applied sciences

MDPI

Article

Biomac3D: 2D-to-3D Human Pose Analysis Model for Tele-Rehabilitation Based on Pareto Optimized Deep-Learning Architecture

Rytis Maskeliūnas [1,*], Audrius Kulikajevas [1], Robertas Damaševičius [1], Julius Griškevičius [2] and Aušra Adomavičienė [3]

[1] Department of Multimedia Engineering, Faculty of Informatics, Kaunas University of Technology, 51368 Kaunas, Lithuania
[2] Department of Biomechanical Engineering, Vilnius Gediminas Technical University, 03224 Vilnius, Lithuania
[3] Department of Rehabilitation, Physical and Sports Medicine, Faculty of Medicine, Vilnius University, 08661 Vilnius, Lithuania
* Correspondence: rytis.maskeliunas@ktu.lt

Abstract: The research introduces a unique deep-learning-based technique for remote rehabilitative analysis of image-captured human movements and postures. We present a ploninomial Pareto-optimized deep-learning architecture for processing inverse kinematics for sorting out and rearranging human skeleton joints generated by RGB-based two-dimensional (2D) skeleton recognition algorithms, with the goal of producing a full 3D model as a final result. The suggested method extracts the entire humanoid character motion curve, which is then connected to a three-dimensional (3D) mesh for real-time preview. Our method maintains high joint mapping accuracy with smooth motion frames while ensuring anthropometric regularity, producing a mean average precision (mAP) of 0.950 for the task of predicting the joint position of a single subject. Furthermore, the suggested system, trained on the MoVi dataset, enables a seamless evaluation of posture in a 3D environment, allowing participants to be examined from numerous perspectives using a single recorded camera feed. The results of evaluation on our own self-collected dataset of human posture videos and cross-validation on the benchmark MPII and KIMORE datasets are presented.

Keywords: Pareto optimization; 2D to 3D; human posture analysis; remote rehabilitation; telehealth

check for updates

Citation: Maskeliūnas, R.; Kulikajevas, A.; Damaševičius, R.; Grixsxkevixcxius, J.; Adomavixcxienė, A. Biomac3D: 2D-to-3D Human Pose Analysis Model for Tele-Rehabilitation Based on Pareto Optimized Deep-Learning Architecture. *Appl. Sci.* **2023**, *13*, 1116. https://doi.org/10.3390/app13021116

Academic Editor: Bruno Chrcanovic

Received: 15 December 2022
Revised: 6 January 2023
Accepted: 10 January 2023
Published: 13 January 2023

1. Introduction

Multiple studies have been conducted to investigate the feasibility and effectiveness of new information-technology tools and their design to facilitate home rehabilitation after stroke or trauma [1–4], a successful recovery that could potentially lead to a positive change in attitudes [5]. Researchers have analyzed the outcomes of computer-assisted therapy [6] or virtual reality (VR) [7,8] in rehabilitation and the effectiveness in the recovery of upper limb motor functions, maintenance of balance of posture and gait, lower limbs, posture, and walking [9]. Additionally, researchers have meticulously examined the clinical effects of tele-rehabilitation, which allows patients to perform therapy with therapists using telecommunication devices in the home environment and has been extensively used for motor and cognitive recovery [10]. This was considered to be one of the most effective approaches to diagnose musculoskeletal issues and rehabilitate patients recovering from numerous impairments through physical-therapy intervention through exercises in specific activities. Participating in physiotherapy and rehabilitation programs is often required and critical in postoperative recovery or in the treatment of a wide range of health problems [11]. However, providing patients with access to a doctor for every rehab session is both impossible and fiscally unjustifiable. As a result, existing health services around the world are structured in such a way that a first portion of rehabilitation programs is performed in an actual hospital under the direct supervision of a clinician, followed by a subsequent portion in which patients often perform a sequence of exercises given at home [12]. Therefore, patients have often carried out such exercises at home, without the presence of specialists or therapists. As a result, patients are unable to receive proper supervision and evaluation of the required activity. On the contrary, home technologies driven by artificial intelligence (AI) have the advantage of allowing flexibility in location and time in rehabilitation therapy, as well as receiving feedback from therapists remotely [13].

Biomechanical analysis of human movements has become an essential tool for fundamental research and therapeutic care of orthopedic and neurological problems [14]. Offline clinical movement analysis has historically been performed by processing collected raw motion and force data [15]. The laboratory or gait report is then sent to the doctor, who then determines treatment options, often including time-series data of biomechanical variables such as joint moments (kinetics) or joint angles (kinematics). In contrast to a standard report created after post-processing, an automated computer-based biomechanical analysis would provide new opportunities for the patient and the kinesitherapist to interact in real time with biomechanical records during patient monitoring or treatment [16].

A meaningful visualization and quantification [17,18] of certain motion factors might be beneficial to clinicians and physical therapists. In addition, kinematics of movement provide limited information about the performance of the movement; however, it can serve as input into numerical biomechanical musculo-skeletal models such as OpenSim [19], AnyBody [20] or Biomechanics of Bodies [21], which allow calculation of kinetic parameters of motion, such as muscle forces and joint torques. Doctors would gain new insights into internal forces and moments, kinematic parameters (ROM, movement speed, acceleration), motion accuracy, muscle force and strength parameters, changes in vital functions (pulse, arterial pressure, glucose level, heart rate) and physical load tolerance and recovery indicators would otherwise be essentially unseen [22]. Moreover, such biomechanical data can be communicated to the patient in real time to help them perform physiotherapy exercises more adequately than following the verbal or tactile input of a kinesitherapist [23].

Full-body pose and motion analysis has become a part of many modern medical solutions, especially in digitized home rehabilitation and telerehabilitation scenarios [24,25]. These often target gait detection and motion estimation [26], as well as physical training monitoring [27]. Limb analysis [28] has also become a common feature [29], leading to custom applications for feedback training employing specific variables, such as a single joint moment or angle. To make such computation possible, approximations that ignore certain mechanical factors, such as inertial components in equations of motion, are frequently utilized, often applying some deep-learning model for compensation. Applications range from computer-vision-based cerebral infarction rehabilitation [30], head neck rehabilitation [31], spinal-cord injury [32], stroke rehabilitation [33], avoidance or alleviation of fracture pains [34], port medicine [35] to robotic-induced systems, where skeleton tracking-based posture assessment can be used for real-time monitoring of upper limb rehabilitation [36]. Naturally, such techniques also apply to a number of other related real-world applications [37], ranging from augmented reality [38] to bad-posture detection [39,40], sitting-posture sensing [41], occluded-pose reconstruction [42] and other modern health-related applications.

It is evident that computer-vision research has produced a wide range of marker-based and markerless technologies in recent years, with the potential to be utilized in a wide range of disciplines and settings. The following issues, however, must still be addressed:

- Requirements for a markerless motion-data acquisition systems remain reliant on the research field and the unique physical-acquisition settings, and so differ between disciplines [43].
- Human motion analysis systems should be very precise to detect minute changes in motion in sports biomechanics and physical rehabilitation tasks, as well as adaptable, noninvasive, and free of constraints [44]. It should be emphasized that resolution

(both spatially and temporally) has the same effect on markerless systems as it does on marker-based systems [45].

- One must also acknowledge that the bulk of the data recorded will be significantly bigger, and so, the markerless systems may need to forego accuracy in order to construct a deployable, rapid system, which may pose difficulty when performing a proper medical analysis. Machine learning for this type of system will almost certainly entail the procurement of a costly graphical processing unit (GPU)-oriented frame processing machine in order to handle vast volumes of video data accurately and efficiently, thus stopping the actual deployment [46].

The main novelty of this paper is a novel deep-learning-based technique for remote rehabilitative analysis of image-captured human movements and postures. We present a proprietary plonynomial Pareto-optimized deep-learning architecture for processing inverse kinematics for sorting out and rearranging human skeleton joints generated by RGB (red, green, blue) image-based two-dimensional (2D) skeleton recognition algorithms, such as blazepose [47], with the goal of producing a full 3D model as a final result. *Our model differs from other implementations in its activity-independent architecture while still ensuring anthropometric regularity and retaining high joint mapping accuracy with smooth motion frames.* The proposed approach allows to extract the entire humanoid character motion curve, which can then be bound to a 3D mesh for preview in near real time. In addition, because the entire video feed is treated as a single entity instead of processing on a frame-by-frame basis, this allows for smooth interpolation between poses, where the interpolation accuracy can be managed by the video-feed sampling rate. The sampling rate can be lowered for faster video preprocessing in exchange for accuracy and vice versa, allowing all calculations to be performed on CPU-based processing systems, rather than expensive GPU farm equipment.

The paper is organized as follows. Section two provides an overview of the state of the art; section three focuses on methods and materials, offering a background for this research, the developed kinematic model of a human skeleton, the advanced joint-topology detection backbone network, and the Pareto optimized deep-learning architecture for inverse kinematic processing, in addition to computer vision processing back-end development. The paper then continues by describing the materials utilized in this assessment, the experimental setup, limitations and examples, metrics employed, and outcomes on both proprietary and benchmark MPII datasets. Following that, a robustness assessment on the KIMORE dataset is provided, and the result section concludes with an evaluation of user experience with such a system. Following that, the paper provides discussion and conclusions.

2. State-of-the-Art Review

State-of-the-art pose assessment applications now drive towards a 3D space, which provides quite an interesting research context. A human person can easily “reconstruct” a 3D shape from a 2D-only image. However, this poses a problem to “classic” or deep-learning processors, as the algorithm has to fit joints to a 3D skeleton [48]. Kinematic features of full 3D human pose representations are high-dimensional and difficult to estimate directly. The more poses a model aims to support, the higher the complexity increase in the distribution of poses, and the higher the dimensionality of the model [49]. Dynamics of motion is another problem often resulting in misalignment of the estimated mesh [50]. This overview offers insights into research on pure 2D image-based conversion, intentionally omitting depth-based approaches [51], as this was not the objective of this study and leads to the limitation of dedicated depth sensors.

Human pose estimation itself can be defined as targeting an estimation of the position and/or spatial location of body key points in an image frame [52]. The result is often a pose composed of joints, depicting the structure of a human body [53]. Traditional (classic) methods often include the pictorial structure model targeting the analysis of a kinematic tree to find the main joints of a human body [54], using the histogram of the orientation gradient (HOG), the hue, saturation, and value (HSV) color model, and other methods to acquire information about shape, color, and other parameters, which are later used to infer the human pose. Such simple but low-performing models are now replaced with those based on deep-learning approaches with the advantage of good robustness and the capability to learn pose characteristics from global space [55]. Still, no matter whether a classic or modern approach, the process of estimating human poses is similar, as first the algorithm must localize the joints of the human body and then group this information into a representation of a pose [56].

One of the most popular approaches is pose machines, which provide sequential prediction methods for learning spatial models [57]. Many convolutional neural network (CNN) approaches were modified to support more dimensions of skeleton keypoint matching [58] and then improved, for example, by applying spatial-temporal graph convolutional networks [59] to learn both spatial and temporal patterns from 2D images. Another approach [60] suggested using probabilistic knowledge of plausible 3D keypoint positions to fine-tune the search to estimate a 3D human pose with a multi-stage CNN architecture. Zou et al. described a graph convolutional network (GCN) that works on regression tasks on graph-structured data, which can improve performance with negligible overhead [61]. Pavvlo et al. [62] suggested using dilated temporal convolutions on 2D keypoints in combination with backprojection for training to use unlabeled video data. Moon et al. [63] showed that a 3D CNN could be used for voxel-to-voxel-like mapping and prediction to solve depth problems, using a 3D voxelized grid for estimation of per-voxel likelihoods for each joint. Gonzales et al. suggested that using depth data to obtain 3D lifted points from 2D information can provide a rough estimate of the true 3D human pose, by separating the 2D pose estimation from the 3D pose refinement [64]. Volumetric representations might work well to achieve good pixel-level localization accuracy, but unfortunately often lead to unrealistic body structures as a result. This can be solved by linking body mesh estimation and 3D keypoint estimation [65]. Most of the above CNNs rely mostly on scale-invariant, translation-invariant, or rotation-invariant operations, such as max-pooling, and have a problem with viewpoint generalization, which can be solved by autoencoder architectures [66]. A combination of GCNs and temporal convolutional networks (TCNs) can be used to estimate multi-person camera-centric 3D poses that do not need to know the camera parameters [67]. As another alternative to CNN, Gartner et al. propose the use of a deep-reinforcement-learning architecture capable of estimating appropriate views in space and time to help with final frame pose estimation [68].

Considering the nature of the approach, conversion from 2D to 3D space leads to numerous artifacts in the results, as well as to limitations in the input data. The occlusion was partially solved by applying a Fisher hierarchical matrix distribution to the relative 3D joint rotation matrices of key points and a Gaussian distribution to the body shape parameters [69]. Cheng et al. [70] also suggested employing the estimated 2D confidence heat maps of joints, together with an optical-flow consistency constraint, then filtering out unreliable estimations of occluded joints. As an alternative to CNN-related problems of depth ambiguity and self-occlusion, Li et al. [71] suggest using a multihypothesis transformer architecture to learn spatiotemporal representations of multiple plausible poses. Ma et al. [72] offer applying transformer-based architectures to solve the problem of occlusions and oblique viewing angles by integrating information from different views. Loss in keypoint information or even segmentation is another big issue, especially when trying to reconstruct clothed figurines, as was shown by Dwiveli et al., who used higher level semantic knowledge about clothing to distinctly penalize the clothed and unclothed image regions [73]. Out-of-domain human pose estimation was performed using the Bilevel Online Adaptation algorithm, which uses temporal constraints to compensate for the unavailable notation, and bilevel optimization procedures to solve the mesh generation [74].

3. Methods and Materials

The purpose of this section is to outline our methodology for the assessment of human posture. Our technique replicates human skeletal postures, deforms surface geometry, and is independent of camera poses at each time step of the depth video. For this objective, an original kinematic model matching to the skeleton-posture data gathered from test individuals utilizing a set of reference systems with a simple link, limiting the degrees of freedom, was required. A joint topology analysis network was created to read and process data for this model, with the purpose of involving a total of 33 human-body keypoints as a superset skeleton topology, allowing for more accurate posture identification. The Pareto optimized deep-learning architecture was introduced to speed up the technique while maintaining good precision for practical assessment. Finally, the methodology concludes with a system-level description of the backbone that we utilised in our approach.

