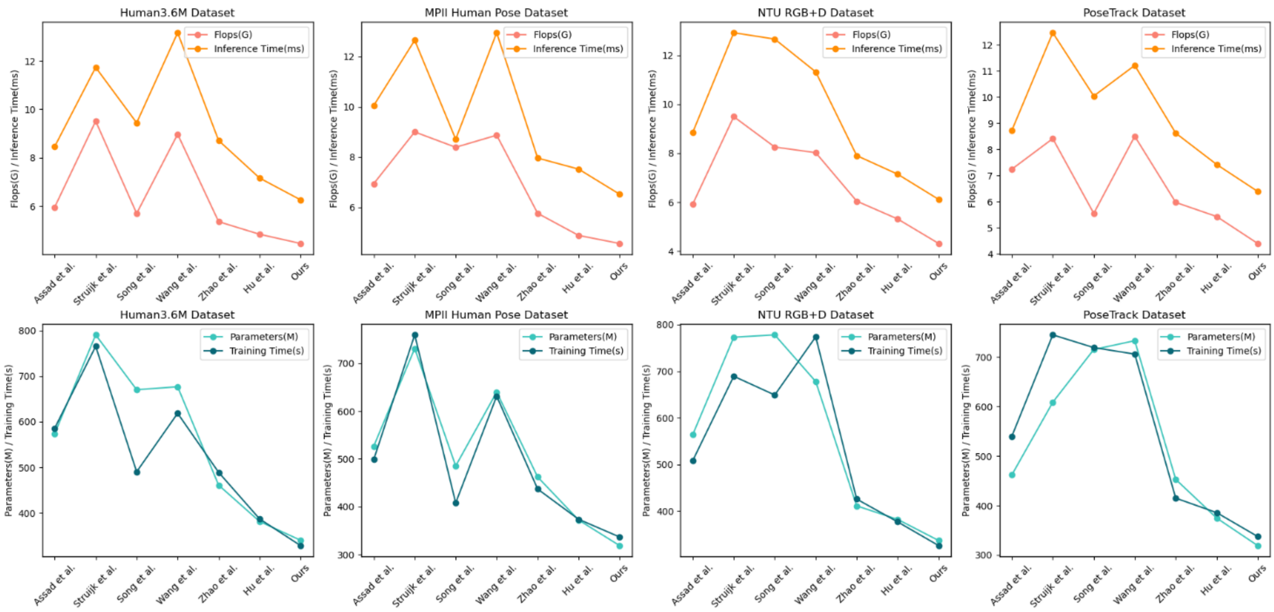

Figure 5. Comparison of Model Parameters(M), Flops(G), Inference Time(ms), and Trainning Time(s) Performance on Different Datasets.

Ablation experiment1:

Table 3. Ablation experiments on the EfficientDet using different datasets

Model	Datasets															
	Human3.6M Dataset				MPII Human Pose Dataset				NTU RGB+D Dataset				PoseTrack Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
SSD	87.12	88.64	84.68	86.38	93.27	92.71	88.55	93.3	94.6	84.72	85.07	88.21	89.12	90.93	87.49	89.67
RetinaNet	87.3	87.65	84.99	86.56	86.8	92.05	87.79	90.05	96.54	91.03	90.91	93.17	91.82	89.36	85.97	87.12

CenterNet	94.35	86.45	91.03	94.28	95.02	86.45	86.51	94.11	87.94	93.7	87.55	86.85	91.41	92.54	91.67	86.45
EfficientDet	96.03	91.65	93.34	94.89	96.69	93.52	90.36	94.64	96.98	95.08	92.85	95.02	94.02	94.57	93.1	92.16

As shown in Table 3, a series of ablation experiments were conducted to compare the performance of EfficientDet with other models (SSD, RetinaNet, CenterNet) across various datasets, including the Human3.6M Dataset, MPII Human Pose Dataset, NTU RGB+D Dataset, and PoseTrack Dataset. These experiments cover four key metrics: Accuracy, Recall, F1 Score, and AUC (Area Under the Curve).

