electronics

MDPI

Article

D-BRANN: A Distributed Bidirectional Recurrent Neural Network for Full-Body Joint Angle Estimation from Sparse Inertial Measurement Units

Kevin Niño-Tejada [1,*], Jhonathan L. Rivas-Caicedo [1], Laura Saldaña-Aristizábal [1] and Juan F. Patarroyo-Montenegro [2,*]

[1] Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00680, USA
[2] Department of Computer Science and Engineering, University of Puerto Rico, Mayaguez, PR 00680, USA
* Correspondence: kevin.nino@upr.edu (K.N.-T.); juan.patarroyo@upr.edu (J.F.P.-M.)

Abstract

Wearable Inertial Measurement Units (IMUs) enable full-body motion tracking studies beyond instrumented laboratory environments, which tend to be complex and rigid. However, current IMU-based motion tracking algorithms require complex neural network models that limit online deployment on the wearable device—or devices—due to hardware constraints. This paper explores the concept of distributing the computational load of a full-body joint-angle estimation algorithm across multiple wearable devices. We present a Distributed Bidirectional Recurrent Neural Network (D-BRANN) scheme that decomposes a sparse-IMU full-body joint angle estimator into five anatomical branches. Each branch processes one local IMU reading and a shared hip/root reference across three incremental stages: leaf position, joint position, and rotation estimation. Then, the distributed inferences of all branches are merged through a residual fusion algorithm. Compared to centralized benchmarks, D-BRANN requires around 66% fewer parameters and floating-point operations (FLOPs) per prediction. Despite this reduction, D-BRANN achieves SIP errors (mean global rotation errors of the upper arms and thighs) and whole-body angular errors that are non-inferior to those of the centralized pipeline benchmark within a 2° margin, with a SIP error of 14.88° compared to 14.95°, while its SIP error under fixed-window online inference increases moderately (18.66° compared to 16.69°). These results show that sparse-IMU pose estimation can be partitioned into substantially smaller computational components while preserving comparable offline SIP accuracy and real-time processing throughput, supporting future deployment across distributed wearable devices.

Keywords: wearable sensors; inertial measurement units; distributed neural inference; human pose estimation; joint angle estimation; motion capture

Check for updates

Academic Editor: George A. Tsihrintzis

Received: 4 October 2026
Revised: 2 November 2026
Accepted: 3 November 2026
Published: 4 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Full-body motion tracking provides quantitative information about how the body performs functional tasks and is widely used in rehabilitation, mobility monitoring, and biomechanics [1,2]. A central component of this analysis is the estimation of joint kinematics, which allows movement to be characterized beyond visual observation and provides information about how individual body segments contribute to certain activity. Optical motion-capture systems remain a reference technology for obtaining these measurements because they can reconstruct body motion with high spatial accuracy [3,4]. However, their dependence on multiple cameras, calibration procedures, and unobstructed line of sight

largely confines their operation to instrumented environments. This dependence becomes a limitation when movement must be monitored beyond the laboratory environment. These limitations have motivated the use of wearable sensing as an alternative for measuring human motion outside instrumented environments. Inertial measurement units (IMUs) are particularly suitable for this purpose because they provide body-segment acceleration and orientation without requiring external cameras [5]. Since the sensors move with the user, inertial systems can extend motion analysis to settings in which fixed optical infrastructure is unavailable or unreliable [6]. Full-body reconstruction, however, introduces a practical trade-off between the devices that must be worn and the required accuracy. Instrumenting additional body segments increases direct observability, but it also increases the amount of sensing hardware that must be placed, configured, synchronized, and processed [7]. Consequently, reducing the number of required sensors has become an important direction toward more practical and ergonomic wearable motion-capture systems.

Sparse-IMU approaches address this trade-off by reconstructing full-body motion from measurements available at only a few body segments. Reducing the sensor count improves the practicality of the sensing configuration, but it simultaneously makes pose estimation more ambiguous because most joints are not observed directly. This ambiguity is illustrated in Figure 1, where different arm configurations can produce similar measurements at the wrist-mounted IMU (S_1).

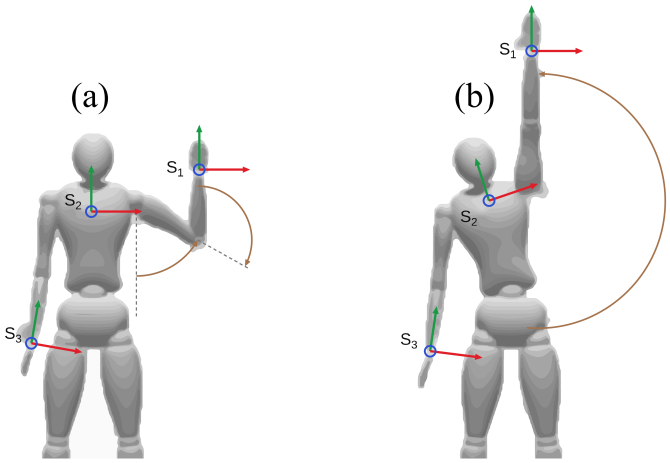


Figure 1. Illustration of pose ambiguity in sparse-IMU motion estimation. Two different upper-limb configurations can correspond to similar measurements (S_1) at the instrumented body locations, while the underlying joint angles (S_2 and S_3) differ. The local coordinate axes indicate the sensor reference frames, and the angular trajectories illustrate the change in arm configuration between poses (**a**,**b**).

Different full-body configurations can therefore be compatible with similar measurements at the instrumented segments, requiring the reconstruction method to infer information that is absent from the sensor inputs. A sparse inertial poser demonstrated that full-body pose can be reconstructed from six IMUs through optimization-based fitting [8], while Deep Inertial Poser (DIP) showed that recurrent neural networks can learn this mapping directly from sparse inertial measurements [9].

Subsequent studies have investigated the sensing configuration itself: joint-angle estimation can be performed with fewer sensors when the problem is restricted to a single limb [10], IMUPoser considers combinations of inertial sensors available in consumer devices [11], and Fürst-Walter et al. analyze the trade-off between sensor number, placement, and reconstruction accuracy [12]. Together, these works show that sparse sensing can

substantially reduce instrumentation, while shifting a larger part of the reconstruction problem to the estimation model.

As sparse sensing has matured, research has increasingly focused on improving the quality of the motion reconstructed from this limited information. One direction incorporates physical or temporal constraints into the estimation process. Physical Inertial Poser (PIP) combines learned kinematics with physics-aware optimization to reduce motion artifacts and improve temporal consistency [13], while Physical Non-inertial Poser (PNP) explicitly models the non-inertial effects introduced by an accelerating or rotating root frame [14]. Transformer Inertial Poser (TIP) uses prediction history and stationary body points to address temporal consistency and motion drift [15]. A second direction uses the structure of the human body itself to organize the learning problem. Dynamic Inertial Poser (DynaIP) divides the body into parts and models their motion dynamics separately [16], whereas HiPoser uses hierarchical learning to capture dependencies between the torso and limbs [17]. Computational cost has also been considered; Faster Deep Inertial Pose Estimation (FDIP), for example, uses simpler recurrent units while retaining a six-sensor configuration [18]. These approaches address different limitations of sparse-IMU reconstruction, including physical consistency, temporal behavior, anatomical organization, and inference efficiency.

Despite these advances, sparse sensing and distributed computation remain largely separated design problems. Most full-body sparse-IMU estimators still collect the available sensor streams and process them through a centralized computational architecture in an offline pipeline. As a result, reducing the number of sensors does not necessarily reduce the computational resources required to perform pose estimation. A system may therefore be physically sparse while still depending on a central device capable of executing the complete neural network. This distinction is particularly important when considering the deployment on resource-constrained wearable devices, where memory, computation, and power availability can be substantially more limited than on workstation or laptop-class hardware. Neural network deployment depends not only on the number of parameters, but also on the operations required during inference, memory access, and the capabilities of the target platform [19]. Reducing the computational burden of full-body pose estimation therefore requires reconsidering how the inference workload itself is organized.

Edge-computing principles provide a possible direction by moving part of the processing closer to the sources that generate the data [20]. In wearable human activity recognition, distributed frameworks have already explored processing measurements locally and exchanging reduced intermediate information instead of continuously transmitting all sensor data to a single computational node [21–23]. Extending this principle to full-body joint-angle estimation, however, presents a different challenge. Activity-recognition systems can combine local features or classification decisions, whereas a distributed pose estimator must produce continuous estimates from different body segments that ultimately describe one physically consistent full-body configuration. The estimates produced from separate parts of the body therefore cannot remain completely independent; some mechanism is required to reconcile them into a coherent pose. Online operation adds another constraint because this reconciliation must occur continuously while satisfying the timing requirements imposed by the incoming sensor stream. These considerations motivate the question of whether the pose-estimation workload itself can be partitioned across body-associated computational components without losing the reconstruction quality required for full-body motion tracking.

To investigate this question, TransPose [24] is selected as the centralized reference architecture because its organization makes the concentration of the pose-estimation workload particularly clear. TransPose reconstructs full-body pose from six IMUs through three

consecutive estimation stages: leaf-joint position, joint position, and joint rotation. Although this staged formulation progressively simplifies the reconstruction problem, the complete sequence of estimators operates as a centralized pipeline. Measurements from all six sensors are therefore provided to a single processing device, which is responsible for receiving sensor information and executing the complete pose-estimation network.

Against this background, this work proposes D-BRANN, a Distributed Bidirectional Recurrent Neural Network, which reorganizes the staged pose-estimation process into five anatomical branches corresponding to the left arm, right arm, left leg, right leg, and trunk-head. Rather than providing the complete six-sensor input to a single pose-estimation network, each branch processes the measurements from one local IMU together with a shared hip/root reference and independently performs the successive leaf-position, joint-position, and rotation-estimation stages for its assigned body segment. The resulting branch rotations are subsequently assembled and corrected through a learned residual fusion module to obtain the full-body pose. In this way, the computation that is concentrated within the centralized reference is divided into smaller body-associated estimation workloads. Figure 2 illustrates this architectural change by contrasting the centralized formulation, in which the sensor streams and complete staged estimation process are handled by a single computational node, with D-BRANN, in which the main pose-estimation workload is partitioned across five anatomical branches before residual fusion.