3.1. Background

Multiple computer-vision problems can be considered and analyzed as convex optimization, which can be addressed by mathematical calculation of the global optimum of a 3D model [75]. However, many of these problems can be non-convex and poorly solved. As a consequence, there may be many optima for which the solution is missing, ambiguous, or unstable, especially under realistic conditions with noisy or corrupted data. In terms of non-expressivity, e.g., computer-vision segmentation can be represented as an energy optimization problem [76], which could then be applied to define an energy function for pixel labels, and the best solution can be found by minimizing the energy [77]. When the given energy function is complicated, finding the energy minimum value precisely is NP-hard, and convex solvers are not able to efficiently check a very large number of local optima without additional constraints. As for the ill-posed problem, many problems require optimization of the parameters of a given numerical model in order to reconstruct observations. Specifically, in computer skeleton tracking problems, different hyper-parameters have to be fine-tuned in order to model "human likeness" [78]. Considering the quantity and quality of available training instances, discovering a parameter setting that can reconstruct the training labels can be nearly impossible.

Our method reproduces the poses of the human skeleton, deforms the surface geometry, and is independent of the camera poses at each time step of the depth video. The skeletal arrangements and camera poses are discovered by solving a joint energy minimization problem that optimizes the matching of RGBZ camera (which acquires both color (RGB) and time-of-flight (ToF) range (Z)) data and the matching of human shape templates to the depth data. The energy function combines geometric matching, indirect scene segmentation, and matching using image features, which has an impact on performance. An extremely large data stream consisting of geometric matching and the integration of human-position and camera-position estimation allows for reliable capture of the results, even though the tested AI methods were trained with data from only two depth sensors. Unlike previous activity capture methods, the algorithm we investigated in this iteration performs well in processing common unsupervised indoor scenes, where a potential patient performs rehabilitation exercises by filming himself with a mobile phone (selfie mode).

3.2. Kinematic Model of Human Skeleton

Formally, the human skeleton kinematic model can be defined using Denavit–Hartenberg (DH) parameters [79] by $\eta(L^H, J^H)$, where L^H and J^H correspond to the sets of skeleton links and joints. The kinematic model corresponds to the skeletal pose data obtained from test subjects using a set of reference systems with a simple connection, limiting the degrees of freedom (DOF) of a 3-D space to four fundamental transformations. A direct kinematic model for a discrete series of joints may be found using this technique. The approach establishes various reference systems, one for the skeleton's basis and one for each of the joint linkages. The geometric transform between consecutive reference systems is achieved by combining four movements: translation a along the x-axis, translation d along the z-axis, rotation α around the x-axis, and rotation θ around the z-axis. In the case of a rotational joint, the variable influenced by the joint would be θ or d in the case of a translational joint.

For each skeleton joint, a transformation matrix can be defined as follows:

$$
\begin{aligned}
{}^{k-1}A_k &= Rot_z(\theta_k)T(0,0,d_k)T(a_k,0,0)Rot_x(\alpha_k) \\
&= \begin{bmatrix} \cos\theta_k & -\cos\alpha_k \sin\theta_k & \sin(\alpha_k)\sin\theta_k & a_k\cos\theta_k \\ \sin\theta_k & \cos\alpha_k\cos\theta_k & -\sin\alpha_k\cos\theta_k & a_k\sin\theta_k \\ 0 & \sin\alpha_k & \cos\alpha_k & d_k \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

By multiplying the matrices for each skeleton joint successively, the coordinates of the endpoints can be achieved as follows.

$$
\begin{gathered}
{}^{B}A_0 = T(0,0,l/2)Rot_y(-\pi/2) \\
{}^{0}A_1 = Rot_z(q_1)T(0,0,0)T(0,0,0)Rot_x(-\pi/2) \\
{}^{1}A_2 = Rot_z(q_2)T(0,0,l)T(0,0,0)Rot_x(\pi/2) \\
{}^{2}A_3 = Rot_z(q_3)T(0,0,0)T(0,0,0)Rot_x(-\pi/2) \\
{}^{3}A_4 = Rot_z(q_4)T(0,0,l)T(0,0,0)Rot_x(0) \\
{}^{B}A_4 =^{B} A_0 \cdot^{0} A_1 \cdot^{1} A_2 \cdot^{2} A_3 \cdot^{3} A_4
\end{gathered}
$$

The human body is divided into 13 rigid segments: head, forearms, upper arms, torso, pelvic segment, thighs, shanks, and feet. In addition, 1-DOF hinge joints depict the ankles, knees, and elbows. The 3-DOF spherical joints depict the hips, shoulders, and joints that connect the pelvic segment to the torso and the torso to the head, respectively. The pelvic section, which can move freely in space and, hence, has six DOF, serves as the mechanism's foundation. The kinematic model comprises a total of 24 articulated rotational DOF, including the 3 rotational DOF and 3 translational DOF of the pelvic segment, which determine the body's location and orientation in relation to the reference coordinate system.

The human arm may be described as a rigid kinematic chain with three joints (shoulder, elbow, and wrist) and seven joints $(q_1, q_2, q_3, q_4, q_5, q_6, q_7)$. The positions (d_1, d_3, d_5, d_7) represent the lengths of the links that connect the torso, shoulder, elbow, wrist, and hand. The mobility of the upper arm is determined by the structure of the shoulder, which consists of three joints (q_1, q_2, q_3), where q_1 controls its forward and backward motion, q_2 controls its downward and upward motion, and q_3 controls its rotation. The extension and flexion of the forearm are determined by the anatomy of the elbow, which has one joint q_4. The structure of the wrist with three joints (q_5, q_6, q_7) regulates the mobility of the hand, where q_5 represents the rotation of the forearm, q_6 represents the extension and flexion of the hand, and q_7 represents the rotation of the hand.

The geometric structure of the arm is used to determine the joint angles and swivel angles of the human arm according to [80]. Suppose that S_L and S_R are the left and right shoulders, M is the midline of the shoulders, T is the torso position, E is the elbow, W is the right wrist, and H is the right hand. θ_j is the angle of joint $q_j, j = 1, 2, \ldots, 7$. ψ is the swivel angle between the reference plane and the arm plane.

The geometry relation is used to compute the corresponding joint angles. The angle between the reference plane and the arm plane is determined by θ_1. θ_2 calculates the angle between the vector $S_R E$ and $S_R T$, θ_3 denotes the rotation angle of the upper arm, and θ_4 calculates the angle between the link $S_R E$ and EW. θ_5 is the rotation angle of the lower arm, and θ_6 is the angle between EW and WH. To simplify, we select $\theta_7 = 0, d_7 = 0$, because the joint angle θ_7 cannot be calculated from the sensor data. Human arm-hand postures are computed using joint angles depending on the link structure's DH parameters.

The forward kinematic problem between end pose ${}^{0}T_7$ and joints coordinates $(j, j = 1, 2, \ldots, 7)$ can be handled by the coordination transformation matrix ${}^{j-1}T_j$ from joint $j - 1$ to joint j, according to the parameters of the human arm DH, where

$$
{}^{j-1}T_j = \begin{bmatrix} \cos\theta_j & -\cos\alpha_j\sin\theta_j & \sin\alpha_j\sin\theta_j & 0 \\ \sin\theta_j & \cos\alpha_j\cos\theta_j & -\sin\alpha_j\cos\theta_j & 0 \\ 0 & \sin\alpha_j & \cos\alpha_j & d_j \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

3.3. *Joint Topology Detection Backbone Network*

One of the industry standards for human posture estimation is *COCO*, which consists of 17 joints tracked on the human body. However, these points are not enough for a certain posture estimation, as they lack scale and orientation information for the limbs that are vital for applications trying to evaluate fitness. For this reason, the inclusion of additional trackable keypoints is crucial for detecting bad human posture and estimating pose. Our solution involves a total of 33 human body keypoints as a superset for the COCO skeleton topology, allowing for more accurate posture detection. All subjects have the same skeleton topology, and suitable bone lengths are computed per subject (by automatically scaling a reference skeleton based on the person's known height).

For pose estimation, a two-stage detector–tracker pipeline was used, where the detector locates the region-of-interest (RoI) for the given image, while the trackers predict the joints within the given RoI. In our model, we establish a series of IMU track targets, each of which is associated to a bone and has the IMU's rotational and translational offsets. IMU orientation readings and derived acceleration measurements are used to express the rotational transform between each IMU reference frame and the global coordinates. The estimation and tracking of joint location estimation, depicted in Figure 1, is performed by training the network on top of the combination of BlazePose and BlazeFace models [47]. Reusing the BlazeFace model allowed us to have a baseline that could be used to track the orientation and position of the human, whereas the joint position baseline was based on that of the BlazePose model.

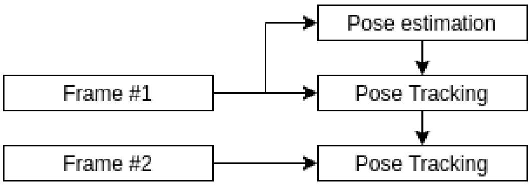

Figure 1. Pose estimation pipeline.

Our network further augments these technologies by substituting missing angles from the original skeletons calculated from the depth data (we used three RealSense depth cameras at different angles). As some of the joints that the network predicts are redundant, e.g., fingers or toes, the reduced joint input is passed into the neural network as an input. However, to bind human joints to a skeleton, their position in 3D space is insufficient. This task requires calculating the joint orientation based on its parent joints, with the complexity increasing with each additional joint, thus requiring the development of a dedicated deep learning-based inverse kinematic model. To obtain real-time performance, each network component must be suitable for real-time applications. To achieve this, the assertion is made that the best recognizable feature of the human torso direction is its face. This assertion allows the solution to simplify pose estimation by using a face detection network capable of real-time performance. Based on the link between the human torso and human head, we can extrapolate the radius of the human body, which enables faster detection of key features such as body center rotation and scale, which are required to estimate complicated poses.

The architecture of the backbone network adapted for our system is given in Figure 2. It takes a 256 × 256 RGB image as input and outputs 33 tracked joint 3D locations along with their visibility. Unlike other methods, which use computationally prohibitive heatmap predictions, this backbone only uses supervised learning combined with heatmap/offset prediction during the training process. The acquired keypoints can then be processed and used in tasks dependent on their location, for example, detecting the performed action or evaluating its deviation from the baseline. Using joints, we extrapolate other required features for posture evaluation, these include: the top and base of the spine; top and base of the neck, and combine them into the final skeleton. Using sequence analysis or manual detection, we determine the action that the subject is about to take and compute the required errors based on such features as neck angle in relation to the shoulders, spine shift based on the vertical axis, or shoulder shift based on the horizontal axes. If a defined error threshold is reached, the user is informed that the performed action was completed improperly and which parts of the body (and by how much) have failed to fit the posture. A constraint is also added for each 2D spatial measurement from the smartphone camera. This function aims to reduce the Euclidean distance between both the measured 2D position in tracked human skeleton coordinates and the calculated (predicted) global track target position mapped.

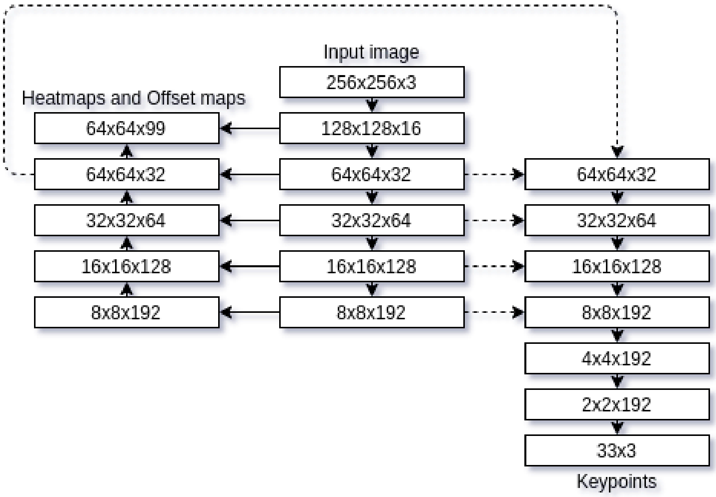

Figure 2. Backbone network architecture.

3.4. Pareto-Optimized Deep-Learning Architecture for Inverse Kinematics Processing

In human movement, there is often no clear preservation of bone lengths and rotations of the limbs are not evaluated. Our method uses keypoint detection over multiple camera frames as input for kinematic optimization to obtain an accurate 3D human pose. This kinematic pose output preserves bone length and can be used directly to perform 3D simulation. Our method attempts to "lift" a 3D pose from a single RGB image by including the knowledge of plausible human poses into the training, refining the detection estimates using a complex method. The 3D human pose is triangulated from a single camera, using CNNs, and a kinematic fitting, using an offline expectation maximization approach over the entire sequence. In our approach, surface grids are applied to single-view images, yielding dense pose correspondences. In unbounded environments, monocular methods are characterized by depth ambiguity, hindering their capability to reconstruct the full pose on absolute scale.

Inverse kinematics (IK) is generally an iterative approach; this makes it hard to apply for real-world tasks. A deep-learning neural network architecture (seen in Figure 3) has been created that allows for single-pass IK prediction of the skeleton as opposed to an iterative one. Here, the input is the joint position in the 3D space, and the output is the joint rotation in the 3D space represented as Euler angles in radians. The first network stage extracts 1024 most useful features from the input that are required for IK calculations, and 23 parallel predictions (one for each joint) are made and combined into a single feature vector. The resulting feature map uses a branched prediction approach to predict joint rotations. Joint-location predictions of the backbone network were used as an input for this inverse kinematics network, for the loss function evaluation, the expected IK results of the model are calculated, giving a ground-truth value. The network was then trained using the created dataset until it converges.

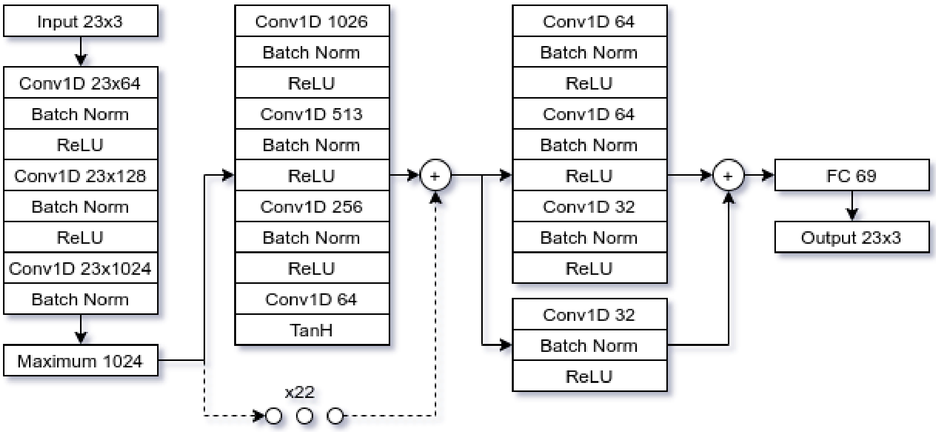

Figure 3. Inverse kinematics of a skeleton, resulting output is Euler angles.