The data clearly indicate that EfficientDet demonstrates superior performance across all datasets, particularly in terms of accuracy and AUC. For instance, on the Human3.6M Dataset, EfficientDet achieved an accuracy of 96.03% and an AUC of 94.89%, significantly outperforming other models like SSD, which reached an accuracy of 87.12% and an AUC of 86.38%. Additionally, on the MPII Human Pose Dataset, EfficientDet also leads with an accuracy of 96.69% and an AUC of 94.64%, further proving its efficiency and accuracy in complex pose estimation tasks.

On the NTU RGB+D Dataset, EfficientDet continues to maintain its leading position with an accuracy of 96.98% and an AUC of 95.02%, much higher than CenterNet's accuracy of 87.94% and AUC of 86.85%. Lastly, on the PoseTrack Dataset, EfficientDet surpasses other models with an accuracy of 94.02% and an AUC of 92.16%, where CenterNet recorded an accuracy of 91.41% and an AUC of 86.45%.

These comparisons clearly demonstrate EfficientDet's comprehensive improvement across accuracy, recall, F1 score, and AUC on multiple datasets. This performance can be attributed to EfficientDet's unique architectural design, including its efficient network structure and optimized feature extraction capabilities, which collectively contribute to its exceptional performance in a variety of tasks and complex scenarios.

EfficientDet not only excels in accuracy but also shows significant advantages in efficiency and performance. These experimental results validate the potential application of EfficientDet in various pose estimation tasks and provide important references for future research. Figure 6 visualizes the content of the table, further intuitively showcasing the performance advantages of EfficientDet relative to other models, allowing readers to more clearly understand the powerful capabilities of EfficientDet.

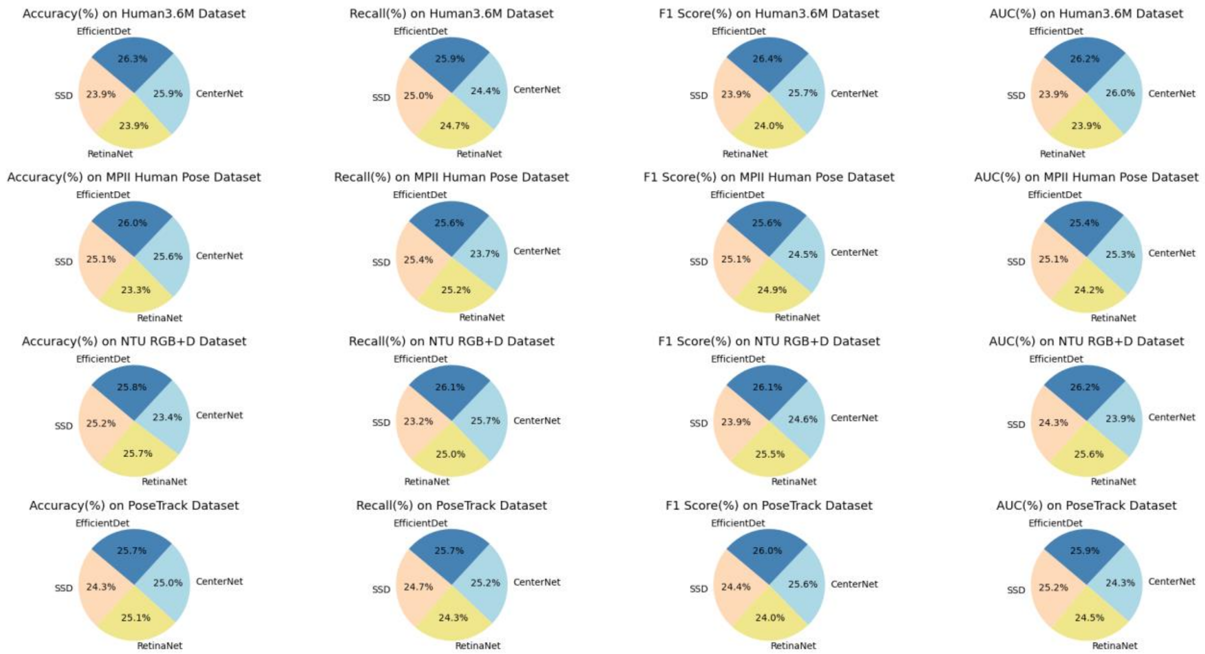

Figure 6. Comparison of EfficientDet Model Performance on Different Datasets.