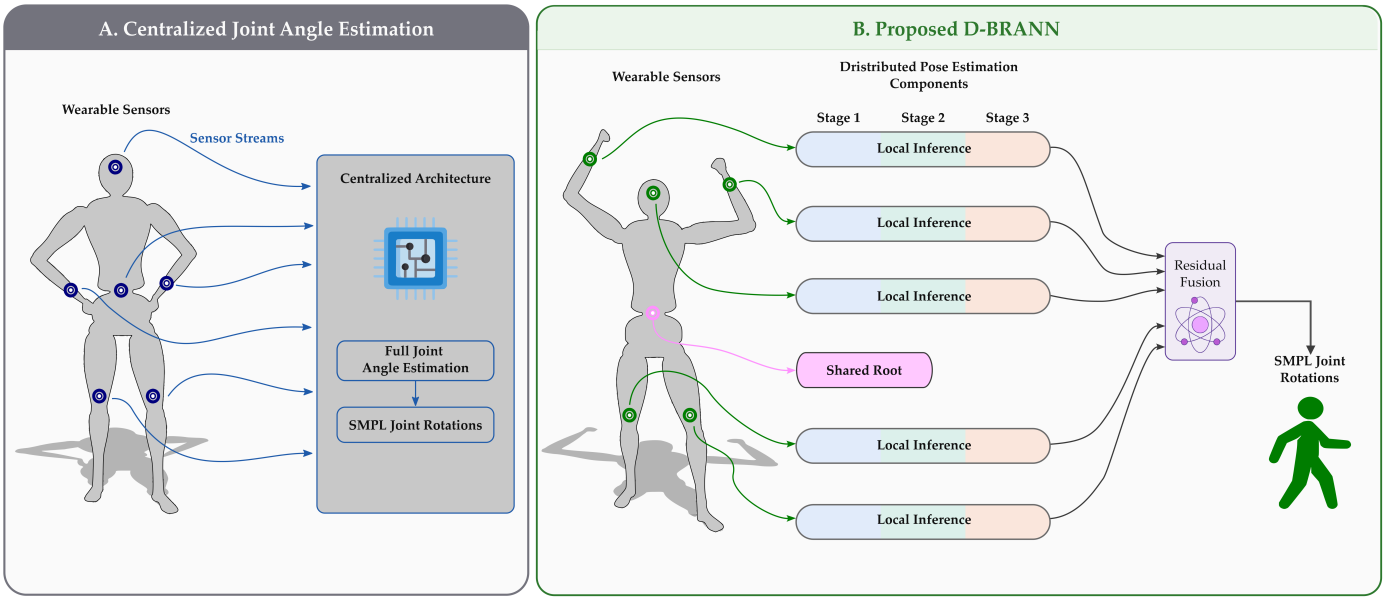


Figure 2. Conceptual comparison between centralized sparse-IMU joint-angle estimation and the proposed D-BRANN architecture. (**A**) In the centralized formulation, sensor streams are collected and processed by a single pose-estimation architecture. (**B**) D-BRANN partitions the pose-estimation workload into five anatomical branches, each using one local IMU together with a shared hip/root reference. The branch estimates are subsequently combined through learned residual fusion to obtain the final human estimation.

In this study, the proposed decomposition is evaluated on a single computational device, where the five anatomical branches are executed as parallel processes. This implementation separates the question addressed in this work from additional factors that would arise in a multi-device deployment, such as inter-node communication, synchronization, and energy consumption. The resulting evaluation therefore provides a computational and reconstruction baseline for a subsequent implementation in which individual branches can be assigned to separate body-worn processing nodes. Accordingly, the term distributed refers in this work to the partitioning of the estimation workload into independent body-

associated components, and the requirements of a physically distributed deployment are discussed in Section 4.

The main contributions of this work are summarized as follows:

- Distributed staged pose-estimation architecture: D-BRANN reorganizes a centralized three-stage sparse-IMU estimator into five anatomical branches, each operating on one local IMU and a shared hip/root reference, followed by learned residual fusion to reconstruct the complete body pose.
- Computational characterization of the distributed decomposition: The proposed architecture is analyzed in terms of parameter count, estimated floating-point operations, throughput, and latency relative to the centralized reference, providing a quantitative baseline for future deployment on resource-constrained wearable hardware.
- Offline and online evaluation of reconstruction performance: D-BRANN is evaluated under complete-sequence offline inference and fixed-window online inference to characterize the relationship between architectural decomposition, joint-angle reconstruction, positional and mesh accuracy, temporal smoothness, and real-time processing performance.
- Robustness and architectural analysis: Controlled sensor-loss and sustained node-disconnection experiments evaluate the sensitivity of the distributed formulation to missing measurements, while ablation experiments examine the contribution of residual fusion and the effect of changing the number of anatomical branches.

The remainder of this paper is organized as follows. Section 2 describes the centralized reference pipeline, the proposed D-BRANN architecture, datasets, training procedure, and evaluation protocols. Section 3 presents the offline and online reconstruction results, ablation analysis, computational performance, robustness experiments, and qualitative comparisons, together with an external evaluation on Xsens–OptiTrack recordings. Section 4 discusses the implications and limitations of the proposed decomposition, and Section 5 summarizes the conclusions and outlines future work toward deployment on physically distributed wearable hardware.

2. Materials and Methods

The methodology follows the data flow of D-BRANN from the inertial measurements to the final full-body pose. Section 2.1 describes the organization of the five anatomical branches and the shared hip/root reference. Section 2.2 details the three successive pose-estimation stages, including the objective, input, recurrent structure, and output of each stage. Section 2.2.4 describes learned residual fusion, Section 2.3 defines offline and fixed-window online inference, and the remaining subsections describe the datasets, training procedure, evaluation, and wearable validation setup.

2.1. Overview of the Distributed Architecture

D-BRANN estimates full-body joint angles from six body-mounted IMUs. Each sensor provides a three-dimensional acceleration vector and a 3×3 orientation matrix, resulting in 12 inertial values per sensor and 72 values for the complete six-sensor configuration at each frame. Five sensors are associated with the left arm, right arm, left leg, right leg, and trunk-head branches. The sixth sensor, located at the hip, does not define an additional branch; instead, it provides a shared root reference available to all five anatomical branches.

Figure 3 summarizes D-BRANN organization and distinguishes the branch-specific estimation paths from the shared root input and subsequent fusion stage. The pose-estimation component preserves the three-stage estimation order of the centralized reference while replacing each centralized estimator with five independent body-associated branches. Within each branch, the stages are executed sequentially because the output of one stage becomes

part of the input to the next: Stage 1 estimates the leaf-joint positions ($x^{(i)}_{\text{leaf}}$, $y^{(i)}_{\text{leaf}}$, and $z^{(i)}_{\text{leaf}}$, with $i = 1, \ldots, 5$), Stage 2 uses that estimate to reconstruct the joint positions assigned to the branch ($x^{(j)}_{\text{joint}}$, $y^{(j)}_{\text{joint}}$, and $z^{(j)}_{\text{joint}}$, with $j = 1, \ldots, 23$), and Stage 3 uses those positions to estimate the corresponding joint rotations ($\varphi^{(k)}_{\text{joint}}$, $\theta^{(k)}_{\text{joint}}$, and $\psi^{(k)}_{\text{joint}}$, with $k = 1, \ldots, 15$). The five branch paths remain independent until their Stage-3 outputs are assembled. A learned residual fusion module then introduces cross-branch information by correcting the assembled rotation representation. The translation-estimation modules inherited from the centralized reference remain unchanged because this work focuses on the pose-estimation component.

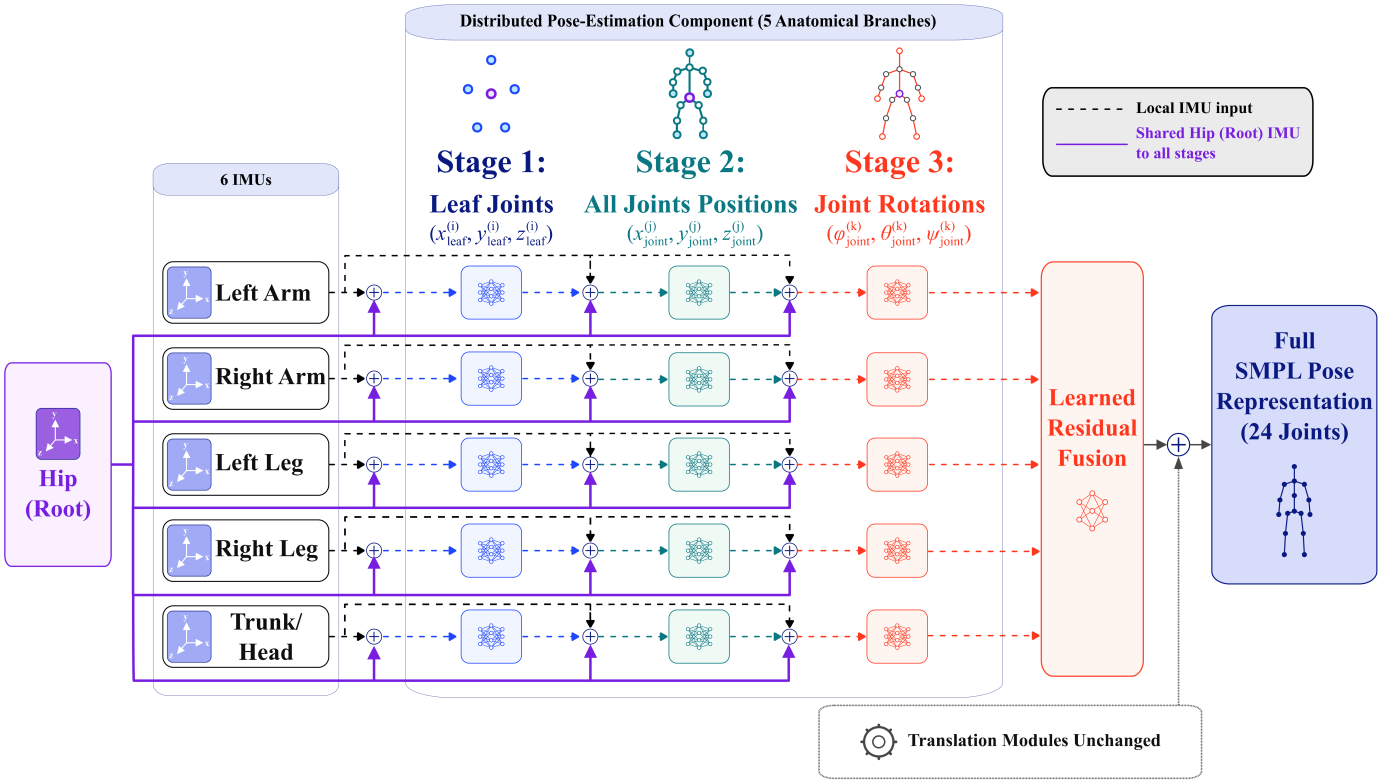


Figure 3. Overview of the D-BRANN architecture for sparse-IMU joint-angle estimation. Five anatomical branches independently perform leaf-joint position, joint-position, and joint-rotation estimation using one local IMU and the shared hip/root reference. The branch rotation outputs are assembled and corrected through learned residual fusion. The translation modules inherited from the centralized reference remain unchanged.