To assist in determining optimal skeleton keypoint values, the Pareto dominance relation was generated as a tuple on the possible skeleton bone state space, so that an optimization model is one that is not dominated by any other solution. The goal is to find the trade-off surface, also known as the Pareto front, which is the collection of non-dominated solution locations. These are also known as Pareto-optimal solution points. The tracked subject's kinematic stance is Pareto-optimized as shown below.

The Pareto set comprises the optimal answers when alternative options that may improve at least one of the objectives without deteriorating any other are not available. For multi-objective problems, there is usually a set of solutions that offers the most information about optimization. After multi-way comparison of solutions, those that meet the fewest number of objectives are labeled as inferior. The decision vector $x^* \in F$ is the Pareto optimal solution to maximize the objective functions k if no alternative decision vectors meet both of the following conditions:

$$\begin{gathered} f_i(\mathbf{x}) \geq f_i(\mathbf{x}^*), \forall i \in \{1,2,\ldots,k\} \\ f_j(\mathbf{x}) > f_j(\mathbf{x}^*), \exists j \in \{1,2,\ldots,k\} \end{gathered}$$

If these requirements are met, the decision vector x dominates the decision vector y in the maximizing problem, as shown by $x > y$. This is written as

$$\begin{gathered} f_i(\mathbf{x}) \geq f_i(\mathbf{y}), \forall i \in \{1,2,\ldots,k\} \\ f_j(\mathbf{x}) > f_j(\mathbf{x}^*), \exists j \in \{1,2,\ldots,k\} \end{gathered}$$

The Pareto-optimal set is defined as follows: if no solution in the search space dominates any member of the set P, then the solutions in the set P create a global Pareto-optimal set. All feasible solutions in the complete solution space are acquired after optimization, eliminating the requirement to combine all objective functions into one. Using the Pareto-optimal frontier sets, one can choose the best solution based on their preferences.

3.5. Image Processing Backend

The network input can be any type of sensor that captures RGB information, e.g., webcam, phone, camera, etc. Each of the video frames is then extracted at the desired frame rate, and the resulting frame is then sent to the estimator, where its processing is performed to extract the object's joint orientation. To playback a video containing the estimated subject posture, firstly we need to preprocess each of the frames by first evaluating the object skeleton in our network; this gives us the estimated object joint positions in 3D space from a flat RGB image. After the joint locations for the poses are estimated, they need to be contracted into an excessive joint model by omitting such features as fingers for better and faster estimation, as these features do not add any important information for the task. Once the joints are selected, an IK neural network is used to estimate the orientations of the joints so that they could be bound to the skeletal mesh. Our approach stores the skeleton's posture as a single multidimensional vector representing the root's 3D global translation. The process is then completed by combining the stacking local joint rotations of each bone (along with the root), which are expressed as 3D angle–axis vectors derived by multiplying the axis of rotation by the angle of rotation in rads. The variable that is optimized is this parameter vector, with the root movement and joint rotations extracted and used in the calculations as needed. Extracted object joints are then bound onto a selected body mesh and saved; the process is repeated for all videos. Once the video is processed, it can be played back and interacted with in 3D space, allowing for the expert to change the camera location to the desired position and to observe the performed action from multiple angles. This entire process is illustrated in the sequence diagram displayed in Figure 4.

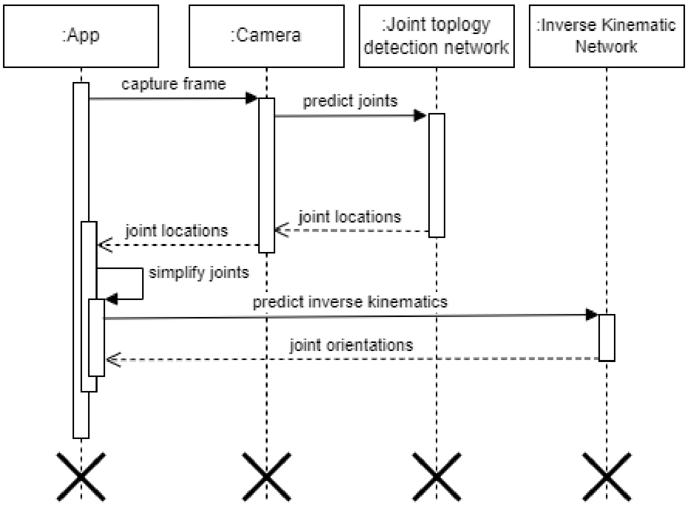

Figure 4. Estimation of a skeleton joint orientation.

Due to the nature of rehabilitation-exercise data, our technique also required variable-length processing. In contrast to generalized computer-vision applications that use sequential data (action/motion categorization in general), rehabilitation-kinesitherapy data show considerable changes within variable-length data. One prominent explanation might be that exercise performers are a broad group of people, ranging from experienced therapists to patients with various disorders and problems. Furthermore, rehabilitation data can be obtained in confined lab/gymnasium settings, indoor/outdoor settings, and home settings. In addition, rehabilitation activities may require a different number of repetitions depending on the therapist's prescription. As a result, individuals direct varying amounts of time to the same exercise with much the same number of iterations.

A simple spatio-temporal approach allows us to take the average or maximum response for a predefined window size while maintaining the time scale and without losing some subtle features, all while keeping the main model for predicting correctness score. Sequential dependencies are represented in the spatiotemporal feature vectors in the processing back-end. It allows us to extract the discriminative traits that have been collected over time. This little information is critical in predicting an exercise's accuracy score.

Dense trajectories have been shown to be useful in recognition of human actions [81]. They represent a person's walking patterns and can be easily computed straight from sensor data; we utilized them to compute skeletal joint characteristics. To calculate trajectories, we choose a group of dense points from a frame and track them in consecutive frames. In a dense optical-flow field, tracking is performed using displacement information. In frame $t+1$, each picked point $P_t = (x_t, y_t)$ in frame at time t is tracked. A trajectory is formed by the concatenation of these tracked sites into successive frames. S represents a displacement vector sequence and is calculated as,

$$S = (\Delta P_t, \Delta P_{t+1}, \ldots, \Delta P_{t+L-1}),$$

where L is the length of the trajectory and $\Delta P_t = (P_{t+1} - P_t)$. The resulting vector S is normalized as follows:

$$S' = \frac{(\Delta P_t, \ldots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|}$$

Given a skeleton pose sequence S' with N frames, defined as $S' = F_t$, where $t = 1, 2, 3, \ldots, N$. Let p_{tj} and p_{tk} be the 3D positions of the j-th and k-th joints in F_t. Then, the inter-joint distance IJD^t_{jk} between p_{tj} and p_{tk} at time sample t is calculated as

$$IJD^t_{jk} = ||\mathbf{p}^t_j - \mathbf{p}^t_k||_2, (t = 1, 2, 3, \ldots, N)$$

The angle of the human joint and the angle variation can be employed as deep spatial characteristics in an action-detection job based on bone data. When combined with the 2D (x, y) coordinates provided in the human-skeleton dataset, the length of the bone and the angle between the bones are easily calculated. We can simply depict the location of connections in the human skeleton using these variables.

It is important to compute the angle created by its two bones. The length of each bone and the angles between the joints are estimated on the basis of the coordinates of each node in the human-bone dataset and its physical relationship.

$$\theta_i = \arccos\left[\frac{\left(x_{pred(i)} - x_i\right) \times \left(x_{succ(i)} - x_i\right) + \left(y_{pred(i)} - y_i\right) \times \left(y_{succ(i)} - y_i\right)}{\sqrt{\left(x_{pred(i)} - x_i\right)^2 + \left(y_{pred(i)} - y_i\right)^2} \times \sqrt{\left(x_{succ(i)} - x_i\right)^2 + \left(y_{succ(i)} - y_i\right)^2}}\right]$$

where x_i and y_i denote the abscissa and ordinate of the node i, $succ(i)$ is the successor node of node i, $pred(i)$ is the predecessor node of node i, and θ_i represents the angle on the node i. A length is calculated as follows:

$$|L_1| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

where $|L_1|$ denotes the length of the bone between the nodes, and its subscript represents the label of the first bone.

The deep-space feature matrix is obtained by extracting the angle between bone length and bone, as well as the angles and lengths between all bones in a single frame.

3.6. Materials

Postural examination is often the initial component of any patient's tests and measurements for any musculoskeletal problem. The therapist looks at the patient from the front, rear and sides. Postural assessment is a critical component of objective evaluation, and ideal static postural alignments have been proposed. However, both static and dynamic postures must be assessed to determine the patient's functional mobility and capacity to self-correct static habits. Scoliosis, postural decompensation, anatomic short leg, previous trauma or surgery, trunk control (after stroke), or particular segmental somatic dysfunctions in areas of the body where asymmetry is present can all be caused by postural misalignment or asymmetries and are very important to assess at the beginning of the rehabilitation program [82,83].

Stroke causes a wide variety of clinical forms of postural instability and pathological conditions. The variety of stroke and the impairments of the patients is one of the primary challenges in the treatment of postural instability after stroke [84]. Workouts, sensors, modalities, actuators, communications, configurations, and end users were all characterized in the literature review study by Hribernik et al. [22], where they summarized that real-time biomechanical input can accelerate physical recovery. During a postural assessment, a medic usually measures and records active ROM at all peripheral joints, especially shoulder/upper limbs. Many pertinent questions about the rehabilitation of postural imbalance after stroke are raised:

- Which physical training method is the most effective?
- What is the most important difference between such a specialized technique aimed at postural imbalance and the generalization benefits of a non-specific approach?
- Is postural imbalance therapy only a sensory approach?
- What is the benefit of technology?
- What is the effectiveness of the training program in relation to the period after the stroke?
- Which workout intensity is the most effective?
- What are the implications for autonomy and quality of life?

Causes of limitations may include pain, weakness, muscle shortening, or swelling. Shoulder ROM limitations, muscle performance and strength deficits may can affect changes in posture or incorrect posture. That is why we used full active ROM of shoulder (extension/flexion/rotations) during assessment, as factors most influencing postural changes or occurring compensatory mechanisms (e.g., full shoulder flexion influences trunk control and leads the hyperextension of the back). Roland et al. [85] explored if bilateral standing with visual feedback treatment increases posture and balance after stroke as compared to traditional therapy, as well as the applicability of the benefits of visual feedback treatment on gait and gait-related tasks. Yu et al.'s experimental research [25] included the muscle stimulation location, stimulation parameters, transition of each stage, and transition between each condition. Their findings demonstrated the usefulness of the control technique for clinical rehabilitation-plan formulation and rehabilitation-training execution. The posture recognition experiment was based on the recognition of six fundamental upper-limb motions. By making comparisons with various classification methods, the viability of a fully connected neural network in posture recognition has been determined, and it can be merged with the rehabilitation assessment of patients in the later stage, and it can serve as a reference basis for the evaluation of the extent of rehabilitation of patients, which is extremely important for the recovery of patients.

To create the dataset, two motion capture databases were used. The first dataset used to train the neural network is *MoVi* [86], it contains vast amounts of motion capture data from multiple camera perspectives. Each of the videos consist of subjects performing multiple different actions; the videos are captured using regular camera sensors.

The second dataset was custom-made by us, using exercise database recorded using two different depth sensors and coordinated by the experts from Vilnius Santara clinics. An example of the dataset can be seen in Figure 5. The first depth sensor *Intel Realsense L515* was placed in front of the subject, the second depth sensor *Intel Realsense D435i* positioned 90° right of the subject. Both of the sensors were placed 1.4 m above ground level and 1.8 m away from the subject. The database consisted of 23 subjects of varying physical conditions, age from 26 to 71, 61% male, 32% female, and the rest identified as other genders). In total, 168 recordings were made for each participant. We provided a demonstration by an expert rehabilitation specialist explaining the purpose and methodology behind the exercises. Each of the subjects performed these action movements from the diagnostic rehabilitation functional assessment methodology [87,88]:

1. Shoulder flexion (90°);
2. Shoulder flexion and internal rotation;
3. Shoulder flexion and internal rotation, elbow flexion;
4. Shoulder extension and internal rotation;
5. Shoulder flexion;
6. Full shoulder flexion (180°);

7. Shoulder adduction.

Each of the videos in both datasets was pre-processed to be used for neural network training. This was performed by extracting every 0.5 s to minimize the number of similar frames in the video feed; a breakdown can be seen in Table 1. The frame was then passed to our network to extract its joints and, subsequently, each of the joint orientations was calculated for IK neural network training. Each person also wore eight HTC VIVE trackers to gather the data to check and tune the accuracy of the kinematics model. The HTC trackers provide an IK solution of their own, allowing us to increase the accuracy of the joints containing the HTC trackers by obtaining a mean orientation of the HTC joints and the calculated joints.

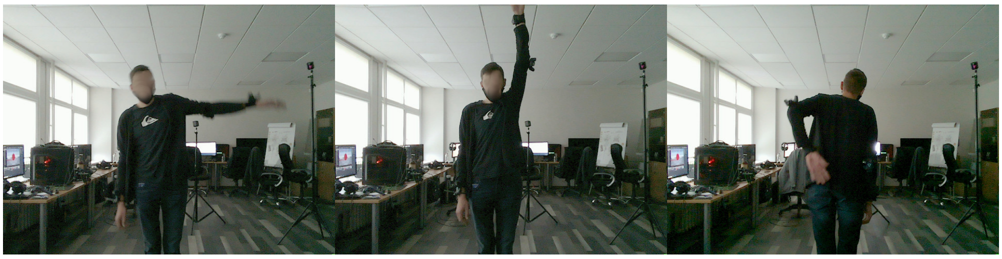

Figure 5. Dataset example.

Table 1. Real-world dataset breakdown captured for a single perspective selfie video (average distribution for different subjects).

Exercise No.	Average Distribution per Subject
Ex. 1	99
Ex. 2	76
Ex. 3	78
Ex. 4	87
Ex. 5	92
Ex. 6	71
Ex. 7	70
Total	**592**

4. Experimental Evaluation and Results

4.1. Experimental Setup

Dealing with a video sequence requires processing large amounts of data; even a single recording of a person performing an exercise can contain thousands of frames, each of which must be evaluated individually. When the task is expanded to deal with multiple camera calibrations, multiple exercises, etc., this quickly becomes an enormous task of big data that cannot be processed manually. For this, tools that would allow to optimize the video-feed evaluation are required. Our method attempts to simplify the big-data task of human posture evaluation by providing such processing tools. During the training process, we used *MoVi* and our own created data sets split using the 80:20 rule where 80% of the data set recordings are used for training and 20% of the remaining recordings are used for testing. However, this solution can still bias the results, as the results can be chosen so that the remaining 20% have good results but not necessarily be a generalized solution fit for all cases. Therefore, during these experiments, a new validation dataset consisting of eight videos was prepared that attempted to encompass all edge cases. For example, these edge cases involve determining if the network can detect user actions such as bending sideways, forward/backward, or touching their face. This validation dataset includes these edge cases because our solution relies on monocular images that lose a lot of depth information that can only be inferred from certain lightning conditions, i.e., shadows and based on subjects' orientation to the camera.