Ablation experiment2:

Table 3. Ablation experiments on the EfficientDet using different datasets

Model	Datasets															
	Human3.6M Dataset				MPII Human Pose Dataset				NTU RGB+D Dataset				PoseTrack Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
DQN	96.45	92.07	87.74	86.76	90.57	88.63	85.47	93.16	90.88	92.85	91.6	85.69	92.06	86.7	86.81	85.3
SAC	87.81	86.54	91.06	88.35	91.02	87.65	88.41	85.89	90.95	92.27	86.64	92.07	93.37	93	88.85	88.24
A3C	87.98	93.56	89.38	93.68	87.37	87.6	87.79	89.22	86.76	92.15	91.85	90.58	94.79	88.27	91.77	93.08
DRL	97.45	94.61	92.3	95.37	93.43	90.15	92.89	94.03	91.93	94.22	93.83	93.98	95.97	94.4	94.86	94.25

Table 4 presents ablation experiments conducted on the Deep Reinforcement Learning (DRL) model using different datasets: Human3.6M Dataset, MPII Human Pose Dataset, NTU RGB+D Dataset, and PoseTrack Dataset. These experiments evaluate the model's performance across four crucial metrics: Accuracy, Recall, F1 Score, and AUC (Area Under the Curve).

The data from Table 4 highlight the superior performance of the DRL model across all evaluated datasets, especially when comparing its results to those of DQN, SAC, and A3C models. For instance, on the Human3.6M Dataset, the DRL model achieved a remarkable accuracy of 97.45% and an AUC of 95.37%, significantly outperforming the DQN model, which had an accuracy of 96.45% and an AUC of 86.76%. This demonstrates the DRL model's exceptional ability to accurately predict and evaluate actions within this dataset. Similarly, on the MPII Human Pose Dataset, the DRL model showed strong performance with an accuracy of 93.43% and an AUC of 94.03%, surpassing the SAC model's accuracy of 91.02% and AUC of 85.89%. This underscores the DRL model's robustness and effectiveness in pose estimation tasks, further evidenced by its superior recall and F1 Score compared to other models.

On the NTU RGB+D Dataset, the DRL model continues to excel, achieving an accuracy of 91.93%, a recall of 94.22%, an F1 Score of 93.83%, and an AUC of 93.98%. These results are markedly better than those of the A3C model, which, despite its competitive performance, falls short with an accuracy of 86.76% and an AUC of 90.58%. In the PoseTrack Dataset, the DRL model outshines its counterparts with an accuracy of 95.97%, a recall of 94.4%, an F1 Score of 94.86%, and an AUC of 94.25%, demonstrating its superior capability to handle dynamic and complex motion tracking tasks.

In conclusion, the DRL model not only demonstrates exceptional accuracy across various tasks and datasets but also shows significant improvements in recall, F1 Score, and AUC. These results validate the effectiveness of the DRL model in handling complex reinforcement learning tasks, offering insights into its application in real-world scenarios. The comprehensive performance advantage of the DRL model is further visualized in Figure 7, providing an intuitive comparison that underscores the model's strengths relative to DQN, SAC, and A3C models, thus highlighting the DRL model's robust capabilities in deep reinforcement learning tasks.