2.2. Distributed Staged Pose Estimation and Residual Fusion

D-BRANN preserves the staged estimation principle of the centralized reference but applies it independently to five anatomical branches. The three stages progressively transform sparse inertial measurements into increasingly pose-specific representations. Stage 1 estimates one leaf-joint position for each branch, Stage 2 expands this information into the positions of the joints assigned to that branch, and Stage 3 uses the estimated skeletal geometry to infer the corresponding joint rotations. The output of each stage therefore provides the intermediate representation required by the following stage.

The same recurrent module structure is used in all three stages. Figure 4 shows the internal structure of one stage module and defines the notation used throughout the three-stage descriptions. For an input dimension d_{in}, a linear layer initially projects the input to a recurrent feature dimension h. A two-layer bidirectional long short-term memory (BLSTM) network then models the temporal sequence in both directions. The forward and backward

hidden representations are concatenated, producing an output of dimension $2h$. A final linear layer maps this representation to the stage-specific output dimension d_{out}.

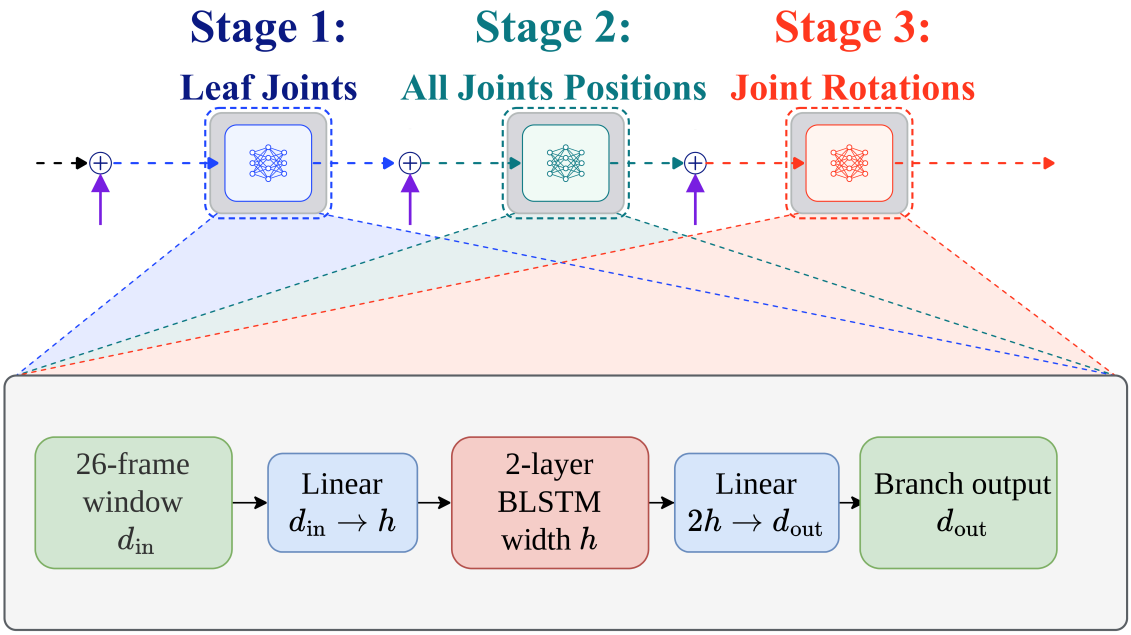


Figure 4. Internal structure of one D-BRANN stage module, corresponding to one stage block in Figure 3. A linear layer maps the stage input from d_{in} to the recurrent width h, a two-layer BLSTM produces a bidirectional representation of dimension $2h$, and a final linear layer maps this representation to the stage-specific output dimension d_{out}.

2.2.1. Stage 1: Leaf-Joint Position Estimation

The objective of Stage 1 is to obtain an initial geometric representation of the body from the sparse inertial measurements. Instead of estimating the complete skeletal configuration directly, each anatomical branch predicts the three-dimensional position of one leaf joint associated with that branch.

Stage 1 uses the recurrent structure shown in Figure 4 with $h = 32$. For each branch, the input contains the measurements from its local IMU and the shared hip/root IMU. Each of the two sensors (local and shared hip/root) contributes three acceleration values and nine orientation-matrix components, resulting in 24 values per frame. The 24-dimensional input is projected to 32 features, processed by the two-layer BLSTM, and mapped from the resulting 64-dimensional bidirectional representation to three output values.

Each branch therefore predicts its own three-dimensional leaf-joint position. Across the five branches, the Stage-1 outputs then comprise a total of 15 values. Each branch passes only its own leaf-joint estimate to its local Stage-2 network, so no information is exchanged between branches at this point.

2.2.2. Stage 2: Joint-Position Estimation

Stage 2 expands the leaf-joint estimate from each branch into a more complete representation of the local skeletal geometry. Its objective is to estimate the positions of the joints assigned to each anatomical branch before joint rotations are inferred.

Stage 2 follows the same Linear–BLSTM–Linear structure with $h = 16$. Each branch receives the same 24 inertial values used by Stage 1 together with the three-dimensional leaf-joint estimate produced by that branch, resulting in 27 values per frame. The 27-dimensional input is projected to 16 features, processed by the two-layer BLSTM, and mapped from the

resulting 32-dimensional bidirectional representation to the joint-position output of the corresponding branch.

The output dimension depends on the anatomical branch. The left- and right-leg branches each estimate four joint positions, producing $4 \times 3 = 12$ values per branch. The trunk-head, left-arm, and right-arm branches each estimate five joint positions, producing $5 \times 3 = 15$ values per branch. After assembly, Stage 2 produces a 69-dimensional representation corresponding to 23 joint positions. Each Stage-3 branch receives only the subset of these joint positions associated with its own body segment.

2.2.3. Stage 3: Joint-Rotation Estimation

Stage 3 converts the positional representation obtained in Stage 2 into the joint rotations required to reconstruct the body pose. Each anatomical branch combines its inertial measurements and shared root reference with the Stage-2 joint positions assigned to that body segment.

Stage 3 follows the same Linear–BLSTM–Linear structure shown in Figure 4, with a recurrent hidden size of $h = 16$. For each leg branch, the input contains the same 24 inertial values used in the preceding stages together with the 12 position values corresponding to four joints, resulting in 36 values per frame. For the trunk-head and arm branches, the 24 inertial values are combined with the 15 position values corresponding to five joints, resulting in 39 values per frame.

The output dimension depends on the number of rotations assigned to each anatomical branch. Each leg branch estimates two joint rotations, producing $2 \times 6 = 12$ values. The trunk-head branch estimates five rotations, producing $5 \times 6 = 30$ values, while each arm branch estimates three rotations, producing $3 \times 6 = 18$ values. After assembly, the five branch outputs form a 90-dimensional representation corresponding to 15 joint rotations.

Each joint rotation is represented using the continuous 6D rotation representation proposed in [25], in which two three-dimensional vectors encode a single rotation. The six predicted values therefore do not correspond to six independent joint angles. This representation is retained at the output of Stage 3 because the subsequent residual fusion module operates directly on the assembled 6D rotations.

2.2.4. Centralized Learned Residual Fusion

The first three stages estimate each anatomical branch independently. Once the five Stage-3 outputs are available, their rotation estimates are assembled into a centralized 90-dimensional representation corresponding to 15 joints represented in 6D. Because the branches are estimated independently, small corrections may arise between their rotation estimates when they are combined into a full-body pose.

The centralized learned residual fusion module is applied to this assembled representation to account for cross-branch relationships. The 90-dimensional input is projected to a 16-dimensional feature space and processed by a single bidirectional LSTM layer with hidden size 16. A final linear layer produces a 90-dimensional correction with the same structure as the assembled Stage-3 representation. This correction is added to the branch estimates, allowing the module to refine the combined pose while preserving the rotations independently estimated by the five anatomical branches.

The residual fusion stage is therefore the first component of the pose-estimation pipeline that jointly processes information from all five branches. After fusion, each corrected 6D rotation is converted into a valid 3×3 rotation matrix using Gram–Schmidt (GS) orthonormalization [26]. Gram–Schmidt transforms non-orthogonal vectors into mutually perpendicular unit vectors. In this case, the two three-dimensional vectors of each 6D rotation are used to construct the orthogonal axes of the rotation matrix required by the Skinned

Multi-Person Linear (SMPL) body model [27]. This conversion follows the continuous 6D rotation formulation introduced by Zhou et al. [25]. The resulting rotation matrices are then incorporated into the complete SMPL pose representation for reconstruction and subsequent evaluation.

The dimensions used throughout the distributed pose-estimation pipeline are summarized in Table 1. Each input dimension follows directly from the inertial measurements and the intermediate representation produced by the preceding stage, while the output dimensions reflect the number of joints assigned to each anatomical branch.

Table 1. Input and output dimensions of the D-BRANN pose-estimation components.