4.2. Limitations and Examples

To evaluate the method edge cases, five common movements evaluating changes in body pose and related to human mobility, motion performance, and posture (walking, climbing the stairs, balance, self-care, or work activities) were selected, where the test subjects are leaning left, leaning right, leaning forward, leaning backward and finally the subject touching their face (seen in Figure 6). As we can see, in most cases where the subject is holding their hand in front of the face, our solution successfully predicted the joint orientation, except for where the subject was instructed to fully touch their face. This is due to the model being unable to accurately predict the user's hand depth relative to the face, and not the IK method itself, suggesting further research could be expanded by a more accurate depth estimation method or depth sensor. The disparity between object joint orientation and IK prediction can be seen in Figure 7; as we can see, some misalignment can occur due to the fact that IK must ensure that the joints create a valid human pose without scaling the human body, as our solution does not, at the moment, account for the person's height.

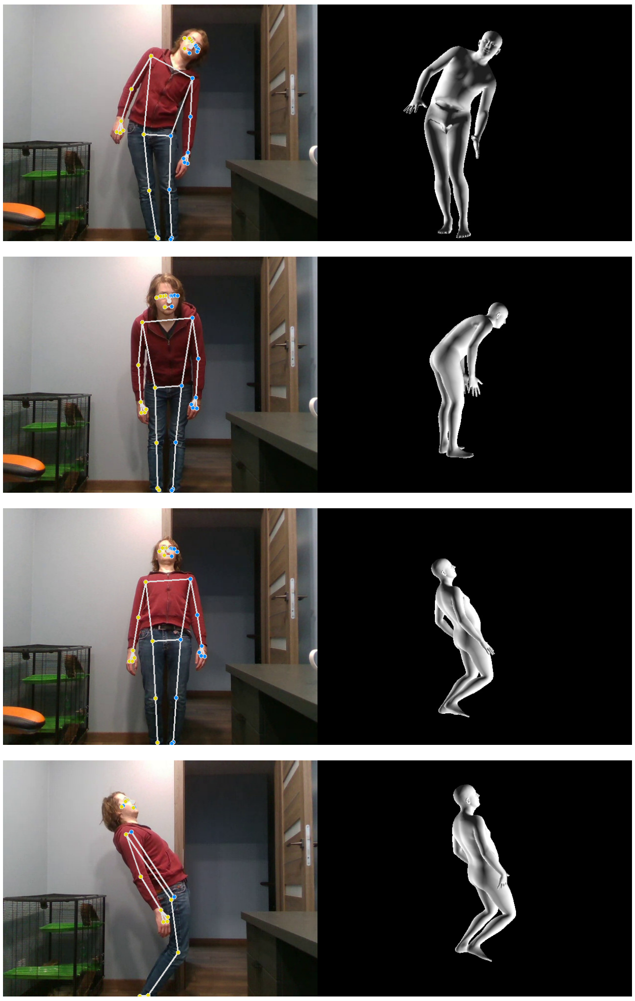

Figure 6. *Cont.*

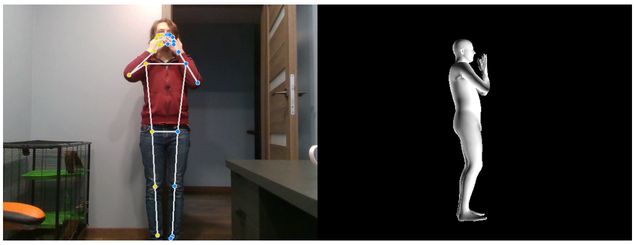

Figure 6. Subject image (**left**) and its pose reconstruction (**right**) (From top to bottom): subject leaning left, subject leaning forward, subject leaning backwards, subject leaning sideways, subject touching face.

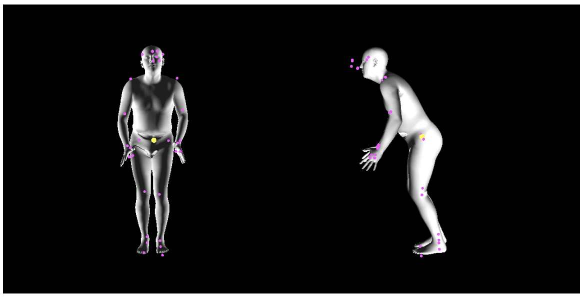

Figure 7. Example of joint overlay.

4.3. Metrics

Object keypoint similarity (*OKS*) [89] was used to calculate the average keypoint similarity between all keypoints of objects, depending on the size of the topic and the distance between the anticipated and ground-truth points. Each keypoint was assigned a similarity value ranging from 0 to 1, and *OKS* was calculated as the average of all these values at all keypoints.

$$OKS = \frac{\sum exp(-d_i^2/2s^2k_i^2)\delta(v_i \geq 0)}{\sum_i \delta(v_i \geq 0)}$$

where d_i is the Euclidean distance between a keypoint and its corresponding ground truth, s is the object scale (subject height), v_i is visibility of the ground truth, and k_i is a constant per keypoint; in our case k_1 is always 1 because we assume that all keypoints are visible as they are inferred from other joint orientation.

The mean average precision (*mAP*) [90] was used to show the accuracy of keypoint detection according to precision, was calculated as a ratio of true positive results to the total positive, and then obtained the mean of the average precision over multiple IoU thresholds throughout the model. The joints included in the evaluations are defined by the MPII dataset [54] which include: (1) right ankle, (2) right knee, (3) right hip, (4) left hip, (5) left knee, (6) left ankle, (7) pelvis, (8) thorax, (9) upper neck, (10) head top, (11) right wrist, (12) right elbow, (13) right shoulder, (14) left shoulder, (15) left elbow, (16) left wrist.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

where AP_i is the AP value for the i-th joint and N is the total number of joints evaluated.

The percentage of correct keypoints (PCK) [91] was used to measure the accuracy of the localization (a higher PCK value means better performance) of different keypoints within a threshold of 50 % of the pose head segment length of each test frame:

$$PCK = \frac{\sum(Joint_{detected} \cap Joint_{groundtruth|vis>0})}{\sum(Joint_{groundtruth|vis>0})}$$

where $Joint_{detected} \cap Joint_{groundtruth|vis>0}$ is the number of joints recognized successfully by the network, and $Joint_{groundtruth|vis>0}$ is the number of visible joints tagged in the ground truth annotations. The PCK of the joints was determined in our implementation by taking visibility > 0 into account.

Mean per joint position error ($MPJPE$) [92] was used as a metric for calculating 3D, by retrieving the Euclidean distance between the calculated 3D joints and the ground truth positions:

$$MPJPE = \frac{1}{N}\sum_{i=1}^{N} ||J_i - J_i^*||^2$$

Here, N is the number of joints, J_i and J_i^* are the ground truth position and the calculated position of the $i_t h$ joint.

Range of motion error ($ROME$) [93] reflects the accuracy in the calculation of the largest amplitude of the motion

$$ROME = [max() - min(x)] - [max(y) - min(y)]$$

where x refers to the ground-truth human joint angles, and y are the joint angles obtained using the proposed method.

The Sprague and Geers' metric [94] was used to measure the similarity of the amplitude and phase between the true and measured curves pointwise. Recently, it was used in a related context for assesing the parameters of children's gait [20].

Suppose that the two signals being compared are represented by p and m. p is the projected data from the simulation model, while m is the observed data from the experiment. The $S\&G$ magnitude error is calculated by

$$M_{SG} \ = \mathrm{sig}(rme)\log_{10}(1 + |rme|)$$

where the relative magnitude error rme is given by

$$rme = \frac{\sum_{i=1}^{N} p_i^2 - \sum_{i=1}^{N} m_i^2}{\sqrt{\sum_{i=1}^{N} p_i^2 \sum_{i=1}^{N} m_i^2}}$$

where N is the number of points. The phase error is defined by

$$P_{SG} = \frac{1}{\pi} cos^{-1}\left(\frac{\sum_{i=1}^{N} p_i m_i}{\sqrt{\sum_{i=1}^{N} p_i^2 \sum_{i=1}^{N} m_i^2}}\right)$$

The $S\&G$ comprehensive error measure is given by

$$C_{SG} = \sqrt{M_{SG}^2 + P_{SG}^2}$$

Given the small number of participants in this study, we used the Wilcoxon–Mann–Whitney test instead of the t-test to find significant differences. The test was carried out on each joint performance measure, and the significance level was set at $p = 0.05$.

The mean and standard deviations (SD) for the verified system and the criteria were calculated. Bland and Altman plots were created to visually show the data's heteroskedasticity and validity of measurements. Bland-Altman bias and limits of agreement (LoA) were estimated according to [95], Pearson's correlation (ρ), the concordance correlation coefficient (CCC) [96] and the intraclass correlation (ICC) [97] to measure agreement between spatio-temporal skeleton characteristics. Pearson's correlation and CCC evaluate the relative and total agreement between the measured values and the ground truth. LoAs were considered excellent, good, moderate, or modest if ρ, CCC, or ICC were greater than 0.9, 0.8, 0.7, or 0.5, respectively.

4.4. Results on Our Self-Collected Dataset

Generally, IK requires multiple iterations to find the appropriate joint orientation for the pose (seen in Figure 8), while the advantage of neural networks is that they can perform the same action in a single iteration, reducing the overall computation time. The ground truths calculated by the *Blender* IK solver were used to train our deep-learning-based IK solver.

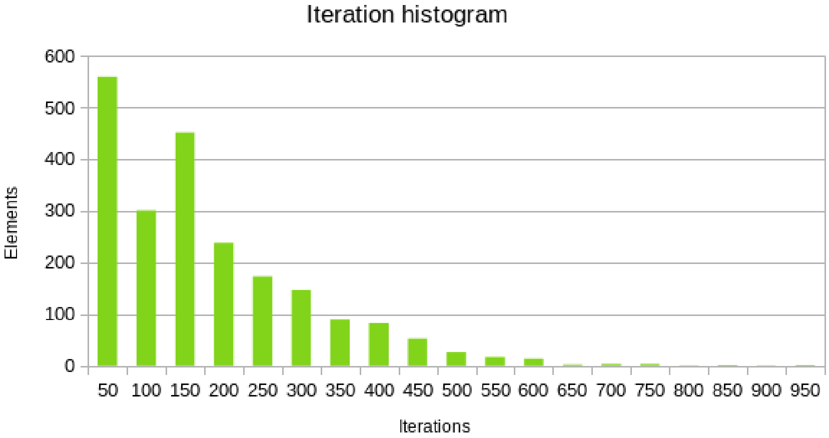

Figure 8. Skeleton inverse kinematics histogram.

The deep-learning-based approach allowed for predicting the joint orientation with high precision. To evaluate the error, the *MPJPE* metric was used, where the resulting mean square error of all skeleton joints is 0.251 with a standard deviation of 0.067, this implies high predictive capacity; a breakdown of joint position error for all poses by exercise can be seen in Figure 9.

Additionally, while predicting the human body pose in space is not a necessary task for human posture evaluation, it is an important feature to preview the action performed in a 3D environment from multiple points of view; the breakdown of error in 3D space offset can be seen in Figure 10, while the breakdown by exercise can be seen in Figure 10. Exercise breakdown (a) shoulder flexion; (b) shoulder flexion and internal rotation; (c) shoulder flexion and internal rotation, elbow flexion; (d) shoulder extension and internal rotation; (e) shoulder flexion; (f) full shoulder flexion; (g) shoulder adduction; and (h) exercises from the unsorted MoVi dataset.

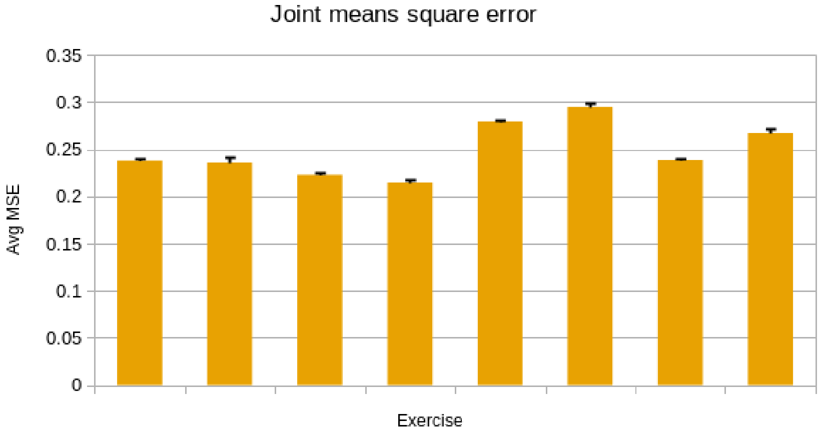

Figure 9. Skeleton inverse kinematics joint MSE by exercise.

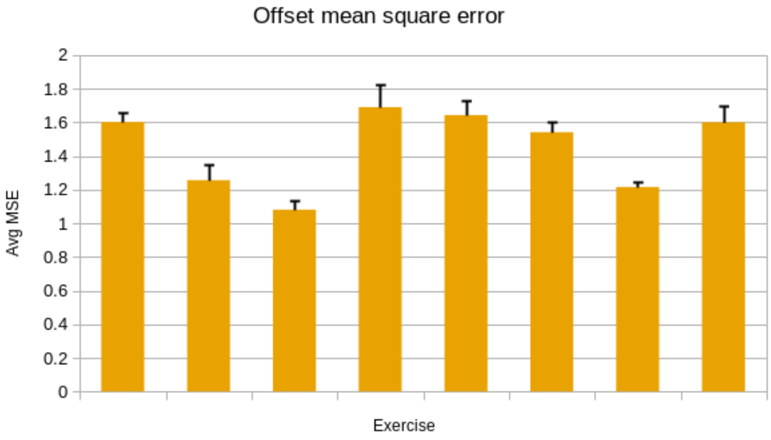

Figure 10. Skeleton inverse kinematics Offset MSE by exercise.

The achieved skeleton inverse kinematics joint MSE was stable accross the whole averaged video-frame sequence as is displayed in Figure 11.

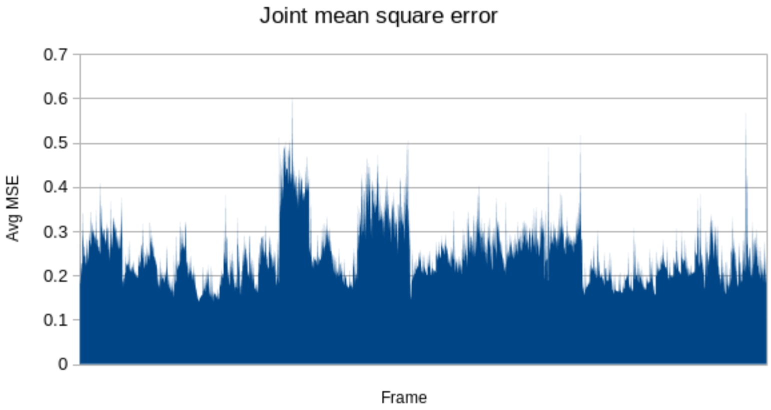

Figure 11. Skeleton IK joint MSE accross video-frame sequence of an exercise.