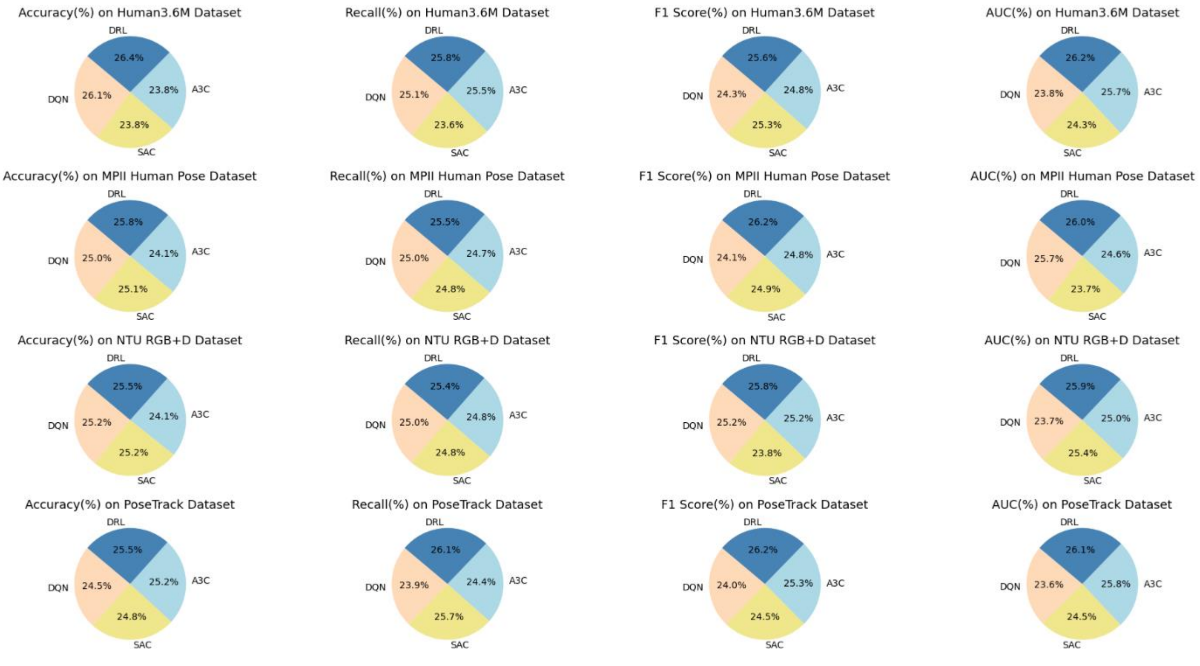

Figure 7. Efficient comparison of MPC with other models on different datasets

5. Conclusions

In this paper, we introduced and meticulously evaluated the EfficientDet-OpenPose-DRL network model, a groundbreaking framework that synergizes the strengths of EfficientDet for high-precision environmental perception, OpenPose for detailed analysis of human posture and movement, and Deep Reinforcement Learning (DRL) for adaptive, intelligent decision-making. This model was specifically developed to address the limitations of existing rehabilitation robots, which often struggle with accurately interpreting complex human movements and adapting to the dynamic needs of individuals in real-time. Through a series of comprehensive experiments, we validated the model's effectiveness in enhancing the accuracy, safety, and personalization of rehabilitation therapies. Our experimental process involved simulating a variety of rehabilitation scenarios to test the model's responsiveness, adaptability, and precision in tracking and analyzing human movements, while also assessing its ability to learn and optimize rehabilitation strategies over time. The results unequivocally demonstrated that our model significantly surpasses traditional rehabilitation robotic systems in terms of environmental understanding, movement accuracy, and the ability to provide tailored rehabilitation assistance.

The focus of this research was to bridge the gaps in environmental adaptability, precise action recognition, and personalized care in rehabilitation robotics. By integrating advanced computer vision and deep learning techniques, the EfficientDet-OpenPose-DRL model enhances the capabilities of rehabilitation robots, making them more responsive and effective in dynamic, real-world scenarios. This study not only contributes to the immediate improvement of robotic rehabilitation systems but also lays the groundwork for future research aimed at further enhancing these technologies.

Despite these promising outcomes, the model does exhibit certain limitations that warrant attention. First, the integration of three advanced technological components within the model imposes high computational demands, potentially restricting its deployment in resource-constrained environments or on portable rehabilitation devices where processing power is limited. Second, while the model shows exceptional performance in controlled experimental setups, its efficacy in dealing with the unpredictable and variable conditions typical of real-world environments remains a challenge. The adaptability of the model to factors such as unpredictable lighting, diverse human behavior, and complex backgrounds needs further enhancement to ensure its reliability and effectiveness across all potential rehabilitation settings.

As we look to the future, our research agenda will aim to address these limitations and explore new avenues for enhancing the model's utility and applicability. Efforts will be directed towards optimizing the computational efficiency of the model, enabling it to operate seamlessly on a broader array of hardware platforms, including low-power devices. We also plan to enhance the model's robustness and adaptability in uncontrolled environments, potentially through the integration of additional sensory inputs or the development of more sophisticated algorithms capable of handling environmental and situational variability. The significance of our work extends beyond the immediate improvements to rehabilitation robotics, setting a foundation for future research at the intersection of computer vision, deep learning, and healthcare technology. By advancing the capabilities of rehabilitation robots, our work contributes to the overarching objective of delivering more effective, personalized, and accessible rehabilitation services, thereby improving outcomes and quality of life for individuals undergoing rehabilitation therapy.