Component	Input Per Frame	Hidden Size h	Output
Stage 1	Local IMU (12) + root IMU (12) = 24	32	3 per branch; 15 assembled
Stage 2	24 inertial values + leaf position (3) = 27	16	12 per leg branch (4 joints × 3); 15 per trunk-head/arm branch (5 joints × 3); 69 assembled
Stage 3	Leg: 24 + 12 joint-position values = 36; trunk-head/arm: 24 + 15 = 39	16	12 per leg (2 rotations × 6); 30 trunk-head (5 × 6); 18 per arm (3 × 6); 90 assembled
Residual fusion	Assembled Stage-3 rotations = 90	16	90-dimensional correction applied to the assembled 6D rotations

Table 1 highlights the progressive transformation performed by D-BRANN: inertial measurements are first converted into leaf positions, then into branch-specific joint positions, and finally into 6D joint rotations. The five anatomical branches remain independent throughout these three stages. Cross-branch information is introduced only after the Stage-3 outputs are assembled and processed by the residual fusion module.

2.3. Offline and Online Inference

D-BRANN is evaluated under complete-sequence offline inference and fixed-window online inference. In both protocols, the three stages within each anatomical branch remain sequential because Stage 2 depends on the Stage-1 leaf-position estimate and Stage 3 depends on the Stage-2 joint-position estimate. During evaluation, predicted intermediate representations are retained throughout the complete inference chain; ground-truth intermediate values are not substituted.

Online inference uses a sliding 26-frame window containing 20 past frames, the current frame, and 5 future frames. The window advances with a stride of one frame, producing one new pose estimate for every newly available frame. Figure 5 illustrates the temporal organization of this input window.

The five anatomical branches are executed in parallel as independent processes on a single computational device. Each branch executes its complete Stage 1 → Stage 2 → Stage 3 sequence independently. Once the five Stage-3 outputs are available, they are assembled into the 90-dimensional rotation representation and passed to the residual fusion module. Fusion therefore provides the first point at which information from all five anatomical branches is processed jointly.

The five future frames introduce an inherent look-ahead delay independent of model execution time. At 60 Hz, this temporal context corresponds to 83.33 ms. Online latency therefore includes this algorithmic delay in addition to the computation required for branch inference and residual fusion. In this work, real-time operation refers to sustaining one pose

estimate per incoming frame at the 60 Hz acquisition rate. Incoming frames are buffered and processed in arrival order: when a prediction exceeds the 16.67 ms frame interval, subsequent frames wait in the buffer, and the resulting queueing delay, which adds to the latency, is recovered once later predictions complete within the interval. No frames are discarded by design.

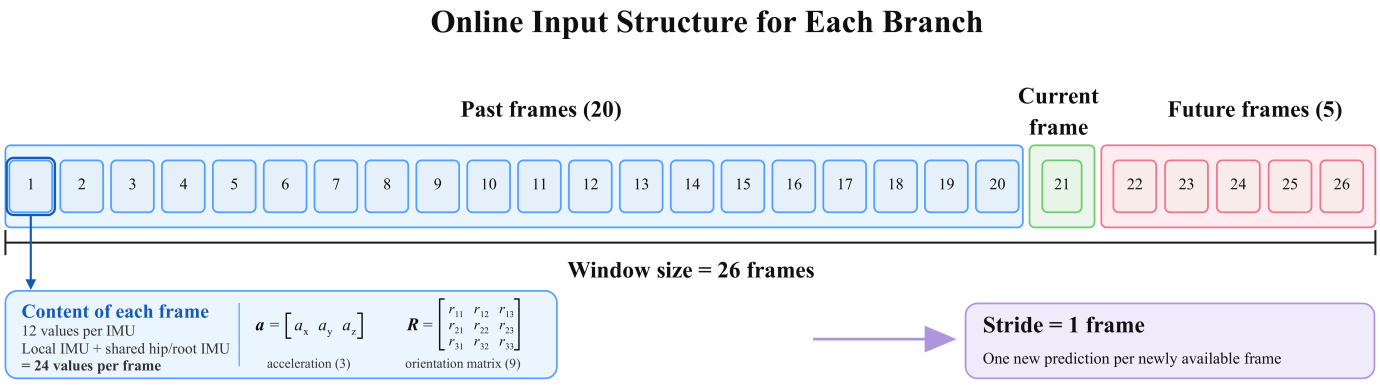


Figure 5. Temporal input structure used for fixed-window online inference. Each anatomical branch processes a 26-frame sequence containing 20 past frames, the current frame, and 5 future frames with a stride of one frame.

2.4. Datasets and Data Preparation

The pose-estimation networks are trained using the Archive of Motion Capture as Surface Shapes (AMASS) [28] and DIP-IMU [9] open-source datasets. AMASS is processed at 60 Hz, and synthetic inertial measurements are generated using the same procedure and preprocessing pipeline as TransPose [24]. Virtual IMUs are attached to the corresponding SMPL body locations, allowing sensor orientations and positions to be derived from the motion sequences. Synthetic accelerations are then obtained from the temporal displacement of the virtual sensor positions using the same smoothing procedure adopted by the reference pipeline.

DIP-IMU is used in full for training and provides real wearable inertial measurements with body-pose annotations. The same sensor ordering, normalization, coordinate transformations, and preprocessing used by TransPose are retained for both AMASS and DIP-IMU. The development dataset is split by complete sequence, with 90% used for parameter optimization and 10% for validation, checkpoint selection, and early stopping.

TotalCapture [29] is not used for training or fine-tuning and is reserved exclusively for the held-out evaluation reported in Section 3.

A separate local dataset was recorded to fine-tune D-BRANN for the Xsens-based acquisition setup and to validate the six-IMU acquisition and streaming workflow. The setup combines six Xsens MTw2 inertial sensors (Xsens Technologies B.V., Enschede, Netherlands) operating at 60 Hz with a 12-camera OptiTrack system (NaturalPoint, Inc., Corvallis, OR, USA) as the optical reference. Eleven recordings of three minutes each provide 33 min of free-form motion, including walking, sitting, squats, arm raising, long steps, and other everyday movements. For each Xsens sensor, the three-dimensional acceleration and 3×3 rotation matrix are stored at each frame.

The local recordings were collected under a protocol approved by the Institutional Review Board of the University of Puerto Rico, Mayagüez Campus (CPHSI/IRB-URPM No. 2024070003).

2.5. Stage-Wise Training Procedure

Figure 6 summarizes the stage-wise training procedure and identifies the inputs used to train each stage and the perturbations applied during training. The three pose-estimation stages are trained independently. Each stage is supervised with ground-truth quantities, so no stage requires the trained output of another stage. Stage 1 is trained to estimate branch-specific leaf-joint positions from the inertial measurements. Stage 2 is trained to estimate the joint positions assigned to each branch from the inertial measurements and the ground-truth leaf-joint positions, and Stage 3 is trained to estimate the reduced 6D joint rotations from the inertial measurements and the ground-truth joint positions. Gaussian noise is added to these ground-truth position inputs so that Stages 2 and 3 tolerate the estimation errors they receive at inference time. After the three stages have been trained, the assembled Stage-3 predictions are used to train the residual fusion module.

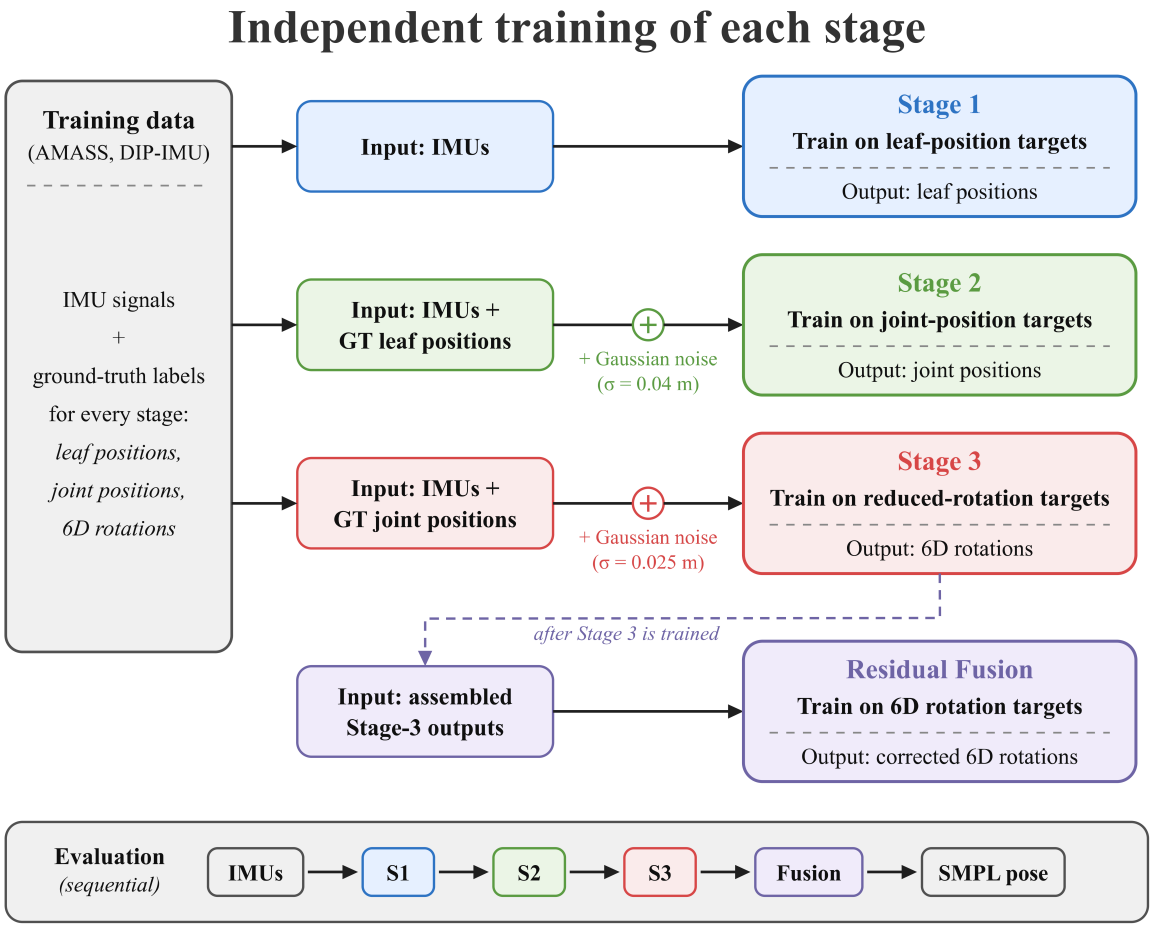


Figure 6. Stage-wise training procedure of D-BRANN. Each stage is trained independently. Stage 1 estimates branch-specific leaf-joint positions from inertial measurements, Stage 2 estimates branch joint positions from inertial measurements and ground-truth leaf-joint positions, and Stage 3 estimates 6D joint rotations from inertial measurements and ground-truth joint positions. Gaussian perturbations with $\sigma = 0.04$ m and $\sigma = 0.025$ m are applied to the ground-truth position inputs during Stage-2 and Stage-3 training, respectively. The residual fusion module is trained after Stage 3 using the assembled rotation estimates.