The average OKS value achieved is 0.989, while its breakdown by exercise can be seen in Figure 12.

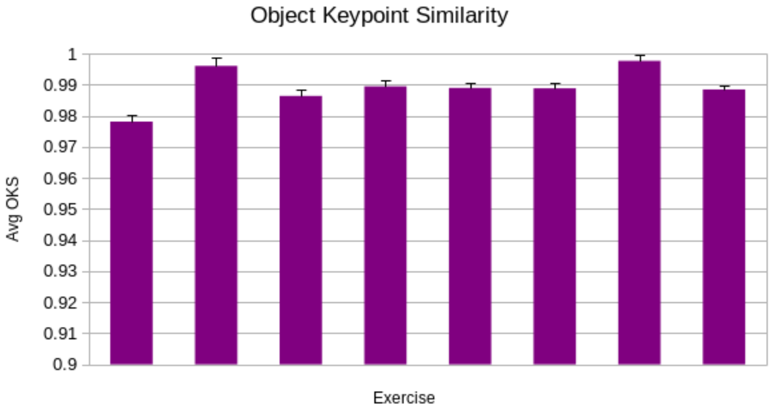

Figure 12. Skeleton inverse kinematics OKS values.

The PCK values thresholded at various detection confidence rates are given in Table 2. Generally, our method allows to achieve better results for joints, which have a lower amplitude of movement and are located at an upper part of the body.

Table 2. Summary of PCK values when thresholding at different detection confidence rates.

Detection Confidence	PCK@0	PCK@0.2	PCK@0.5
Head	0.96	0.98	0.99
Shoulder	0.90	0.93	0.94
Hands	0.82	0.87	0.92
Knees	0.68	0.75	0.82
Legs	0.55	0.66	0.74
All keypoints	0.88	0.92	0.95

Table 3 displays the RMSE and ROME for all keypoints during all exercises. In all joints, the mean values of RMSE and ROME were less than 5°, which is considered clinically valid.

Table 3. Summary of RMSE and ROME over all subjects and 95% confidence interval (CI)

Exercise	RMSE (Deg)	ROME (Deg)
Leaning left	3.2964 (1.8306–4.7622)	1.9293 (1.0408–2.8178)
Leaning forward	3.5323 (2.5728–4.4919)	1.4839 (0.3116–2.6562)
Leaning backwards	2.9320 (2.3884–3.4757)	2.1103 (1.4680–2.7526)
Leaning sideways	2.9112 (2.3884–4.1522)	1.6518 (0.5339–2.7698)
Touching face	3.0301 (1.8922–4.1679)	2.1526 (1.2196–3.0856)

To demonstrate the agreement between recorded and reconstructed signals, we used the Sprague and Geers' (*S&G*) metric. The overall error is presented by combined *S&G* error between recorded signals and the IK-reconstructed signals have a more than 97% Pearson's correlation, indicating that the recorded and reconstructed signals are very similar. Table 4 summarizes the *S&G* error as well as the Pearson's correlation for various exercises performed by all subjects.

Table 4. Mean Sprague and Geers' error and Pearson's correlation between recorded and IK-reconstructed joint positions for all subjects.

Exercise	*S&G* Error	Pearson's Correlation
Leaning left	0.019	0.982
Leaning forward	0.025	0.975
Leaning backwards	0.032	0.984
Leaning sideways	0.024	0.979
Touching face	0.029	0.971

4.5. Results on MPII Dataset

To evaluate the mean average precision of the approach, we use the MPII data set and achieved an average mAP of 0.950. The breakdown per joint is shown in Figure 13.

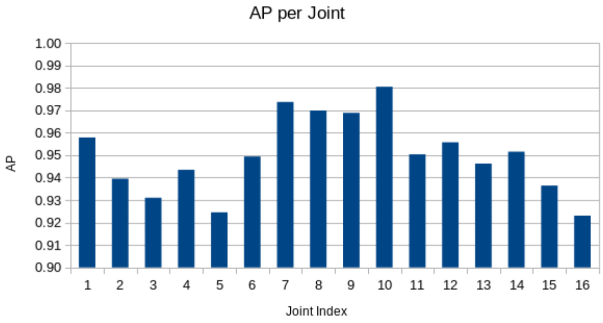

Figure 13. Mean average precision (mAP) per joint.

To further validate our results, we performed the analysis using different cross-validation schemes. As most machine-learning models improve their performance as the training set is expanded, the k-fold cross-validation may perform somewhat better than the cross-validation estimate suggests. As the difference in size between the training set used in each fold and the entire dataset is only a single pattern, leave-one-out cross-validation (LOOCV) is nearly unbiased. The results are presented in Figure 14. However, LOOCV has a high variance (so you would obtain very different estimates in repeated tests). As the estimator error is a combination of bias and variance, whether LOOCV is superior to 10-fold cross-validation depends on both quantities. This observation matches well with other similar studies [98].

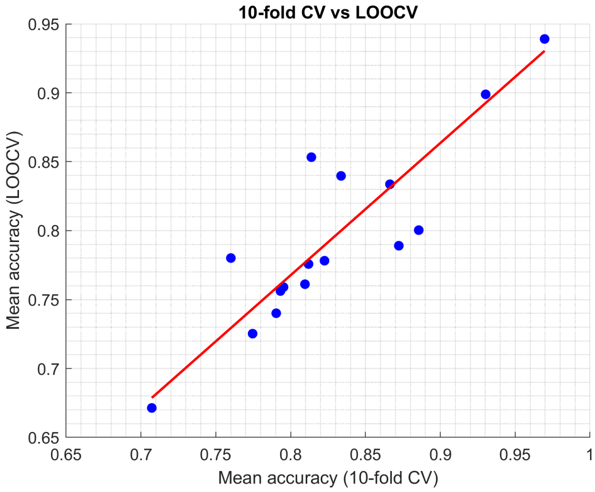

Figure 14. The cross-validation graph of accuracy distribution: 10-fold cross-validation vs. LOOCV.

A Bland–Altman analysis of agreement [95] was performed between the spatio-temporal features. The results (see Figure 15 for the vertical line characterizing the posture, and Figure 16 for the horizontal line characterizing the posture) show good agreement between the spatiotemporal features of the gait for subjects calculated using RealSense and IK analysis. The measure of validity is how close the results are to the truth on average (i.e., the higher the percentage of error on average, the lower the validity). A 100% validity would result in a mean difference of 0, resulting in a relative systematic error of 0.

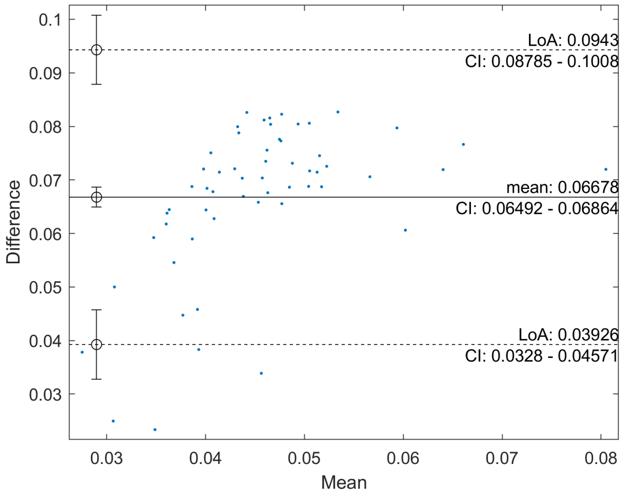

Figure 15. Bland–Altman plot shows the agreement between spatiotemporal features (horizontal deviation in the position of the vertical neck line betwen Heads and SpineBase joints) for subjects calculated using RealSense and IK analysis. The reference line at the mean is represented by the red solid line, and the top and lower limits of agreement are represented by the two black dashed lines.

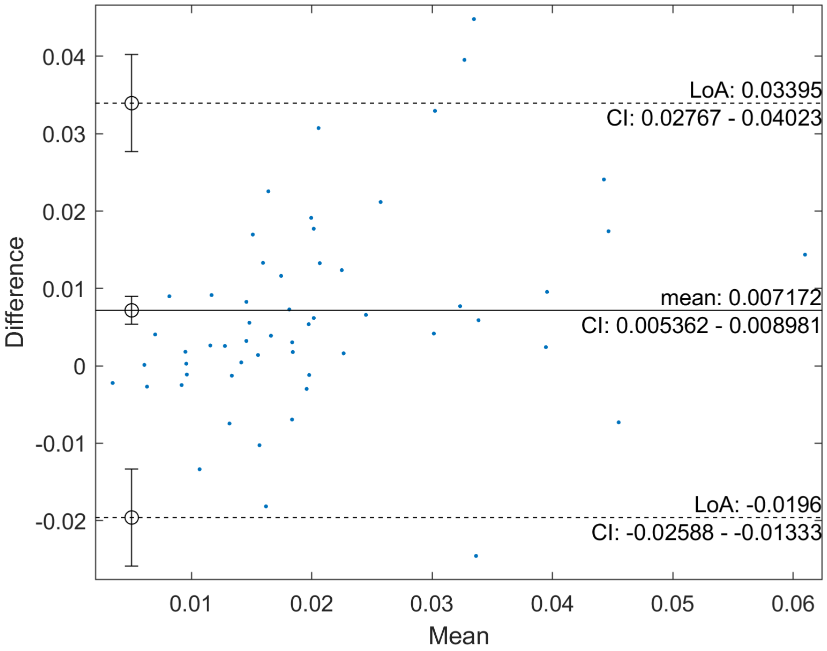

Figure 16. Bland–Altman plot shows the agreement between spatiotemporal features (vertical deviation in the position of the horizontal shoulder line between ShoulderLeft and ShoulderRight joints) for subjects calculated using RealSense and IK analysis. The reference line at the mean is represented by the solid line, and the top and lower limits of agreement are represented by the two black dashed lines.

The correlation between the changes in position between the video sequences of the frames is represented in Figure 17. the presented graph confirms the validity of position estimation using IK, while only for 7 frames out of 58, the correlation value was not significant ($p > 0.05$).

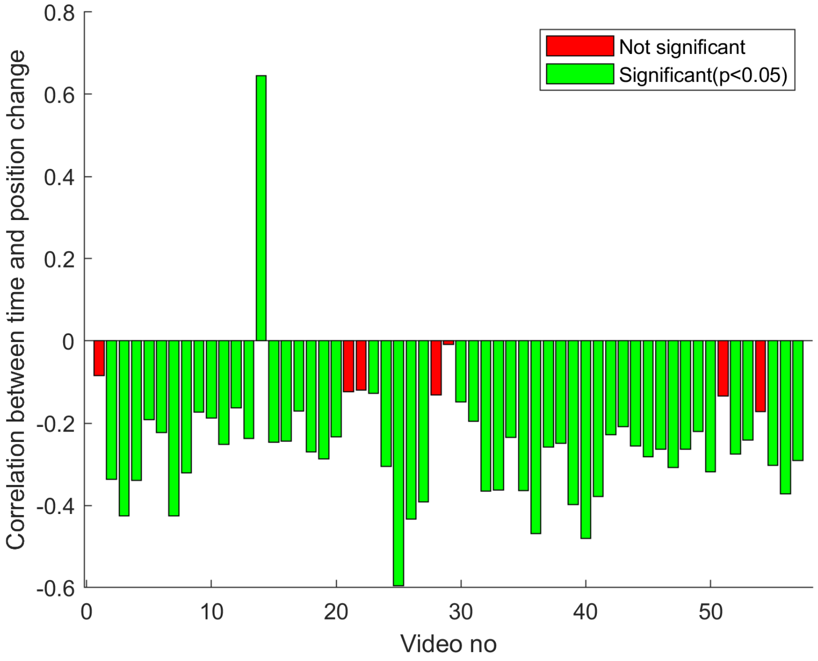

Figure 17. Correlation between position changes.

4.6. Robustness Evaluation and Crossvalidation on KIMORE Dataset

The robustness of our system was evaluated using the kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation (KIMORE) dataset [99], which includes RGB depth videos, as well as skeleton joint position and orientations, including specified characteristics, as well as physician assessment or score. We used only RGB materials from the dataset to evaluate performance. No material from this dataset was used to train our model. The KIMORE dataset includes a large diverse population of 78 participants separated into two groups. Group I (n = 34) participants with motor dysfunctions (back pain, after stroke, and Parkinson's disease) and group II (n = 44) healthy participants. Each participant completes five distinct workouts. We did not classify per these three classes as the dataset was imbalanced in this regard, the main focus was to check the accuracy of 3D recreation of each movements performed. The movements are as follows:

1. Arms being lifted;
2. A lateral tilt of the trunk with arms extended;
3. Rotation of trunks;
4. Transverse plane pelvic rotations;
5. Kneeling.

The results are displayed in Table 5 and the plot of correctly and incorrectly performed exercizes is displayed in Figure 18.

Table 5. The Pearson's correlation, CCC and ICC (with respective confidence intervals), Bland-Altman bias and LoA's, percentage error (PE), and repeatability coefficients.

Exercise	Correlation	CCC (95% CI)	ICC (95% CI)	Bias (95% LoA)	PE (%)	Repeatability (%)
Arms being lifted	0.81	0.77 (0.56–0.89)	0.87 (0.72–0.94)	−2.81 (−17.15–11.54)	23.98	13.08
Lateral tilt	0.94	0.88 (0.77–0.94)	0.93 (0.83–0.96)	−0.91 (−3.79–1.97)	13.05	15.14
Rotation of trunks	0.85	0.80 (0.62–0.90)	0.89 (0.76–0.95)	−2.24 (−13.16–8.68)	20.71	13.69
Pelvic rotations	0.84	0.80 (0.62–0.90)	0.89 (0.75–0.94)	−2.32 (−13.69–9.06)	21.15	13.61
Kneeling	0.79	0.76 (0.54–0.88)	0.86 (0.71–0.94)	−3.02 (−18.66–12.62)	25.22	12.84

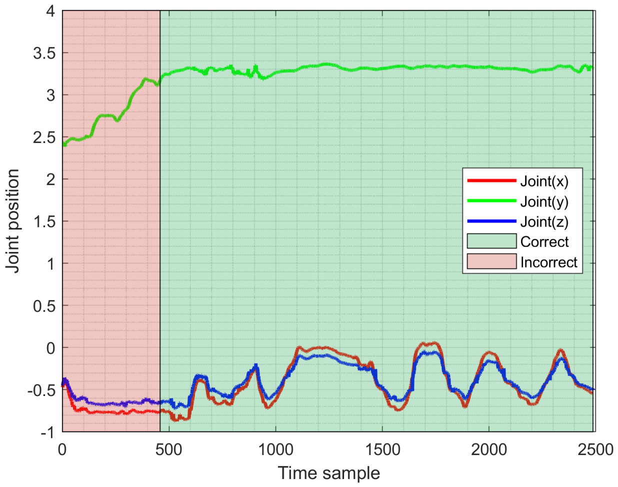

Figure 18. Plot of correctly and incorrectly performed exercises.