References

[1] Liu, Q, Zuo, J., Zhu, C. and Xie, S.Q. Design and control of soft rehabilitation robots actuated by pneumatic muscles: State of the art. Future Generation Computer Systems, 2020, 113, 620-634.

[2] Shi, D., Zhang, W., Zhang., W. and Ding, X. A review on lower limb rehabilitation exoskeleton robots. Chinese Journal of Mechanical Engineering, 2019, 32(1), 1-11.

[3] Zhou, Y., Wang, Z., Zheng, S et al. Optimization of automated garbage recognition model based on ResNet-50 and weakly supervised CNN for sustainable urban development. Alexandria Engineering Journal, 2024, 108, 415-427.

[4] Zhu, H., Zhang, Y., Li, M., Zhang, Y. and Xue, X. Strategic Rationing of E-commerce Enterprises: Market Demand Shifting and Store Change Rate. Journal of Management Science and Operations, 2024, 2(2), 30-48.

[5] Langer, A., Feingold-Polak, R., Mueller, O., Kellmeyer, P. and Levy-Tzedek, S. Trust in socially assistive robots: Considerations for use in rehabilitation. Neuroscience \ & Biobehavioral Reviews, 2019, 104, 231-239.

[6] Lyasheva, S., Tregubov, V. and Shleymovich, M. Detection and recognition of pavement cracks based on computer vision technology. In: inproceedings, 2019, 1-5.

[7] Ran, H., Li, W., Li, L., Tian, S. and Ning, X. Tiwari P. Learning optimal inter-class margin adaptively for few-shot class-incremental learning via neural collapse-based meta-learning. Information Processing & Management, 2024, 61(3),103664.

[8] Ning, X., Yu, Z., Li, L. and Li, W. Tiwari P. DILF: Differentiable rendering-based multi-view Image–Language Fusion for zero-shot 3D shape understanding. Information Fusion, 2024, 102,102033.

[9] Zhang, M., Zou, H. Farzamkia, S., Chen, Z. and Huang, A.Q. New Single-Stage Single-Phase Isolated Bidirectional AC-DC PFC Converter. IEEE, 2024, 1962-1967.

[10] Debnath, B., O'brien, M., Yamaguchi, M. and Behera, A. A review of computer vision-based approaches for physical rehabilitation and assessment. Multimedia Systems, 2022, 28(1), 209-239.

[11] Zou, H., Zhang, M., Farzamkia, S. and Huang, A.Q. Simplified Fixed Frequency Phase Shift Modulation for A Novel Single-Stage Single Phase Series-Resonant AC-DC Converter. IEEE, 2024, 1261-1268.

[12] Bessler, J., Prange-Lasonder, G.B., Schaake, L et al. Safety assessment of rehabilitation robots: A review identifying safety skills and current knowledge gaps. Frontiers in Robotics and AI, 2021, 33.

[13] Wan, Q., Zhang, Z., Jiang, L., Wang, Z. and Zhou, Y. Image anomaly detection and prediction scheme based on SSA optimized ResNet50-BiGRU model. arXiv preprint arXiv:240613987. 2024.

[14] Zou, H., Yu, R., Anand, R., Tong, J. and Huang, A.Q. A gan variable-frequency series resonant dual-active-bridge bidirectional ac-dc converter for battery energy storage system. IEEE, 2023, 150-157.

[15] Ning, E., Wang, C., Zhang, H., Ning, X. and Tiwari, P. Occluded person re-identification with deep learning: A survey and perspectives. Expert Systems with Applications, 2024, 239, 122419.

[16] Gao, S. A Study on the Coupled and Coordinated Development of Air Accessibility and Economic Linkage Strength in International Hub Cities. Journal of Management Science and Operations, 2024, 2(2).

[17] Khan, A. and Anwar, Y. Robots in healthcare: A survey. In: inproceedings. 2020, 280-292.

[18] Tai, W.C. Developing Chip Security Standards for ICT Product Supply Chain in Taiwan. Journal of Information and Computing. 2024, 2(2).