All estimation stages use Adam with an initial learning rate of 10^{-3}, gradient clipping at 1.0, and early stopping based on validation performance. Stage 1 uses a batch size of 128, a patience of 10 epochs, and a maximum budget of 150 epochs. Stage 2 uses a batch size of 256 and a patience of 10 epochs, while Stage 3 uses a batch size of 64 and a patience of 12 epochs; both use a maximum budget of 100 epochs. The residual fusion module

follows the Stage-3 training configuration while using the single recurrent layer described in Section 2.2.4. The recurrent widths reported in Section 2.2 were selected experimentally by progressively reducing the hidden size while monitoring the trade-off between validation error and computational cost, and then kept fixed across the corresponding experiments. All training runs use the same fixed random seed.

Gaussian noise is added to the ground-truth position inputs used to train Stages 2 and 3 to reduce sensitivity to the errors propagated from the preceding estimator at inference time. Stage 2 receives ground-truth leaf-position inputs perturbed with $\sigma = 0.04$ m. Stage 3 receives ground-truth joint-position inputs perturbed with $\sigma = 0.025$ m. These noise levels were selected after evaluating their effect on performance in the online implementation. No such perturbation is applied when training the residual fusion module. Both the recurrent widths and the noise levels were selected on the validation split.

During final evaluation, the complete prediction chain is used without ground-truth intermediate representations, so errors introduced at an earlier stage propagate naturally through the subsequent estimators.

2.6. Evaluation Metrics and Computational Characterization

Pose reconstruction is evaluated using the same metrics for offline and online inference. SIP error is the global rotation error of the upper arms and upper legs, and angular error is the global rotation error of all body joints, both in degrees. Positional error and mesh error are the mean Euclidean errors of the joint positions and of the SMPL mesh vertices, respectively, with the root joint aligned, both in centimeters. Jitter is the mean jerk of the body joints and characterizes temporal smoothness rather than accuracy; it is reported in 10^2 m/s^3.

Computational characterization considers model size and inference workload. Parameter counts are obtained directly from the implemented networks, and model size is calculated using the FP32 representation. Floating-point operations (FLOPs) are computed from the Linear and recurrent (LSTM) operations executed during one complete 26-frame online prediction, counting one multiply–accumulate operation as two FLOPs; the same convention is applied to TransPose and D-BRANN. FLOPs are calculated for the pose-estimation network, for the complete pipeline, and separately for each anatomical branch to characterize the computational workload that would be assigned to an individual processing node in a future distributed implementation.

Runtime performance is characterized using sustained throughput in predictions per second, mean processing time, and 95th- and 99th-percentile processing times. TransPose and D-BRANN are measured with the same benchmarking procedure, in which predictions are generated frame by frame with a batch size of one over the 26-frame window, after 30 warm-up predictions per sequence, for a total of 10,493 predictions per model. For these measurements, the five anatomical branch networks are executed as five separate parallel processes on the same GPU, with one process assigned to each distributed branch. Each process independently executes the complete Stage 1 → Stage 2 → Stage 3 inference path of its corresponding branch. TransPose is executed as a single process on the same devices.

Measurements are performed on a workstation equipped with an NVIDIA RTX PRO 6000 Blackwell Max-Q GPU and a consumer laptop equipped with an NVIDIA GeForce RTX 4060 Laptop GPU, both using PyTorch 2.10.0 with CUDA 13.0. Since the inertial acquisition operates at 60 Hz, 16.67 ms per frame is used as the reference processing-time budget for online operation.

2.7. Evaluation Protocol

The primary benchmark compares D-BRANN with the centralized reference under the offline and online protocols defined in Section 2.3 using the metrics in Section 2.6. Additional experiments examine how the architectural decomposition affects the accuracy–complexity trade-off and how the system responds to missing sensor measurements.

2.7.1. Architectural Ablation Design

Two architectural ablations are evaluated under offline inference. The first isolates the contribution of residual fusion by bypassing the fusion module and evaluating the assembled Stage-3 rotation representation directly. Because the branch checkpoints remain unchanged, the comparison isolates the correction introduced by residual fusion.

The second ablation evaluates the effect of anatomical partitioning by comparing the proposed five-branch architecture with three-branch and single-branch variants. Both variants are trained from scratch using the same datasets, training procedure, random seed, and recurrent widths as D-BRANN.

The three-branch model groups the peripheral sensors into legs, trunk-head, and arms. The legs branch combines the left- and right-leg sensors and estimates eight lower-limb joints. The trunk-head branch uses the head sensor and estimates five trunk and head joints. The arms branch combines the left- and right-arm sensors and estimates ten upper-limb joints. Together, these branches preserve the complete partition of the five peripheral sensors, the 23 positional joints, and the 15 reduced-rotation joints.

The single-branch variant removes the anatomical partition and processes all six IMUs through one network per stage. Its Stage-1 input therefore contains the complete 72-dimensional inertial vector. The recurrent widths remain fixed at 32 for Stage 1 and 16 for Stages 2 and 3 so that the variants are compared at identical recurrent widths. Because the total parameter count still changes with the number of branches, the comparison does not fully separate anatomical partitioning from overall network capacity.

For each partitioning strategy, Stages 1–3 and the residual fusion module are retrained using the procedure described in Section 2.5. SIP, angular, positional, mesh, and jitter metrics are computed together with parameter count.

2.7.2. Sensor-Loss Robustness

The robustness evaluation examines intermittent sensor loss and sustained node disconnection. Both protocols use zero-order hold (ZOH), which retains the most recent valid measurement whenever a new sample is unavailable. Both are evaluated under fixed-window online inference on a subset of 10 TotalCapture sequences per repetition, so the no-loss baseline reported for these experiments corresponds to this subset rather than to the complete 45-sequence benchmark. The same corrupted inputs are applied to TransPose, so that both architectures are compared under identical loss conditions.

The intermittent-loss protocol independently removes samples from each of the six sensor streams at nominal probabilities of 10%, 30%, 60%, 80%, and 95%. Each condition is repeated over 10 seeded runs. This protocol represents scattered communication or acquisition losses in which missing samples are interspersed with valid measurements.

Figure 7 illustrates the difference between scattered sample loss and sustained node disconnection. The sustained-disconnection protocol removes new measurements from one sensor node or a selected pair of nodes over contiguous intervals. Each outage window spans 10% of the corresponding recording duration, with a minimum length of one frame. Four non-overlapping outage windows are applied per recording, resulting in a total affected duration equivalent to 40% of each sequence. Defining the outage duration proportionally rather than using a fixed time interval provides a consistent protocol across TotalCapture

sequences of different lengths while allowing multiple non-overlapping interruptions to be evaluated. During each outage, ZOH retains the last valid acceleration and orientation measurements from the affected sensor. The six possible nodes correspond to the five anatomical sensors and the shared hip/root reference.



Figure 7. Sensor-failure protocols used for robustness evaluation. Intermittent loss independently removes individual sensor samples throughout the sequence, whereas sustained disconnection removes new measurements from one or more nodes over four non-overlapping intervals, each covering 10% of the recording duration. Zero-order hold retains the last valid measurement during missing intervals.

A supplementary comparison evaluates first-order hold (FOH) and second-order hold (SOH). FOH extrapolates a linear trend from preceding valid samples, whereas SOH uses quadratic extrapolation. The comparison examines whether following the recent sensor trajectory improves reconstruction relative to retaining the last valid measurement.

2.7.3. Statistical Analysis

Offline and online comparisons between D-BRANN and the centralized reference use paired per-sequence results because each TotalCapture sequence is evaluated by both methods. In addition to aggregate means and standard deviations, the paired differences are evaluated using a two-sided Wilcoxon signed-rank test [30,31]. This non-parametric test is appropriate for paired observations and examines whether the sequence-level differences between the two methods are systematically centered around zero without requiring the paired differences to follow a normal distribution. Performing the comparison at the sequence level also avoids treating temporally correlated frames from the same motion as independent observations.

Because five metrics are compared under two inference protocols on the same sequences, the ten Wilcoxon p-values are adjusted for multiple comparisons using the Holm–Bonferroni procedure [32], which controls the family-wise error rate while being uniformly more powerful than the Bonferroni correction. A non-significant Wilcoxon result, however, does not by itself establish comparable performance. SIP and angular errors are therefore also evaluated with a one-sided Wilcoxon signed-rank non-inferiority test on the paired differences (D-BRANN minus TransPose), with the null hypothesis that their median is at least 2°. This margin was chosen with reference to the 2° level commonly regarded as acceptable for kinematic measurement error in clinical gait analysis [33], rather than as a validated threshold for differences between estimators. The same adjustment is applied to these four non-inferiority tests.

For each metric, the median paired difference is reported with a 95% confidence interval estimated from 100,000 bootstrap resamples over sequences. For the robustness experiments, each seeded repetition is summarized as the mean over the same fixed subset of recordings, and 95% confidence intervals are estimated from 10,000 percentile-bootstrap resamples of these repetition means; they therefore reflect variability across corruption patterns rather than across sequences.

2.8. Per-Vertex Error Visualization

The quantitative metrics summarize reconstruction performance across complete sequences but do not show where errors occur on the body. A complementary per-vertex visualization is therefore used to examine the spatial distribution of mesh error for selected poses, following the procedure adopted by the centralized reference pipeline [24].