We tested our final model with fixed-length input after training it with variable-length data. We generate four fixed-length versions of identical test data (100, 200, 300, and 400 frames) as suggested by [100]. These fixed-length versions depict rehabilitative exercises done at various (slow/fast) speeds and repetition counts. As input, we extract the latent feature representation for variable-length and fixed-length test data.

To visualize multidimensional feature data, we use t-SNE [101]. t-SNE is a nonlinear dimension reduction method that allows high-dimensional data to be visualized by assigning each data point a location on a lower-dimensional projection. Each multidimensional sample is modeled by the t-SNE so that similar samples are mapped to nearby points, and dissimilar samples are mapped to distant points. The algorithm can capture the local structure of the dataset while revealing its global structure, such as any clusters. The 2D t-SNE plot of numerous workouts from the KIMORE dataset is shown in Figure 19. Our model gives equivalent feature representations for varying input durations, as can be shown. Based on the results of our pilot study, which compared the movements of disabled and healthy participants, it is clear that our model can successfully assess physical rehabilitation exercises/movements regardless of participant anthropometric parameters, the presence of imbalances or incorrect posture, compensatory mechanisms or slow motion. The model accurately assesses deviations from the norm or accuracy of the movement regardless of how many repetitions or how slowly the subjects performed the motion sequence.

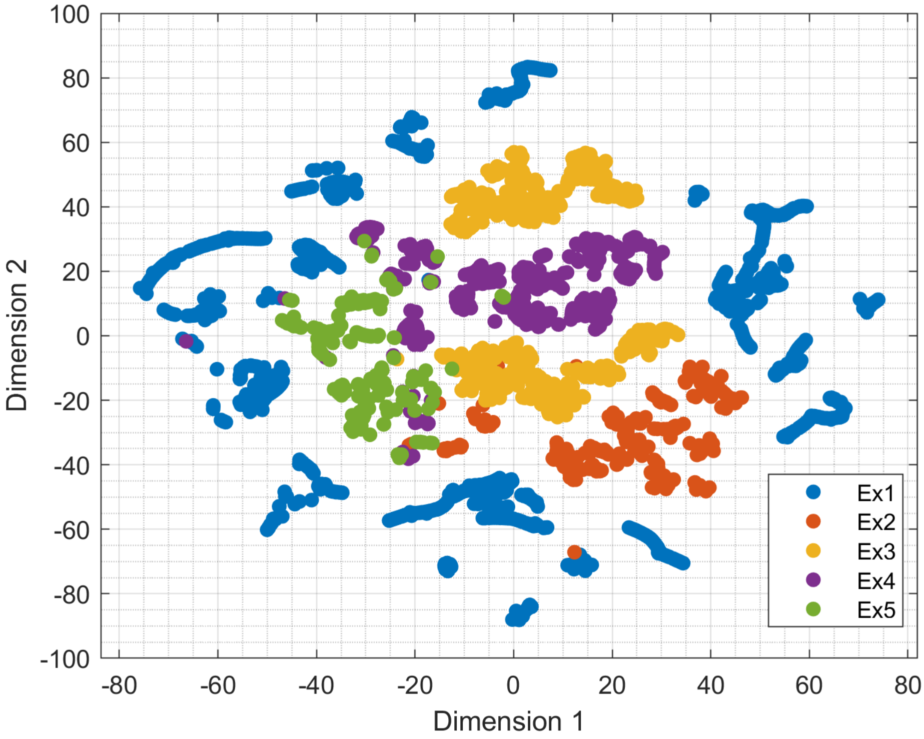

Figure 19. t-SNE plot of 5 workouts (exercises) from the KIMORE dataset.

To evaluate the reliability of our model performing on the KIMORE dataset, we used a five-fold cross-validation. Cross-validation is a statistical approach for evaluating machine learning models with a prediction aim. A round of cross-validation divides the data into complimentary subsets, with training on one set and validation testing on the other. To decrease variability, we performed five-fold cross-validation (a larger number of folds was not reasonable, considering the limited data in KIMORE).

4.7. Computation Performance

As the target solution does not require real-time performance, our solution does not require a GPU and is instead CPU-based. The average processing time achieved was 0.494 s per frame on a *Linux Mint 20.2* computer with a *Intel i5–10500* CPU and *16 GB DDR4* RAM, an improvement would be achieved by using hardware-accelerated computing. The processing time per file breakdown can be seen in Figure 20.

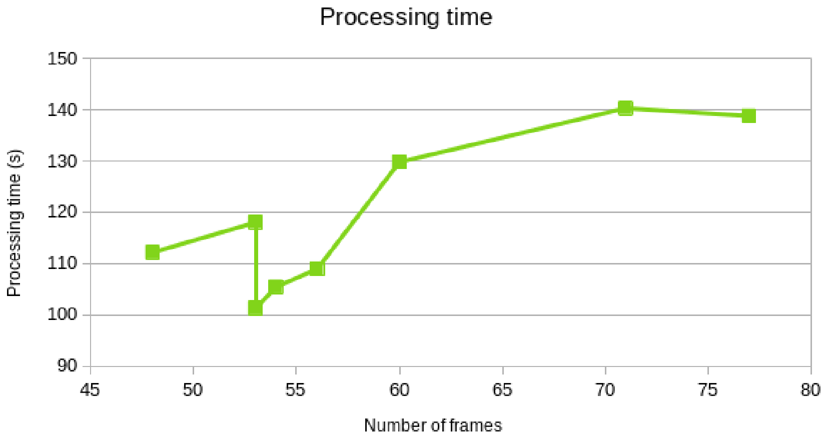

Figure 20. Processing time

4.8. User Experience Evaluation of a Tele-Rehabiliation Tool

The User Experience Questionnaire (UEQ), a validated subjective user experience (UX) measuring instrument [102], was used to interview 8 physicians and 11 potential end users (people doing exercises for our dataset, 11 responses were received from the original group of 31 participants). UEQ assesses the appeal of a tool and includes goal-directed characteristics (i.e., dependability, efficiency, perspicacity) and not goal-directed quality (i.e., stimulation, novelty).

Responses are presented in Figure 21 and Table 6. The quality levels were the following: exceptional (top 10%), good (10–25%), above average (25–50%), below average (50–75%), and terrible (bottom 25%). The results discussed above indicate that our approach may be able to satisfy the expectations of its intended user base. The survey participants noted the dependability and novelty aspects.

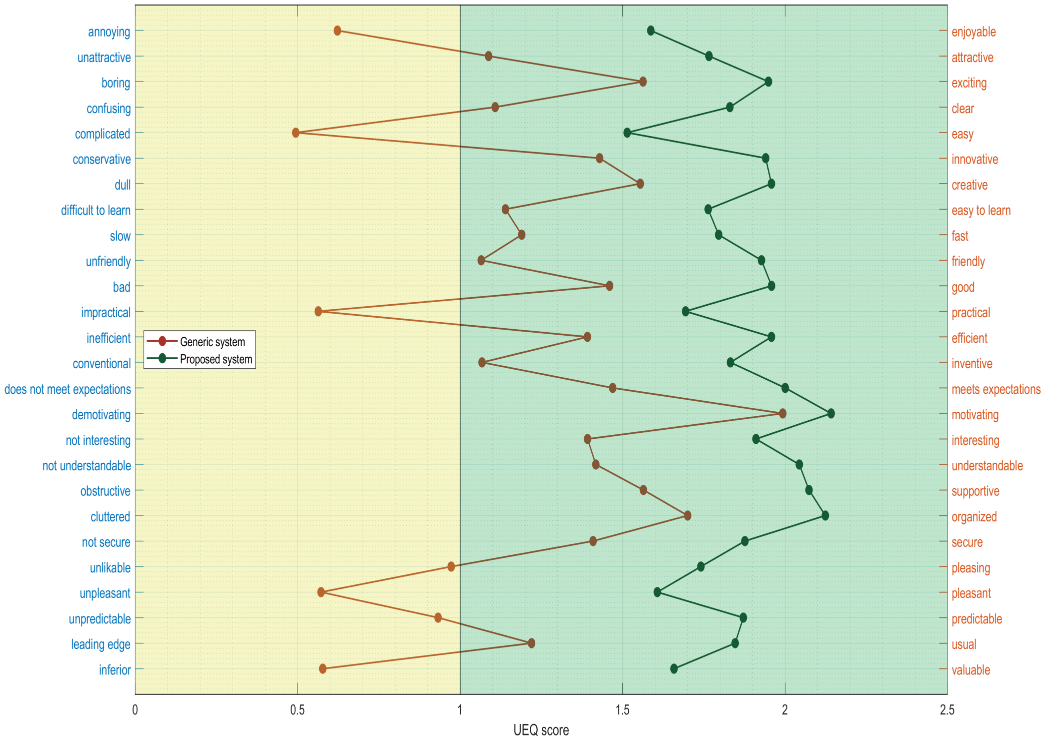

Figure 21. The results of UEQ survey.

The majority of them were enthusiastic about the efficacy of the product and eager to test it. Except for Question 9, which asks whether 3D body analysis was quick or sluggish to interact, all questions got a mean score of 5 or above, suggesting a somewhat favorable experience (4 is neutral). The findings obtained for the subscales indicate a satisfactory experience with the system. These reported results from the ongoing investigation may vary when the project is completed. Following the end of the actual prototype development, a full analysis of the various outcome measures and subjective input from the various participants will be performed. The UEQ survey allowed us to get feedback on the practical validity of such a system. Researchers delivered a brief demonstration of our technique at the beginning of the investigation. Users were advised to spend some time at the start of a session getting acquainted with the program environment.

Table 6. Summary of UEQ dimensions.

Dimension	Mean	Variance	Rating
Attractiveness	1.3887	1.4490	Above average
Perspicuity	1.4581	1.4890	Above average
Efficiency	1.3602	1.2962	Above average
Dependability	1.6233	1.4434	Good
Stimulation	1.4736	1.2523	Above average
Novelty	1.5207	1.4131	Good

4.9. Comparison to Other Approaches

The best performing methods are summarized in Table 7, comparing methods with 90 % or higher precision on the MPII dataset [54]. Our approach managed to achieve an mAP of 0.950 for the task of predicting the joint position of a single subject using the MPII dataset as a benchmark. Therefore, the proposed method achieved state-of-the-art accuracy when comparing the mAP with other existing state-of-the-art approaches. Additionally, the proposed solution allows for a seamless evaluation of posture in a 3D environment, allowing for the subjects performed to be evaluated from multiple angles from a single filmed camera feed. Further incorporation of depth sensors or multiple camera perspectives could even further improve the results when dealing with depth precision.

Table 7. Model performance using MPII dataset.

Model	Feature Introduced	Accuracy (%)
Regression-based approaches		
Zhang et al. [103]	distribution-aware coordinate representation of joints	90.60
Luvison et al. [50]	using soft-argmax operation	91.20
Detection-based approaches		
Sun et al. [104]	reducing variations of human poses statistically	91.00
Yang et al. [105]	using pyramid residuals to enhance the invariance in scales	92.00
Chen et al. [106]	structure-aware convolutional network	92.10
Ke et al. [107]	Deep conv-deconv hourglass model	92.10
Artacho et Savakis [108]	LSTM architecture for pose estimation	92.70
Nie et al. [109]	parsing induced learner to exploit parsing information	92.40
Bulat et al. [110]	gated skip connection combined with HourGlass and U-Net	94.10
Our approach	**activity-independent architecture for anthropometric regularity**	**95.0**

5. Discussion and Conclusions

Various home-based rehabilitation programs worldwide have been introduced to increase participation rates, and evidence shows that they are often an effective alternative. Unfortunately, there is no home or community rehabilitation system in Lithuania. Therefore, no user-friendly environment and smart tools have been created to be used safely in the framework of home rehabilitation. The idea of it is only gradually growing among scientists and medical staff.

The creation of such a technological system which can enable patients after stroke or trauma to independently undertake rehabilitation exercises at home is a very important part of today's health science and it is still a big challenge. The success and good results of the home training system are determined by a system that is properly and comfortably adapted to the patient: the prototype must be able to accurately detect movements when performing prescribed exercises or activities. Such data allow further analysis of the user's movements, positions, incorrect posture, and training progress. Our presented system demonstrated a very promising result, provides accurate measurements not only for these demonstrated movements of posture, but also for complex evaluation of all body movements (upper and lower limbs, incorrect posture, balance, and gait parameters); the complex movements measurements are very accurate and correct, which potentially allows patients to perform correctly selected movements as suggested by their physiotherapists. It is very important that precalculated movement patterns are correlated and matched with patient movements. Consequently, the system must estimate the perceived difficulty of complex movements by the patient, so that automatic adjustments are made to ensure a highly engaging and adaptive experience of serious movements.

Our newly developed system allows the user and rehabilitation specialist to receive biofeedback. Outside of the hospital, a biofeedback-based solution has the potential to provide feedback on exercise technique and track progress. This, in turn, can increase rehabilitation participation and improve clinical outcomes. Biofeedback allows a patient to receive real-time information about his body state to improve health or movement performance, and it provides people with useful information to help them recreate a sense of body position in space. The system can precisely assess movements of all body segments or changes in posture regardless of user anthropometric differences, motions, and mobility changes associated with dysfunction, movement repetitions, accuracy, and speed. Furthermore, real-time visual and auditory biofeedback improved motor learning while people performed specific motor tasks or exercises as instructed. Personalized cognitive and movement therapy paired with lifestyle counseling can have a greater and longer lasting impact than standard treatments, and biofeedback from movement sensors can improve results.

The overall results illustrate the efficacy of predicting human kinematics using a fully supervised learning technique in reconstructing human posture from unconstrained selfie videos as an alternative to marker-based motion capture. Our solution is supported by a Pareto-optimized inverse kinematics model, as well as dedicated joint detection models for building skeleton components. As a result, like other markerless techniques, it is free of the shortcomings of marker-based systems [111]. For example, our technique is not limited to a costly laboratory setup with regulated lighting, reflectance, and other environmental factors. The deep learning approach has achieved a near real-time performance of user action tracking by first extracting the 33 key joint features required for pose estimation and then evaluating the performed action; this approach allows for further evaluation and reporting its error using the estimated joint metrics. Additionally, the extracted joints are then passed to the IK prediction network where the predicted joint rotations can be bound to a human skeleton for further visualization. Naturally, a number of other issues exist that must be addressed in the near future, in order to further improve the robustness of the suggested approach. For example, models may generate pose candidates with symmetrical limbs or lose hands if they are concealed behind the back. Although the model output might still provide a realistic 3D pose, it may not be appropriate for thorough medical assessment. Finally, occlusions and self-occlusions provide the greatest challenge in pose estimation; however, they may be addressed in part by using a specialized model or semantic segmentation and data restoration [112,113] or by using a partial UV map to depict an object-occluded human shape [114].