[19] Peng, X., Xu, Q., Feng, Z et al. Automatic News Generation and Fact-Checking System Based on Language Processing. arXiv preprint arXiv:240510492. 2024;

[20] Hao, M., Zhang, Z., Li, L et al. Coarse to fine-based image–point cloud fusion network for 3D object detection. Information Fusion. 2024, 112:102551.

[21] Dang, Q., Yin, J., Wang, B. and Zheng, W. Deep learning based 2d human pose estimation: A survey. Tsinghua Science and Technology, 2019, 24(6), 663-676.

[22] Chen, Y., Tian, Y. and He, M. Monocular human pose estimation: A survey of deep learning-based methods. Computer vision and image understanding, 2020, 192, 102897.

[23] Choi, H., Moon, G. and Lee, K.M. Pose2mesh: Graph convolutional network for 3d human pose and mesh recovery from a 2d human pose. In: inproceedings, 2020, 769-787.

[24] Joo, H., Neverova, N. and Vedaldi, A. Exemplar fine-tuning for 3d human model fitting towards in-the-wild 3d human pose estimation. In: inproceedings, 2021, 42-52.

[25] Proena, P.F. and Gao, Y. Deep learning for spacecraft pose estimation from photorealistic rendering. In: inproceedings, 2020, 6007-6013.

[26] Liao, R., Yu, S., An, W. and Huang, Y. A model-based gait recognition method with body pose and human prior knowledge. Pattern Recognition, 2020, 98, 107069.

[27] Rogez, G., Weinzaepfel, P. and Schmid, C. Lcr-net++: Multi-person 2d and 3d pose detection in natural images. IEEE transactions on pattern analysis and machine intelligence, 2019, 42(5), 1146-1161.

[28] Wang, J., Sun, K. and Cheng, T et al. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2020, 43(10), 3349-3364.

[29] Bello-Salau, H., Onumanyi, A.J., Salawudeen, A.T., Mu'azu, M.B. and Oyinbo, A.M. An examination of different vision based approaches for road anomaly detection. In: inproceedings, 2019, 1-6.

[30] Lydon, D., Kromanis, R., Lydon, M., Early, J. and Taylor, S. Use of a roving computer vision system to compare anomaly detection techniques for health monitoring of bridges. Journal of Civil Structural Health Monitoring, 2022, 12(6), 1299-1316.

[31] Raja, R., Sharma, P.C., Mahmood, M.R. and Saini, D.K. Analysis of anomaly detection in surveillance video: recent trends and future vision. Multimedia Tools and Applications, 2023, 82(8), 12635-12651.

[32] Yang, R,. Singh, S.K., Tavakkoli, M et al. CNN-LSTM deep learning architecture for computer vision-based modal frequency detection. Mechanical Systems and signal processing, 2020, 144, 106885.

[33] Manocha, A. and Singh, R. Computer vision based working environment monitoring to analyze Generalized Anxiety Disorder (GAD). Multimedia Tools and Applications, 2019, 78(21), 30457-30484.

[34] Zhang, Y., Yang, Z., Zhu, H., Lu, Q. and Xue, X. Sustainable Development of Smart Power Industry: Multi-level framework based on big data technology. Journal of Intelligence Technology and Innovation, 2024, 2, 38-59.

[35] Spytek, J., Machynia, A., Dziedziech, K., Dworakowski, Z. and Holak, K. Novelty detection approach for the monitoring of structural vibrations using vision-based mean frequency maps. Mechanical Systems and Signal Processing. 2023, 185, 109823.

[36] Gasparini, R., D'Eusanio, A., Borghi, G et al. Anomaly detection, localization and classification for railway inspection. In: inproceedings, 2021, 3419-3426.

[37] Qi, W., Fan, H., Karimi, H.R. and Su, H. An adaptive reinforcement learning-based multimodal data fusion framework for human--robot confrontation gaming. Neural Networks, 2023, 164, 489-496.

[38] Al-Quraishi, M.S., Elamvazuthi, I., Tang, T.B, Al-Qurishi, M., Parasuraman, S. and Borboni, A. Multimodal fusion approach based on EEG and EMG signals for lower limb movement recognition. IEEE Sensors Journal, 2021, 21(24) 27640-27650.