Figure 8 provides an example of this representation before the complete qualitative comparisons are presented in Section 3.5. For each selected frame, the reconstructed SMPL mesh is aligned at the root joint with the corresponding ground-truth mesh. The Euclidean distance between corresponding vertices is then computed, and each reconstructed vertex is colored according to its reconstruction error. A common color scale is used across D-BRANN and the centralized reference.

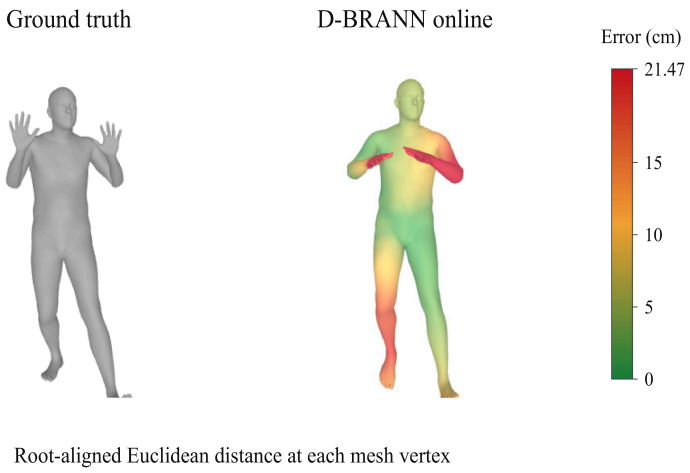


Figure 8. Example of the per-vertex error visualization used for qualitative analysis. The reconstructed SMPL mesh is root-aligned with the corresponding ground truth, and each vertex is colored according to its Euclidean reconstruction error in centimeters.

The qualitative comparison includes both complete-sequence offline inference and fixed-window online inference. Selected motions include free-form movement, walking, squats, and raised-arm configurations. Equivalent reconstructions from TransPose are generated using the same validated reference checkpoint.

2.9. Experimental Setup and Wearable Validation

The practical acquisition workflow is evaluated using six Xsens MTw2 IMUs operating at 60 Hz. The sensors follow the same six-IMU configuration used throughout the study, and each sensor provides a three-dimensional acceleration measurement together with a 3×3 rotation matrix. The live pipeline covers sensor acquisition, D-BRANN inference, residual fusion, and transmission of the reconstructed pose for visualization.

The eleven local recordings described in Section 2.4 were acquired from a single participant and are used to fine-tune D-BRANN for the Xsens-based acquisition setup and to verify continuous acquisition and inference; they are not part of the TotalCapture evaluation. Before each session, the sensors were calibrated in two steps: the root sensor was aligned with the body reference frame, and the participant then held a T-pose for 5 s. Because the Xsens and OptiTrack systems do not share a clock, each recording begins and ends with a marked gesture, and the two streams are aligned by a linear time mapping fitted to the gesture peaks detected in both modalities.

Figure 9 summarizes the experimental configuration and paired-avatar visualization. A 12-camera OptiTrack system is used simultaneously to establish the experimental basis for optical comparison under the same movements. Reflective markers are tracked with Motive while the participant wears the six Xsens sensors. The D-BRANN estimate and the corresponding optical pose drive two independently rigged avatars in a common Unity scene.

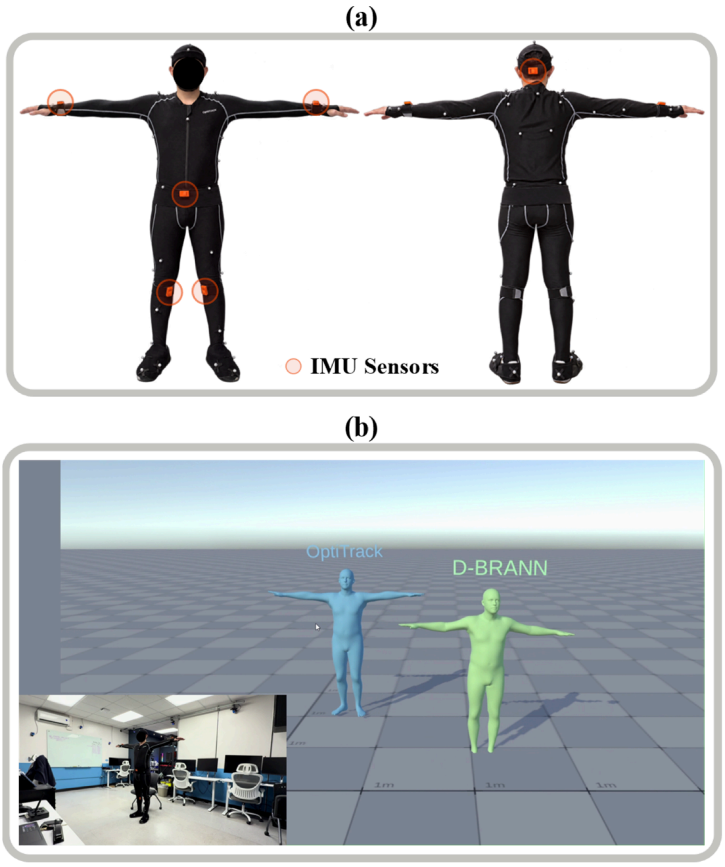


Figure 9. Experimental setup used for wearable validation and subsequent optical comparison. (**a**) Six Xsens MTw2 inertial sensors are recorded simultaneously with a 12-camera OptiTrack system. (**b**) The D-BRANN estimate obtained from the inertial measurements and the OptiTrack reference drive separate avatars in a common Unity visualization.

The optical recordings are used here to validate the acquisition and visualization workflow rather than as part of the quantitative TotalCapture benchmark. Beyond this workflow validation, the synchronized recordings also support an external quantitative assessment. Because all eleven recordings are used for fine-tuning, this assessment uses the D-BRANN and TransPose checkpoints evaluated on TotalCapture, which were trained only on AMASS and DIP-IMU and have not been exposed to Xsens data. Both models are evaluated under fixed-window online inference against an SMPL reference obtained by mapping the Motive skeleton rotations to the corresponding SMPL joints with a neutral body shape. Because this reference is derived from bone rotations rather than from a subject-specific surface fit, the assessment is intended as a feasibility evaluation on real wearable data.

3. Results

This section evaluates D-BRANN from four complementary perspectives: pose-reconstruction performance under offline and online inference, the contribution of the main architectural design choices, computational performance, and robustness to missing sensor measurements. The primary reconstruction and ablation experiments use the complete 45-sequence TotalCapture test set. Runtime is measured as described in Section 2.6, and the robustness experiments follow the repeated-subset protocol defined in Section 2.7.2. The section concludes with a qualitative analysis of the spatial distribution of reconstruction errors and an external evaluation on the Xsens–OptiTrack recordings.

3.1. Offline and Online Reconstruction Performance

TotalCapture is used as the held-out benchmark for reconstruction evaluation and is not included in model training. The complete test set contains 45 sequences and provides the common evaluation basis for D-BRANN, DIP, and the centralized TransPose reference. Tables 2 and 3 report the reconstruction metrics under complete-sequence offline inference and fixed-window online inference, respectively.

Table 2. Offline pose-reconstruction performance over the complete 45-sequence TotalCapture test set (mean ± SD). Jitter is reported in 10^2 m/s^3.

Method	SIP Err. (deg)	Ang. Err. (deg)	Pos. Err. (cm)	Mesh Err. (cm)	Jitter (10^2 m/s^3)
DIP	18.79 ± 11.85	17.77 ± 9.51	9.61 ± 5.76	11.34 ± 6.45	28.86 ± 29.18
TransPose	14.95 ± 6.90	**12.26 ± 5.59**	**5.57 ± 3.09**	**6.36 ± 3.47**	1.57 ± 2.93
D-BRANN	**14.88 ± 8.23**	13.24 ± 6.53	6.52 ± 3.72	7.46 ± 4.23	**0.98 ± 1.80**

Bold indicates the lowest value in each column.

Table 3. Fixed-window online pose-reconstruction performance over the complete TotalCapture test set (mean ± SD). Jitter is reported in 10^2 m/s^3.

Method	SIP Err. (deg)	Ang. Err. (deg)	Pos. Err. (cm)	Mesh Err. (cm)	Jitter (10^2 m/s^3)
DIP	18.93 ± 12.44	17.50 ± 10.10	9.57 ± 5.95	11.40 ± 6.87	35.94 ± 34.45
TransPose	**16.69 ± 8.79**	**12.93 ± 6.15**	**6.61 ± 3.93**	**7.49 ± 4.35**	9.44 ± 13.57
D-BRANN	18.66 ± 11.96	15.25 ± 8.54	8.93 ± 5.85	10.30 ± 6.74	**5.18 ± 7.92**

Bold indicates the lowest value in each column.

Under offline inference, D-BRANN reaches a SIP error of 14.88°, compared with 14.95° for TransPose, while producing lower jitter. TransPose retains lower angular, positional, and mesh errors, and D-BRANN reports lower values than DIP across all five metrics. These re-

sults show that the distributed decomposition closely preserves the offline SIP performance of the centralized reference while substantially reducing the pose-estimation workload.

Under the fixed-window online protocol, D-BRANN achieves a SIP error of 18.66°, compared with 16.69° for TransPose, and maintains lower jitter. The differences in the reconstruction metrics become more noticeable than in the offline setting, while D-BRANN remains below DIP across all five metrics. This comparison reflects the accuracy–smoothness trade-off introduced by restricting the bidirectional estimator to the 26-frame online context.

Aggregate means describe the overall reconstruction behavior, but they do not indicate whether the differences between D-BRANN and TransPose occur consistently across the same motion sequences. Because both methods are evaluated on the same 45 TotalCapture sequences, Tables 4 and 5 report the paired Wilcoxon signed-rank analysis described in Section 2.7.3 for offline and online inference, respectively. Differences are calculated as D-BRANN minus TransPose; positive values therefore indicate larger errors for D-BRANN, whereas negative values indicate smaller errors.

Table 4. Offline paired comparison between D-BRANN and TransPose over 45 TotalCapture sequences. Confidence intervals are obtained from 100,000 bootstrap resamples of the sequence-level differences.