One significant shortcoming of such evaluation systems is that they still need users to conduct exercises while standing (or sitting) in front of a smartphone selfie camera. Musculoskeletal sufferers may find it challenging to operate the equipment and do exercises in such a restricted environment. In such circumstances, real patients' compliance may be low. The accuracy level of all marker-less human posture estimate systems, including BIOMAC, is a crucial concern. A potential decline in accuracy is possible in uncontrolled real-world home usage, which can be large due to the broad and vastly different expressions of aberrant or impaired human motions recuperating from major health issues.

Furthermore, the system cannot forecast what will happen when the user's hands are behind his back or his torso is twisted at an angle that is not visible to the camera, and the AI technique lacks data to adjust and recompensate within the recovered 3D model, for example when measuring the angle of the knee joint, where a person may stand up with his or her side towards the camera. The model would first calculate the coordinates of the hip, knee, and ankle joints. These coordinates form a triangular construct, and after computing the Euclidean distances between all coordinates, the rule of cosines is used to estimate all corners. The accuracy with which such joint coordinates are computed is critical, since it might affect the overall representation and analysis of rehabilitation performance.

In the near future, additional in-depth research into the movement and function of more sophisticated joints, such as hips, wrists, and shoulders, will be required to increase the accuracy of 3D posture estimate in this respect. There are also certain unresolved issues and gaps between research and actual applications, such as the impact of body part occlusion and congested backgrounds of several people. Future studies ought to look at both global and local settings for more prejudicial human body characteristics, while also including human body elements into the deep learning network for antecedent limitations.

Finally, there is the issue of data variety. It is possible to collect more data for certain complicated circumstances, and there are various ways to enhance current datasets. While there is an attribute divide between artificial data and real data, GANs may potentially produce an infinite amount of data. Cross-dataset augmentation, particularly supplementing 3D datasets with 2D datasets, has the potential to alleviate the problem of inadequate diversity of training data and this is another aim on our investigation list.

Author Contributions: Conceptualization, J.G. and A.A.; Data curation, A.K., R.D. and A.A.; Formal analysis, R.M., R.D. and J.G.; Funding acquisition, R.M.; Investigation, R.M. and J.G.; Methodology, R.M., R.D. and A.A.; Project administration, R.M. and J.G.; Resources, A.K.; Software, A.K.; Supervision, R.M.; Validation, R.M., A.K. and A.A.; Visualization, A.K. and R.D.; Writing—original draft, R.M.; Writing—review and editing, R.M., R.D. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by EU Structural Funds project financed by the European Regional Development Fund under the 2014–2020 Operational Programme for the Investment of European Union Funds (2014–2020) Measure No 1.2.2-CPVA-K-703 "Promotion of Centres of Excellence and Centres for Innovation and Technology Transfer", project number 01.2.2-CPVA-K-703-03-0022.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Vilnius tech faculty committee 64-2221.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
GPU	Graphical Processing Unit
HOG	Histogram of Oriented Gradient
HSV	Hue, Saturation and Value
CNN	convolutional neural network
RGBZ	Red Green Blue Depth (Z buffer)
TOF	time-of-fligh
DH	Denavit-Hartenberg
DOF	degrees of freedom
COCO	Skeleton topology
ROI	Region-of-Intere
IMU	Inertial measurement unit
IK	Inverse Kinematics
ROM	Range of motion
MoVi	Dataset MoVi
HTC VIVE	Virtual reality equipment
OKS	Object Keypoint Similarity
mAP	Mean Average Precision
PCk	Percentage of Correct Keypoints
MPJPE	Mean Per Joint Position Error
ROME	Range of Motion Error
SG	Sprague and Geers' metric
SD	Standard Deviation
CPU	Central Processing Unit
UEQ	User Experience Questionnaire

References

1. Shaikh, T.A.; Dar, T.R.; Sofi, S. A data-centric artificial intelligent and extended reality technology in smart healthcare systems. *Soc. Netw. Anal. Min.* **2022**, *12*, 122. [CrossRef] [PubMed]
2. Nazar, P.S.; Pott, P.P. Ankle Rehabilitation Robotic Systems for domestic use—A systematic review. *Curr. Dir. Biomed. Eng.* **2022**, *8*, 65–68. [CrossRef]
3. Gelineau, A.; Perrochon, A.; Daviet, J..; Mandigout, S. Compliance with Upper Limb Home-Based Exergaming Interventions for Stroke Patients: A Narrative Review. *J. Rehabil. Med.* **2022**, *54*, jrm00325. [CrossRef] [PubMed]
4. Edwards, D.; Williams, J.; Carrier, J.; Davies, J. Technologies used to facilitate remote rehabilitation of adults with deconditioning, musculoskeletal conditions, stroke, or traumatic brain injury: An umbrella review. *JBI Evid. Synth.* **2022**, *20*, 1927–1968. [CrossRef]
5. Malleret, T.; Schwab, K. *COVID-19*; ISBN Agentur Schweiz: Zürich, Switzerland, 2020.
6. Loureiro, R.C.V.; Harwin, W.S.; Nagai, K.; Johnson, M. Advances in upper limb stroke rehabilitation: A technology push. *Med. Biol. Eng. Comput.* **2011**, *49*, 1103–1118. [CrossRef]
7. Amini Gougeh, R.; Falk, T.H. Head-Mounted Display-Based Virtual Reality and Physiological Computing for Stroke Rehabilitation: A Systematic Review. *Front. Virtual Real.* **2022**, *3*, 889271. [CrossRef]
8. Truijen, S.; Abdullahi, A.; Bijsterbosch, D.; van Zoest, E.; Conijn, M.; Wang, Y.; Struyf, N.; Saeys, W. Effect of home-based virtual reality training and telerehabilitation on balance in individuals with Parkinson disease, multiple sclerosis, and stroke: A systematic review and meta-analysis. *Neurol. Sci.* **2022**, *43*, 2995–3006. [CrossRef]
9. Avers, D. Functional Performance Measures and Assessment for Older Adults. In *Guccione's Geriatric Physical Therapy*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 137–165. [CrossRef]
10. Peretti, A.; Amenta, F.; Tayebati, S.K.; Nittari, G.; Mahdi, S.S. Telerehabilitation: Review of the State-of-the-Art and Areas of Application. *JMIR Rehabil. Assist. Technol.* **2017**, *4*, e7. [CrossRef]
11. Alexander, M., Ed. *Telerehabilitation*; Elsevier—Health Sciences Division: Philadelphia, PA, USA, 2022.
12. Stucki, G.; Zampolini, M.; Selb, M.; Ceravolo, M.G.; Delargy, M.; Donoso, E.V.; Kiekens, C.; and, N.C. European Framework of Rehabilitation Services Types: The perspective of the Physical and Rehabilitation Medicine Section and Board of the European Union of Medical Specialists. *Eur. J. Phys. Rehabil. Med.* **2019**, *55*, 411–417. [CrossRef]
13. Chen, Y.; Abel, K.T.; Janecek, J.T.; Chen, Y.; Zheng, K.; Cramer, S.C. Home-based technologies for stroke rehabilitation: A systematic review. *Int. J. Med. Inform.* **2019**, *123*, 11–22. [CrossRef]
14. Halilaj, E.; Rajagopal, A.; Fiterau, M.; Hicks, J.L.; Hastie, T.J.; Delp, S.L. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J. Biomech.* **2018**, *81*, 1–11. [CrossRef] [PubMed]

15. Nagymáté, G.; Kiss, R.M. Application of OptiTrack motion capture systems in human movement analysis. *Recent Innov. Mechatronics* **1970**, *5*, 1–9. [CrossRef]
16. Silva, N.; Zhang, D.; Kulvicius, T.; Gail, A.; Barreiros, C.; Lindstaedt, S.; Kraft, M.; Bölte, S.; Poustka, L.; Nielsen-Saines, K.; et al. The future of General Movement Assessment: The role of computer vision and machine learning—A scoping review. *Res. Dev. Disabil.* **2021**, *110*, 103854. [CrossRef] [PubMed]
17. Aliaj, K.; Henninger, H. Kinematics-vis: A Visualization Tool for the Mathematics of Human Motion. *J. Open Source Softw.* **2021**, *6*, 3490. [CrossRef] [PubMed]
18. Pogrzeba, L.; Neumann, T.; Wacker, M.; Jung, B. Analysis and Quantification of Repetitive Motion in Long-Term Rehabilitation. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1075–1085. [CrossRef] [PubMed]
19. Pizzolato, C.; Reggiani, M.; Modenese, L.; Lloyd, D.G. Real-time inverse kinematics and inverse dynamics for lower limb applications using OpenSim. *Comput. Methods Biomech. Biomed. Eng.* **2016**, *20*, 436–445. [CrossRef]
20. Ziziene, J.; Daunoraviciene, K.; Juskeniene, G.; Raistenskis, J. Comparison of kinematic parameters of children gait obtained by inverse and direct models. *PLoS ONE* **2022**, *17*, e0270423. [CrossRef]
21. Shippen, J.; May, B. BoB—Biomechanics in MATLAB. In Proceedings of the 11th International Conference Biomdlore 2016, Druskininkai, Lithuania, 20–22 October 2016; VGTU Technika: Vilnius, Lithuania, 2016. [CrossRef]
22. Hribernik, M.; Umek, A.; Tomažič, S.; Kos, A. Review of Real-Time Biomechanical Feedback Systems in Sport and Rehabilitation. *Sensors* **2022**, *22*, 3006. [CrossRef]
23. Annaswamy, T.M.; Pradhan, G.N.; Chakka, K.; Khargonkar, N.; Borresen, A.; Prabhakaran, B. Using Biometric Technology for Telehealth and Telerehabilitation. *Phys. Med. Rehabil. Clin. N. Am.* **2021**, *32*, 437–449. [CrossRef]
24. Vanagas, G.; Engelbrecht, R.; Damaševičius, R.; Suomi, R.; Solanas, A. EHealth Solutions for the Integrated Healthcare. *J. Healthc. Eng.* **2018**, *2018*, 3846892. [CrossRef]
25. Yu, X.; Xiao, B.; Tian, Y.; Wu, Z.; Liu, Q.; Wang, J.; Sun, M.; Liu, X. A Control and Posture Recognition Strategy for Upper-Limb Rehabilitation of Stroke Patients. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 6630492. [CrossRef]
26. Chen, Y.L.; Yang, I.J.; Fu, L.C.; Lai, J.S.; Liang, H.W.; Lu, L. IMU-Based Estimation of Lower Limb Motion Trajectory With Graph Convolution Network. *IEEE Sens. J.* **2021**, *21*, 24549–24557. [CrossRef]
27. Ryselis, K.; Petkus, T.; Blažauskas, T.; Maskeliūnas, R.; Damaševičius, R. Multiple Kinect based system to monitor and analyze key performance indicators of physical training. *Hum.-Centric Comput. Inf. Sci.* **2020**, *10*, 51. [CrossRef]
28. Chang, W.J.; Su, J.P.; Chen, L.B.; Hsu, C.H.; Lin, C.P.; Yang, T.C.; Chen, M.C.; Ou, Y.K. BodyTracker: A Deep Learning Based 3D Limb Trajectory Tracking System for Rehabilitation. In Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; pp. 383–384. [CrossRef]
29. Milosevic, B.; Leardini, A.; Farella, E. Kinect and wearable inertial sensors for motor rehabilitation programs at home: State of the art and an experimental comparison. *BioMed. Eng. Online* **2020**, *19*, 25. [CrossRef]
30. Fan, Y. Cerebral Infarction Rehabilitation Evaluation with Posture Analyses. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *612*, 022082. [CrossRef]
31. Mihajlovic, Z.; Popovic, S.; Brkic, K.; Cosic, K. A system for head-neck rehabilitation exercises based on serious gaming and virtual reality. *Multimed. Tools Appl.* **2017**, *77*, 19113–19137. [CrossRef]
32. Visée, R.J.; Likitlersuang, J.; Zariffa, J. An Effective and Efficient Method for Detecting Hands in Egocentric Videos for Rehabilitation Applications. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 748–755. [CrossRef]
33. Piraintorn, P.; Sa-ing, V. Stroke Rehabilitation based on Intelligence Interaction System. In Proceedings of the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June 2020; pp. 648–651. [CrossRef]
34. Lin, C.Y.; Chen, P.Y.; Bai, P.Z.; Xu, Z.X.; Liao, W.X.; Tsai, C.L. A Mechanism For Solving The Elderly Posture Problem. In Proceedings of the 2019 IEEE International Conference on Smart Cloud (SmartCloud), Tokyo, Japan, 10–12 December 2019; pp. 175–180. [CrossRef]
35. Gao, S.; He, T.; Zhang, Z.; Ao, H.; Jiang, H.; Lee, C. A Motion Capturing and Energy Harvesting Hybridized Lower-Limb System for Rehabilitation and Sports Applications. *Adv. Sci.* **2021**, *8*, 2101834. [CrossRef]
36. Passon, A.; Schauer, T.; Seel, T. Hybrid Inertial-Robotic Motion Tracking for Posture Biofeedback in Upper Limb Rehabilitation*. In Proceedings of the 2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob), Enschede, The Netherlands, 26–29 August 2018; pp. 1163–1168. [CrossRef]
37. Luvizon, D.C.; Picard, D.; Tabia, H. Multi-Task Deep Learning for Real-Time 3D Human Pose Estimation and Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 2752–2764. [CrossRef]
38. Layona, R.; Yulianto, B.; Tunardi, Y. Web based Augmented Reality for Human Body Anatomy Learning. *Procedia Comput. Sci.* **2018**, *135*, 457–464. [CrossRef]
39. Kulikajevas, A.; Maskeliunas, R.; Damaševičius, R. Detection of sitting posture using hierarchical image composition and deep learning. *PeerJ Comput. Sci.* **2021**, *7*, e442. [CrossRef] [PubMed]
40. Ogundokun, R.O.; Maskeliūnas, R.; Damaševičius, R. Human Posture Detection Using Image Augmentation and Hyperparameter-Optimized Transfer Learning Algorithms. *Appl. Sci.* **2022**, *12*, 10156. [CrossRef]
41. Li, M.; Jiang, Z.; Liu, Y.; Chen, S.; Wozniak, M.; Scherer, R.; Damasevicius, R.; Wei, W.; Li, Z.; Li, Z. Sitsen: Passive sitting posture sensing based on wireless devices. *Int. J. Distrib. Sens. Netw.* **2021**, *17*, 1–11. [CrossRef]