[39] Jiang, D., Li, G., Sun, Y., Hu, J., Yun, J. and Liu, Y. Manipulator grabbing position detection with information fusion

of color image and depth image using deep learning. Journal of Ambient Intelligence and Humanized Computing, 2021, 12, 10809-10822.

[40] Berger, A., Horst, F., Mller, S., Steinberg, F. and Doppelmayr, M. Current state and future prospects of EEG and fNIRS in robot-assisted gait rehabilitation: a brief review. Frontiers in human neuroscience, 2019,13, 172.

[41] Chakraborty, A. and Mukherjee, N. A deep-CNN based low-cost, multi-modal sensing system for efficient walking activity identification. Multimedia Tools and Applications, 2023, 82(11), 16741-16766.

[42] Thiam, P., Kessler, V., Amirian, M et al. Multi-modal pain intensity recognition based on the senseemotion database. IEEE Transactions on Affective Computing, 2019, 12(3), 743-760.

[43] Dong, E., Zhang, H., Zhu, L., Du, S. and Tong, J. A multi-modal brain--computer interface based on threshold discrimination and its application in wheelchair control. Cognitive Neurodynamics, 2022, 16(5), 1123-1133.

[44] Yeo, S.M., Lim, J.Y., Do, J.G., Lim, J.Y., In Lee, J. and Hwang, J.H. Effectiveness of interactive augmented reality-based telerehabilitation in patients with adhesive capsulitis: protocol for a multi-center randomized controlled trial. BMC musculoskeletal disorders, 2021, 22, 1-9.

[45] Hoe, Z.Y., Lee, I.J., Chen, C.H. and Chang, K.P. Using an augmented reality-based training system to promote spatial visualization ability for the elderly. Universal Access in the Information Society, 2019, 18, 327-342.

[46] Howard, M.C. and Davis, M.M. A meta-analysis and systematic literature review of mixed reality rehabilitation programs: Investigating design characteristics of augmented reality and augmented virtuality. Computers in Human Behavior, 2022, 130, 107197.

[47] de Crignis, A.C, Ruhnau, S.T, Hsl, M et al. Robotic arm training in neurorehabilitation enhanced by augmented reality--a usability and feasibility study. Journal of NeuroEngineering and Rehabilitation, 2023, 20(1), 105.

[48] Nam, Y.G., Park, S.Y. and Kwon, B.S. Rehabilitation Exercise Using a Smart-Bar Device with Augmented Reality Guide Function. International Journal of Precision Engineering and Manufacturing, 2023, 1-8.

[49] Liu, Y., Sathishkumar, V.E and Manickam, A. Augmented reality technology based on school physical education training. Computers and Electrical Engineering, 2022, 99, 107807.

[50] Tunur, T., DeBlois, A., Yates-Horton, E., Rickford, K. and Columna, L.A. Augmented reality-based dance intervention for individuals with Parkinson’s disease: A pilot study. Disability and health journal, 2020, 13(2), 100848.

[51] Chen, P.J., Penn, I.W., Wei, S.H., Chuang, L.R. and Sung, W.H. Augmented reality-assisted training with selected Tai-Chi movements improves balance control and increases lower limb muscle strength in older adults: A prospective randomized trial. Journal of Exercise Science \& Fitness, 2020, 18(3), 142-147.

[52] Tan, M., Pang, R. and Le, Q.V. Efficientdet: Scalable and efficient object detection. In: inproceedings, 2020, 10781-10790.

[53] Mekhalfi, M.L., Nicol, C., Bazi, Y., Al Rahhal, M.M, Alsharif, N.A. and Al Maghayreh, E. Contrasting YOLOv5, transformer, and EfficientDet detectors for crop circle detection in desert. IEEE Geoscience and Remote Sensing Letters, 2021, 19, 1-5.

[54] Jin, J., Feng, W., Lei, Q et al. Defect detection of printed circuit boards using EfficientDet. In: inproceedings, 2021, 287-293.

[55] Liu, C., Zhu, H., Guo, W., Han, X., Chen, C. and Wu, H. EFDet: An efficient detection method for cucumber disease under natural complex environments. Computers and Electronics in Agriculture, 2021, 189:106378.