Metric	Median Δ	Wilcoxon p	95% CI
SIP (deg)	−0.14	0.867	[−0.46, 0.54]
Angular (deg)	+0.91	8.0×10^{-9}	[0.50, 1.26]
Positional (cm)	+0.90	1.1×10^{-13}	[0.83, 1.14]
Mesh (cm)	+1.09	1.7×10^{-13}	[0.98, 1.17]
Jitter (10^2 m/s^3)	−0.63	5.7×10^{-14}	[−0.72, −0.42]

Table 5. Online paired comparison between D-BRANN and TransPose over 45 TotalCapture sequences. Confidence intervals are obtained from 100,000 bootstrap resamples of the sequence-level differences.

Metric	Median Δ	Wilcoxon p	95% CI
SIP (deg)	+1.72	3.8×10^{-10}	[1.46, 2.20]
Angular (deg)	+1.91	8.0×10^{-13}	[1.55, 2.07]
Positional (cm)	+1.76	5.7×10^{-14}	[1.63, 1.95]
Mesh (cm)	+2.05	5.7×10^{-14}	[1.90, 2.55]
Jitter (10^2 m/s^3)	−3.65	5.7×10^{-14}	[−4.71, −2.85]

In the offline setting, SIP shows the closest agreement between the two methods. The median paired difference is only −0.14°, the 95% confidence interval spans zero, and the Wilcoxon test does not identify a systematic difference across sequences ($p = 0.867$). Angular, positional, and mesh differences remain positive, whereas jitter shows a negative paired difference, consistent with the lower jitter observed for D-BRANN in Table 2. The non-inferiority test described in Section 2.7.3 further shows that D-BRANN is non-inferior to TransPose within the 2° margin for both SIP and angular error (Holm-adjusted $p < 10^{-6}$). All significant differences in Table 4 also remain significant after Holm–Bonferroni correction.

Under online inference, the paired SIP difference shifts to +1.72°, with a 95% confidence interval entirely above zero and a significant Wilcoxon result. Angular, positional, and mesh errors follow the same positive direction, whereas jitter remains lower for D-BRANN. The paired analysis therefore shows a clear contrast between the close offline SIP agreement and the more directional sequence-level difference observed under the restricted online window. Non-inferiority within the 2° margin is not established in this setting (Holm-

adjusted $p = 0.28$ for SIP and $p = 0.26$ for angular error), consistent with the upper limits of the confidence intervals in Table 5 exceeding the margin. All differences remain significant after Holm–Bonferroni correction.

Figure 10 provides a direct sequence-level visualization of this contrast for the SIP metric. Each point represents the SIP errors obtained by TransPose and D-BRANN on the same TotalCapture sequence, while the dashed diagonal indicates equal performance between the two methods. In the offline comparison, D-BRANN obtains lower SIP error in 24 of the 45 sequences, while TransPose obtains lower error in 21. The points are consequently distributed on both sides of the equality line, visually supporting the small median difference and the non-significant Wilcoxon result. In the online comparison, D-BRANN obtains lower SIP error in 6 sequences and TransPose in 39, producing the clearer displacement above the equality line reflected by the positive median difference in Table 5. Figure 10 therefore complements the statistical analysis by showing how the paired SIP behavior changes from a balanced offline comparison to a more directional online comparison.

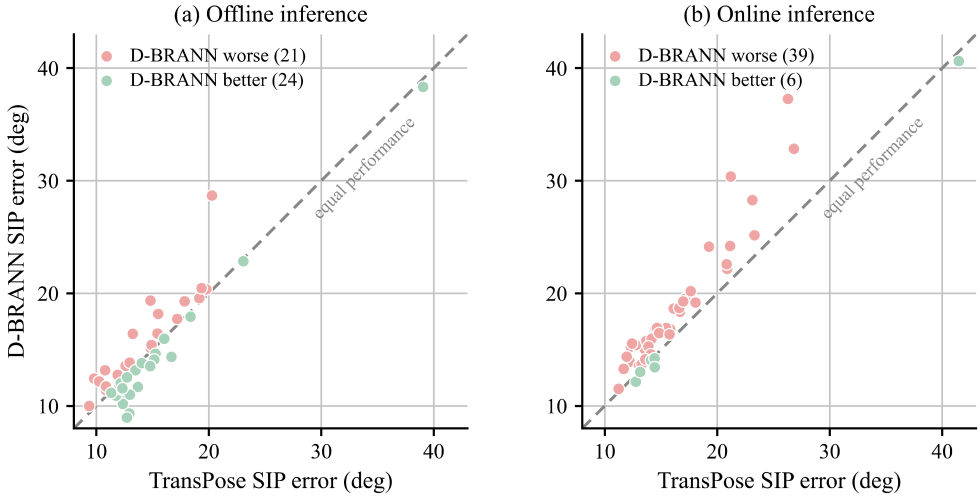


Figure 10. Per-sequence SIP error comparison between D-BRANN and TransPose under (**a**) offline and (**b**) online inference. The dashed line indicates equal SIP error.

3.2. Architectural Ablation

The ablation experiments evaluate two questions: whether learned residual fusion improves the assembled pose and how reconstruction changes as the anatomical partition is reduced. Table 6 reports the offline results for the complete five-branch architecture, the same model without residual fusion, and the independently trained three-branch and single-branch variants.

Table 6. Offline architectural ablation on the 45-sequence TotalCapture test set. All partition variants use the same recurrent widths. Positional error is also computed by the evaluation pipeline but is not included in this compact summary.

D-BRANN Variants	SIP (deg)	Angular (deg)	Mesh (cm)	Jitter (10^2 m/s^3)	Params. (M)
Five branches	**14.88**	**13.24**	**7.46**	**0.98**	0.34
Without residual fusion	15.42	14.05	8.01	0.98	0.33
Three branches	15.90	13.47	7.79	1.13	0.21
Single branch, same width	17.53	16.05	9.02	1.33	0.08

Bold indicates the lowest value for each reconstruction metric.

Bypassing residual fusion increases SIP error from 14.88° to 15.42°, angular error from 13.24° to 14.05°, and mesh error from 7.46 to 8.01 cm, while jitter remains essentially unchanged. The result indicates that residual fusion primarily improves the spatial and rotational consistency of the assembled branch estimates rather than temporal smoothness.

Reducing the number of branches produces a larger degradation. The three-branch configuration reaches 15.90° SIP error, whereas the single-branch model reaches 17.53°. The same progression is observed for angular error, mesh error, and jitter. The five-branch configuration therefore provides the lowest reconstruction errors among the tested variants at the selected recurrent widths. Because the partition variants also differ in parameter count, this progression reflects the combined effect of partitioning and model capacity.

3.3. Computational Complexity and Real-Time Performance

Table 7 compares the parameter count, computational cost, and FP32 model size of D-BRANN with the centralized reference. The comparison distinguishes the pose-estimation network, which is modified in this work, from the complete pipeline that also includes the unchanged translation modules.

Table 7. Parameter count, FLOPs per 26-frame prediction, and FP32 model size of the centralized reference and D-BRANN.

Component	TransPose	D-BRANN	Change
Pose-estimation network (parameters)	3,537,198	341,784	−90.34%
Pose-estimation network (MFLOPs)	183.16	17.20	−90.6%
Complete pipeline (parameters)	4,798,771	1,603,357	−66.59%
Complete pipeline (MFLOPs)	248.42	82.46	−66.8%
Model size (FP32, MiB)	18.31	6.12	−66.6%

The distributed pose-estimation component reduces the parameter count by 90.34% relative to the centralized pose network. Because the translation modules are unchanged, the reduction for the complete pipeline is smaller but remains 66.59%, with the FP32 model size decreasing from 18.31 to 6.12 MiB. The same pattern holds for computation: the pose-estimation network requires 90.6% fewer FLOPs per prediction than the centralized pose network (17.20 versus 183.16 MFLOPs), and the complete pipeline requires 66.8% fewer (82.46 versus 248.42 MFLOPs).

The implemented D-BRANN pose estimator requires 17.196 MFLOPs per 26-frame online prediction, of which 16.758 MFLOPs correspond to the five anatomical branches and 0.438 MFLOPs to residual fusion. Individual branches require 3.337–3.375 MFLOPs per prediction, with an average of 3.352 MFLOPs. At the 60 Hz input rate, this corresponds to approximately 0.20 GFLOP/s per branch. These per-branch values provide the computational baseline most directly relevant to a future deployment in which each anatomical branch is assigned to a separate wearable processor. In a distributed deployment, these per-branch workloads would be complemented by the components that remain centralized: residual fusion and, when global translation is required, the unchanged translation modules, which account for approximately 65.3 MFLOPs per prediction in both pipelines. Including residual fusion, the centralized workload of D-BRANN is approximately 65.7 MFLOPs, or 0.438 MFLOPs when only joint angles are estimated.

Table 8 then reports the measured runtime performance of TransPose and D-BRANN on the workstation and laptop platforms.

Table 8. Sustained throughput and computation-time percentiles of TransPose and D-BRANN for the complete 26-frame online pipelines. The 60 Hz input stream corresponds to 16.67 ms between consecutive frames.

	Workstation		Laptop	
Metric	**TransPose**	**D-BRANN**	**TransPose**	**D-BRANN**
Throughput (predictions/s)	402	119	226	91
Mean processing time (ms)	2.49	8.43	4.42	11.03
P95 processing time (ms)	2.50	9.06	5.42	12.02
P99 processing time (ms)	2.52	9.31	5.91	12.70
Predictions above 16.67 ms (%)	0	0	0	0.03

Both pipelines sustain more than 60 predictions/s on the two platforms, and their 99th-percentile processing times remain below the 16.67 ms inter-frame budget [23]; on the laptop, only 0.03% of D-BRANN predictions exceed this interval. TransPose is faster on both GPUs in this implementation. The advantage of the decomposition therefore does not lie in latency on a shared GPU but in the reduced workload assigned to each node, which determines whether each node fits an embedded processor.