42. Kulikajevas, A.; Maskeliunas, R.; Damasevicius, R.; Scherer, R. Humannet—A two-tiered deep neural network architecture for self-occluding humanoid pose reconstruction. *Sensors* **2021**, *21*, 3945. [CrossRef] [PubMed]
43. Desmarais, Y.; Mottet, D.; Slangen, P.; Montesinos, P. A review of 3D human pose estimation algorithms for markerless motion capture. *Comput. Vis. Image Underst.* **2021**, *212*, 103275. [CrossRef]
44. Colyer, S.L.; Evans, M.; Cosker, D.P.; Salo, A.I.T. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sport. Med.-Open* **2018**, *4*, 24. [CrossRef]
45. Zago, M.; Luzzago, M.; Marangoni, T.; Cecco, M.D.; Tarabini, M.; Galli, M. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Front. Bioeng. Biotechnol.* **2020**, *8*, 181. [CrossRef]
46. Savadjiev, P.; Chong, J.; Dohan, A.; Vakalopoulou, M.; Reinhold, C.; Paragios, N.; Gallix, B. Demystification of AI-driven medical image interpretation: Past, present and future. *Eur. Radiol.* **2018**, *29*, 1616–1624. [CrossRef]
47. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.L.; Zhang, F.; Grundmann, M. BlazePose: On-device Real-time Body Pose tracking. *arXiv* **2020**, arXiv:2006.10204.
48. Gosztolai, A.; Günel, S.; Lobato-Ríos, V.; Pietro Abrate, M.; Morales, D.; Rhodin, H.; Fua, P.; Ramdya, P. LiftPose3D, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nat. Methods* **2021**, *18*, 975–981. [CrossRef]
49. Zhang, Z.; Tang, J.; Wu, G. Simple and Lightweight Human Pose Estimation. *arXiv* **2019**, arXiv:1911.10346.
50. Luo, Z.; Hachiuma, R.; Yuan, Y.; Kitani, K. Dynamics-Regulated Kinematic Policy for Egocentric Pose Estimation. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–14 December 2021.
51. Zhou, Y.; Dong, H.; Saddik, A.E. Learning to Estimate 3D Human Pose From Point Cloud. *IEEE Sens. J.* **2020**, *20*, 12334–12342. [CrossRef]
52. Kreiss, S.; Bertoni, L.; Alahi, A. OpenPifPaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 13498–13511. [CrossRef]
53. Liu, Y.; Xu, Y.; Li, S.b. 2-D Human Pose Estimation from Images Based on Deep Learning: A Review. In Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 25–27 May 2018; pp. 462–465. [CrossRef]
54. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 15–17 June 2014.
55. Ben Gamra, M.; Akhloufi, M.A. A review of deep learning techniques for 2D and 3D human pose estimation. *Image Vis. Comput.* **2021**, *114*, 104282. [CrossRef]
56. Tang, W.; Yu, P.; Wu, Y. Deeply Learned Compositional Models for Human Pose Estimation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
57. Wei, S.E.; Ramakrishna, V.; Kanade, T.; Sheikh, Y. Convolutional pose machines. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
58. Martinez, J.; Hossain, R.; Romero, J.; Little, J.J. A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
59. Yan, S.; Xiong, Y.; Wang, J.; Lin, D. MMSkeleton. Available online: https://github.com/open-mmlab/mmskeleton (accessed on 1 November 2022).
60. Tome, D.; Russell, C.; Agapito, L. Lifting From the Deep: Convolutional 3D Pose Estimation From a Single Image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
61. Zou, Z.; Tang, W. Modulated Graph Convolutional Network for 3D Human Pose Estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 11477–11487.
62. Pavllo, D.; Feichtenhofer, C.; Grangier, D.; Auli, M. 3D human pose estimation in video with temporal convolutions and semi-supervised training. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
63. Moon, G.; Chang, J.; Lee, K.M. V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 18–20 June 2018.
64. Martínez-González, A.; Villamizar, M.; Canévet, O.; Odobez, J.M. Residual Pose: A Decoupled Approach for Depth-Based 3D Human Pose Estimation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 24 October 2020–24 January 2021.
65. Li, J.; Xu, C.; Chen, Z.; Bian, S.; Yang, L.; Lu, C. HybrIK: A Hybrid Analytical-Neural Inverse Kinematics Solution for 3D Human Pose and Shape Estimation. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
66. Trumble, M.; Gilbert, A.; Hilton, A.; Collomosse, J. Deep Autoencoder for Combined Human Pose Estimation and Body Model Upscaling. In Proceedings of the European Conference on Computer Vision (ECCV'18), Munich, Germany, 8–14 September 2018.
67. Cheng, Y.; Wang, B.; Yang, B.; Tan, R.T. Graph and Temporal Convolutional Networks for 3D Multi-person Pose Estimation in Monocular Videos. *AAAI Conf. Artif. Intell.* **2021**, *35*, 1157–1165. [CrossRef]

68. Gärtner, E.; Pirinen, A.; Sminchisescu, C. Deep Reinforcement Learning for Active Human Pose Estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.
69. Sengupta, A.; Budvytis, I.; Cipolla, R. Hierarchical Kinematic Probability Distributions for 3D Human Shape and Pose Estimation from Images in the Wild. In Proceedings of the International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021.
70. Cheng, Y.; Yang, B.; Wang, B.; Wending, Y.; Tan, R. Occlusion-Aware Networks for 3D Human Pose Estimation in Video. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 723–732. [CrossRef]
71. Li, W.; Liu, H.; Tang, H.; Wang, P.; Van Gool, L. MHFormer: Multi-Hypothesis Transformer for 3D Human Pose Estimation. *arXiv* **2021**, arXiv:2111.12707.
72. Ma, H.; Chen, L.; Kong, D.; Wang, Z.; Liu, X.; Tang, H.; Yan, X.; Xie, Y.; Lin, S.Y.; Xie, X. TransFusion: Cross-view Fusion with Transformer for 3D Human Pose Estimation. In Proceedings of the British Machine Vision Conference, Online, 22–25 November 2021.
73. Dwivedi, S.K.; Athanasiou, N.; Kocabas, M.; Black, M.J. Learning To Regress Bodies From Images Using Differentiable Semantic Rendering. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 11250–11259.
74. Shanyan, G.; Jingwei, X.; Yunbo, W.; Bingbing, N.; Xiaokang, Y. Bilevel Online Adaptation for Out-of-Domain Human Mesh Reconstruction. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
75. Izatt, G.; Dai, H.; Tedrake, R. Globally Optimal Object Pose Estimation in Point Clouds with Mixed-Integer Programming. In *Robotics Research*; Springer Proceedings in Advanced Robotics; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 695–710. [CrossRef]
76. Tsai, C.C.; Li, W.; Hsu, K.J.; Qian, X.; Lin, Y.Y. Image Co-Saliency Detection and Co-Segmentation via Progressive Joint Optimization. *IEEE Trans. Image Process.* **2019**, *28*, 56–71. [CrossRef]
77. Liu, K.; Zhao, Y.; Nie, Q.; Gao, Z.; Chen, B.M. Weakly Supervised 3D Scene Segmentation with Region-Level Boundary Awareness and Instance Discrimination. In Proceedings of the Computer Vision—ECCV 2022, Tel Aviv, Israel, 23–27 October 2022; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2022; pp. 37–55. [CrossRef]
78. Lippi, V.; Maurer, C.; Mergner, T. Human-Likeness Indicator for Robot Posture Control and Balance. In *Communications in Computer and Information Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 98–113. [CrossRef]
79. Denavit, J.; Hartenberg, R.S. A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices. *J. Appl. Mech.* **1955**, *22*, 215–221. [CrossRef]
80. Su, H.; Enayati, N.; Vantadori, L.; Spinoglio, A.; Ferrigno, G.; Momi, E.D. Online human-like redundancy optimization for tele-operated anthropomorphic manipulators. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1–13. [CrossRef]
81. Papadopoulos, K.; Demisse, G.; Ghorbel, E.; Antunes, M.; Aouada, D.; Ottersten, B. Localized Trajectories for 2D and 3D Action Recognition. *Sensors* **2019**, *19*, 3503. [CrossRef]
82. Eng, J.J.; Pastva, A.M. Advances in Remote Monitoring for Stroke Recovery. *Stroke* **2022**, *53*, 2658–2661. [CrossRef] [PubMed]
83. Georgiadis, C.; Karvounis, E.; Koritsoglou, K.; Votis, K.; Tzovaras, D.; Dimopoulos, D.; Varvarousis, D.; Plouims, A. A remote rehabilitation training system using Virtual Reality. In Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Preveza, Greece, 24–26 September 2021; pp. 1–4. [CrossRef]
84. Hugues, A.; Marco, J.D.; Janiaud, P.; Xue, Y.; Pires, J.; Khademi, H.; Cucherat, M.; Bonan, I.; Gueyffier, F.; Rode, G. Efficiency of physical therapy on postural imbalance after stroke: study protocol for a systematic review and meta-analysis. *BMJ Open* **2017**, *7*, e013348. [CrossRef] [PubMed]
85. Peppen, R.V.; Kortsmit, M.; Lindeman, E.; Kwakkel, G. Effects of visual feedback therapy on postural control in bilateral standing after stroke: A systematic review. *J. Rehabil. Med.* **2006**, *38*, 3–9. [CrossRef]
86. Ghorbani, S.; Mahdaviani, K.; Thaler, A.; Körding, K.P.; Cook, D.J.; Blohm, G.; Troje, N.F. MoVi: A Large Multipurpose Motion and Video Dataset. *arXiv* **2020**, arXiv:2003.01888.
87. Liao, Y.; Vakanski, A.; Xian, M.; Paul, D.; Baker, R. A review of computational approaches for evaluation of rehabilitation exercises. *Comput. Biol. Med.* **2020**, *119*, 103687. [CrossRef]
88. Cramer, S.C.; Dodakian, L.; Le, V.; See, J.; Augsburger, R.; McKenzie, A.; Zhou, R.J.; Chiu, N.L.; Heckhausen, J.; Cassidy, J.M.; et al. Efficacy of Home-Based Telerehabilitation vs In-Clinic Therapy for Adults After Stroke. *JAMA Neurol.* **2019**, *76*, 9. [CrossRef] [PubMed]
89. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755.
90. Jung, H.Y.; Lee, S.; Heo, Y.S.; Yun, I.D. Random tree walk toward instantaneous 3D human pose estimation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2467–2474. [CrossRef]

91. Gu, Y.; Zhang, H.; Kamijo, S. Multi-Person Pose Estimation using an Orientation and Occlusion Aware Deep Learning Network. *Sensors* **2020**, *20*, 1593. [CrossRef]
92. Mehta, D.; Rhodin, H.; Casas, D.; Fua, P.; Sotnychenko, O.; Xu, W.; Theobalt, C. Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision. In Proceedings of the 2017 Fifth International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017. [CrossRef]
93. Torricelli, D.; Cortés, C.; Lete, N.; Bertelsen, Á.; Gonzalez-Vargas, J.E.; del Ama, A.J.; Dimbwadyo, I.; Moreno, J.C.; Florez, J.; Pons, J.L. A Subject-Specific Kinematic Model to Predict Human Motion in Exoskeleton-Assisted Gait. *Front. Neurorobot.* **2018**, *12*, 18. [CrossRef]
94. Sprague, M.A.; Geers, T.L. Spectral elements and field separation for an acoustic fluid subject to cavitation. *J. Comput. Phys.* **2003**, *184*, 149–162. [CrossRef]
95. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Int. J. Nurs. Stud.* **2010**, *47*, 931–936. [CrossRef]
96. Lin, L.I.K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **1989**, *45*, 255. [CrossRef] [PubMed]
97. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [CrossRef] [PubMed]
98. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef] [PubMed]
99. Capecci, M.; Ceravolo, M.G.; Ferracuti, F.; Iarlori, S.; Monteriu, A.; Romeo, L.; Verdini, F. The KIMORE Dataset: KInematic Assessment of MOvement and Clinical Scores for Remote Monitoring of Physical REhabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1436–1448. [CrossRef]
100. Deb, S.; Islam, M.F.; Rahman, S.; Rahman, S. Graph Convolutional Networks for Assessment of Physical Rehabilitation Exercises. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 410–419. [CrossRef]
101. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
102. Schankin, A.; Budde, M.; Riedel, T.; Beigl, M. Psychometric Properties of the User Experience Questionnaire (UEQ). In Proceedings of the CHI Conference on Human Factors in Computing Systems, ACM, New Orleans, LA, USA, 29 April–5 May 2022. [CrossRef]
103. Zhang, F.; Zhu, X.; Dai, H.; Ye, M.; Zhu, C. Distribution-Aware Coordinate Representation for Human Pose Estimation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [CrossRef]
104. Sun, K.; Lan, C.; Xing, J.; Zeng, W.; Liu, D.; Wang, J. Human Pose Estimation Using Global and Local Normalization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
105. Yang, W.; Li, S.; Ouyang, W.; Li, H.; Wang, X. Learning Feature Pyramids for Human Pose Estimation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
106. Chen, Y.; Shen, C.; Wei, X.S.; Liu, L.; Yang, J. Adversarial PoseNet: A Structure-Aware Convolutional Network for Human Pose Estimation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
107. Ke, L.; Chang, M.C.; Qi, H.; Lyu, S. Multi-Scale Structure-Aware Network for Human Pose Estimation. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 731–746. [CrossRef]
108. Artacho, B.; Savakis, A. UniPose: Unified Human Pose Estimation in Single Images and Videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
109. Nie, X.; Feng, J.; Zuo, Y.; Yan, S. Human Pose Estimation with Parsing Induced Learner. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
110. Bulat, A.; Kossaifi, J.; Tzimiropoulos, G.; Pantic, M. Toward fast and accurate human pose estimation via soft-gated skip connections. In Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 16–20 November 2020. [CrossRef]
111. Needham, L.; Evans, M.; Cosker, D.P.; Wade, L.; McGuigan, P.M.; Bilzon, J.L.; Colyer, S.L. Human Movement Science in The Wild: Can Current Deep-Learning Based Pose Estimation Free Us from The Lab? *bioRxiv* **2021**, bioRxiv:2021.04.22.440909. [CrossRef]
112. Huang, B.; Zhang, T.; Wang, Y. Object-Occluded Human Shape and Pose Estimation with Probabilistic Latent Consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *early access*, 1–16. [CrossRef]
113. Angelini, F.; Fu, Z.; Long, Y.; Shao, L.; Naqvi, S.M. 2D Pose-Based Real-Time Human Action Recognition With Occlusion-Handling. *IEEE Trans. Multimed.* **2020**, *22*, 1433–1446. [CrossRef]
114. Zhang, T.; Huang, B.; Wang, Y. Object-Occluded Human Shape and Pose Estimation From a Single Color Image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.