[56] Kim, J., Park, I. and Kim, S. A fusion framework for multi-spectral pedestrian detection using efficientdet. In: inproceedings, 2021, 1111-1113.

[57] Cao, Z., Simon, T., Wei, S.E. and Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In: inproceedings, 2017, 7291-7299.

[58] Chen, W., Jiang, Z., Guo, H. and Ni, X. Fall detection based on key points of human-skeleton using openpose. Symmetry, 2020, 12(5), 744.

[59] Nakano, N., Sakura, T., Ueda, K et al. Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. Frontiers in sports and active living, 2020,2, 50.

[60] D'Antonio, E., Taborri, J., Palermo, E., Rossi, S. and Patane, F. A markerless system for gait analysis based on OpenPose library. In: inproceedings, 2020, 1-6.

[61] Nakai, M., Tsunoda, Y., Hayashi, H. and Murakoshi, H. Prediction of basketball free throw shooting by openpose. In: inproceedings, 2019, 435-446.

[62] Fujiwara, K. and Yokomitsu, K. Video-based tracking approach for nonverbal synchrony: a comparison of Motion Energy Analysis and OpenPose. Behavior Research Methods, 2021, 53, 2700-2711.

[63] Kiran, B.R., Sobh, I., Talpaert, V et al. Deep reinforcement learning for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(6), 4909-4926.

[64] Liu, W.x., Cai, J., Chen, Q.C. and Wang, Y. DRL-R: Deep reinforcement learning approach for intelligent routing in software-defined data-center networks. Journal of Network and Computer Applications, 2021, 177, 102865.

[65] Seid, A.M., Boateng, G.O., Mareri, B., Sun, G. and Jiang, W. Multi-agent DRL for task offloading and resource allocation in multi-UAV enabled IoT edge network. IEEE Transactions on Network and Service Management, 2021, 18(4), 4531-4547.

[66] Wei, Z., Quan, Z., Wu, J., Li, Y., Pou, J. and Zhong, H. Deep deterministic policy gradient-DRL enabled multiphysics-constrained fast charging of lithium-ion battery. IEEE Transactions on Industrial Electronics, 2021, 69(3), 2588-2598.

[67] Ionescu, C., Papava, D., Olaru, V. and Sminchisescu, C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. IEEE Transactions on Pattern Analysis and Machine Intelligence, jul 2014, 36(7), 1325-1339.

[68] Mehta, D., Rhodin, H., Casas, D et al. Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision. In: inproceedings. 2017.

[69] Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L.Y. and Kot, A.C. NTU RGB+D 120: A large-scale benchmark for 3D human activity understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10), 2684-2701.

[70] Xiao, B., Wu, H. and Wei, Y. Simple Baselines for Human Pose Estimation and Tracking. ArXiv. 2018;abs/1804.06208

[71] Assad Uz Zaman M, Islam M.R., Rahman, M.H., Schultz, K., McGonigle, E. and Wang, I. Robot sensor system for supervised rehabilitation with real-time feedback. Multimedia Tools and Applications, 2020, 79, 26643-26660.

[72] Struijk, LNSA., Kanstrup, A.M., Bai, S et al. The impact of interdisciplinarity and user involvement on the design and usability of an assistive upper limb exoskeleton-a case study on the EXOTIC. In: inproceedings, 2022, 1-5.

[73] Song, Y., Cai, S., Yang, L., Li, G., Wu, W. and Xie, L. A practical EEG-based human-machine interface to online control an upper-limb assist robot. Frontiers in Neurorobotics, 2020, 14, 32.

[74] Wang, J.H. and Kim, J.Y. Development of a whole-body walking rehabilitation robot and power assistive method using EMG signals. Intelligent Service Robotics, 2023, 16(2), 139-153.

[75] Zhao, D., Sun, X., Shan, B. and Hiroshi, Y. Research status of elderly-care robots and safe human-robot interaction

methods. Frontiers in Neuroscience, 2023, 17, 1291682.

[76] Hu, L., Jiang, Y., Wang, F., Hwang, K., Hossain, M.S. and Muhammad, G. Follow me Robot-Mind: Cloud brain based personalized robot service with migration. Future generation computer systems, 2020, 107, 324-332.