The replay measurements characterize model computation on preloaded sequences and do not include the 83.33 ms look-ahead imposed by the five future frames. Adding this look-ahead to the 95th-percentile processing time of D-BRANN yields a combined look-ahead and processing time of approximately 92 ms on the workstation and 95 ms on the laptop, excluding any queueing delay. During live Xsens acquisition on the laptop, the measured P95 processing time increases to 23.12 ms, which exceeds the 16.67 ms inter-frame interval and reflects the additional acquisition and streaming workload; the corresponding combined look-ahead and processing time is approximately 106 ms, and predictions that exceed the interval are handled by the buffering described in Section 2.3. The fine-tuned D-BRANN model was also integrated into the live pipeline to drive the Unity avatar from the Xsens measurements alongside the OptiTrack visualization. This experiment verifies the end-to-end acquisition and inference workflow; the local recordings are not used for the TotalCapture benchmark reported in Section 3.1.

Because the 83.33 ms look-ahead dominates this combined time, Table 9 evaluates both models with fewer future frames, using the same checkpoints without retraining and keeping the 20 past frames, so that the window length decreases from 26 to 21 frames. The configuration with five future frames corresponds to Table 3.

Reducing the look-ahead increases the reconstruction errors of both models, and the increase is larger for D-BRANN: without future frames, its SIP error rises from 18.66° to 22.56°, compared with an increase from 16.69° to 17.72° for TransPose. D-BRANN nevertheless retains lower jitter at every look-ahead, whereas the jitter of TransPose increases from 9.44 to 14.48 as future context is removed. Because both models were trained as bidirectional estimators on 26-frame windows, these results quantify the cost of truncating their future context rather than the performance attainable by a causal design.

Table 9. Effect of the number of future frames in the online window on reconstruction performance over the complete TotalCapture test set. SIP and angular errors are reported in degrees, positional and mesh errors in cm, and jitter in 10^2 m/s^3.

		D-BRANN					TransPose				
Future Frames	**Look-Ahead (ms)**	**SIP (deg)**	**Ang. (deg)**	**Pos. (cm)**	**Mesh (cm)**	**Jitter (10^2 m/s^3)**	**SIP (deg)**	**Ang. (deg)**	**Pos. (cm)**	**Mesh (cm)**	**Jitter (10^2 m/s^3)**
5	83.33	18.66	15.25	8.93	10.30	5.18	16.69	12.93	6.61	7.49	9.44
3	50.00	19.41	15.95	9.32	10.81	5.31	16.99	13.18	6.87	7.78	10.67
1	16.67	20.90	17.28	9.98	11.59	5.82	17.42	13.62	7.30	8.28	12.90
0	0.00	22.56	18.66	10.83	12.39	6.73	17.72	13.86	7.62	8.64	14.48

3.4. Robustness to Sensor Loss

The robustness experiments examine two distinct failure modes: intermittent missing samples and sustained loss of complete sensor streams. The first determines how rapidly reconstruction degrades as isolated measurements become unavailable, while the second evaluates the effect of losing specific anatomical inputs for extended periods.

3.4.1. Intermittent Sensor Loss

Table 10 reports whole-body joint-angle error for D-BRANN and TransPose when individual samples are independently removed from the six sensor streams and replaced using zero-order hold.

Table 10. Whole-body joint-angle error under independent random sensor loss with zero-order-hold replacement for D-BRANN and TransPose. Confidence intervals are obtained by bootstrapping the ten repetition means over a fixed set of sequences.

	D-BRANN			TransPose		
Condition	**Mean Error (deg)**	**Δ vs. Baseline (deg)**	**95% CI**	**Mean Error (deg)**	**Δ vs. Baseline (deg)**	**95% CI**
Baseline	13.133	—	—	11.265	—	—
10% loss	13.183	+0.049	[13.180, 13.186]	11.302	+0.037	[11.298, 11.305]
30% loss	13.333	+0.200	[13.326, 13.341]	11.425	+0.160	[11.420, 11.432]
60% loss	13.929	+0.796	[13.913, 13.948]	11.978	+0.714	[11.962, 11.996]
80% loss	15.475	+2.341	[15.441, 15.511]	13.566	+2.301	[13.522, 13.611]
95% loss	22.046	+8.913	[21.887, 22.205]	21.134	+9.869	[20.949, 21.316]

The error changes only slightly under low and moderate random loss: 10% loss adds 0.049°, while 30% adds 0.200° relative to the no-loss baseline. The effect becomes progressively larger at 60% and 80% loss and increases sharply at 95%, where the mean error reaches 22.046°. The results therefore distinguish a low-loss region in which scattered missing measurements have limited impact from severe conditions in which reconstruction deteriorates substantially. Under identical corruption, TransPose follows the same trend, with comparable increases up to 80% loss and a larger increase at 95% loss (+9.869° versus +8.913°).

To determine whether extrapolation can improve the replacement of missing samples, Table 11 compares ZOH with first- and second-order hold for both models.

Table 11. Whole-body joint-angle error under random sensor loss using zero-, first-, and second-order hold for D-BRANN and TransPose.

	D-BRANN			TransPose		
Loss Rate	**ZOH (deg)**	**FOH (deg)**	**SOH (deg)**	**ZOH (deg)**	**FOH (deg)**	**SOH (deg)**
30%	13.333	13.228	13.286	11.425	11.407	11.562
60%	13.929	13.805	14.488	11.978	12.137	13.456
95%	22.046	28.370	34.382	21.134	29.895	40.302

FOH provides small improvements over ZOH at 30% and 60% loss, whereas SOH improves on ZOH only at 30%. At 95% loss, both extrapolation methods degrade sharply: FOH reaches 28.370° and SOH 34.382°, compared with 22.046° for ZOH. The results show that extrapolation can provide marginal gains when valid measurements remain frequent, but ZOH is substantially more stable under the most severe tested loss condition. Under the same conditions, TransPose shows a larger increase relative to its no-loss baseline than D-BRANN with both extrapolation methods at every loss rate, and at 95% loss its absolute error also exceeds that of D-BRANN (29.895° with FOH and 40.302° with SOH). ZOH therefore remains the most stable replacement strategy for both architectures under severe loss.

3.4.2. Sustained Node Disconnection

Table 12 evaluates sustained disconnection of individual sensors and selected sensor pairs for both models using the four 10%-duration outage windows defined in Section 2.7.2. The reported value is the whole-body joint-angle error, so the effect of a node failure is measured on the reconstructed pose rather than only on its associated anatomical segment.

Table 12. Whole-body joint-angle error under sustained node disconnection for D-BRANN and TransPose using zero-order hold.

	D-BRANN		TransPose	
Node(s) Disconnected	**Mean Error (deg)**	**Δ vs. Baseline (deg)**	**Mean Error (deg)**	**Δ vs. Baseline (deg)**
Baseline	13.13	—	11.26	—
Left arm	16.13	+3.00	15.04	+3.77
Right arm	16.05	+2.91	14.90	+3.63
Left leg	14.07	+0.94	12.73	+1.46
Right leg	14.27	+1.14	12.90	+1.63
Head	13.97	+0.84	12.39	+1.12
Root (hip)	18.16	+5.03	18.47	+7.21
Left arm + left leg	17.05	+3.92	16.50	+5.24
Right arm + right leg	17.17	+4.04	16.37	+5.11
Left arm + right arm	19.06	+5.93	18.84	+7.58
Left leg + right leg	15.21	+2.08	14.33	+3.07

Table 12. *Cont.*

	D-BRANN		TransPose	
Node(s) Disconnected	**Mean Error (deg)**	**Δ vs. Baseline (deg)**	**Mean Error (deg)**	**Δ vs. Baseline (deg)**
Left arm + root	18.98	+5.84	18.92	+7.66
Right arm + root	18.60	+5.47	18.56	+7.29
Left leg + root	18.43	+5.30	18.66	+7.40
Right leg + root	18.49	+5.35	18.93	+7.66
Head + root	18.16	+5.03	18.54	+7.27

The hip/root sensor is the most consequential single-node failure, increasing whole-body joint-angle error by 5.03°. This contrasts with the head and leg sensors, whose individual disconnections increase the error by approximately 0.8–1.1°. Among the evaluated pairs, simultaneous loss of both arm sensors produces the largest degradation, reaching 19.06° (+5.93°). Pairwise failures involving the hip/root sensor also consistently produce increases above 5°. These results show that sustained failures are strongly dependent on sensor location, with the shared root reference and upper-limb measurements having the largest effect among the tested conditions. Under the same outages, TransPose shows a larger increase in error than D-BRANN in all fifteen conditions. Its lower baseline keeps its absolute error below that of D-BRANN when the root sensor remains available, whereas D-BRANN reaches a lower absolute error when the root sensor is lost alone or together with the head or a leg sensor; when an arm sensor and the root are lost together, both models reach nearly identical errors.

3.5. Qualitative Visual Error Analysis

To complement the sequence-level metrics obtained with complete sensor input, Figure 11 visualizes the spatial distribution of reconstruction error for four selected TotalCapture poses corresponding to freestyle motion, walking, a squat, and an arms-raised configuration. Each example includes the ground-truth mesh together with D-BRANN and TransPose reconstructions under offline and online inference. Table 13 reports the corresponding mean and maximum vertex errors.

Table 13. Mean and maximum whole-mesh vertex errors for the selected examples (cm).

Example	D-BRANN Offline	D-BRANN Online	TransPose Offline	TransPose Online
Freestyle (S1)	**5.77/21.23**	6.94/26.34	6.03/26.28	13.17/30.71
Walking (S3)	**3.59/11.07**	6.78/26.69	4.71/11.84	7.93/19.80
Squat (S3)	**6.88/25.70**	19.33/53.35	22.71/59.87	22.35/64.88
Arms raised (S3)	10.38/32.17	13.13/41.81	**7.32/21.37**	19.43/52.36

Bold indicates the lowest mean and maximum errors for each example.

In three of the four selected poses, D-BRANN offline yields the lowest mean vertex error among the four reconstruction conditions, whereas TransPose offline is lowest for the arms-raised example. D-BRANN's online error is higher than its offline error in all four poses, with the largest difference occurring for the squat. Within these selected examples, D-BRANN also produces lower mean online mesh error than TransPose. These examples illustrate individual poses and do not alter the aggregate comparison reported in Section 3.1.

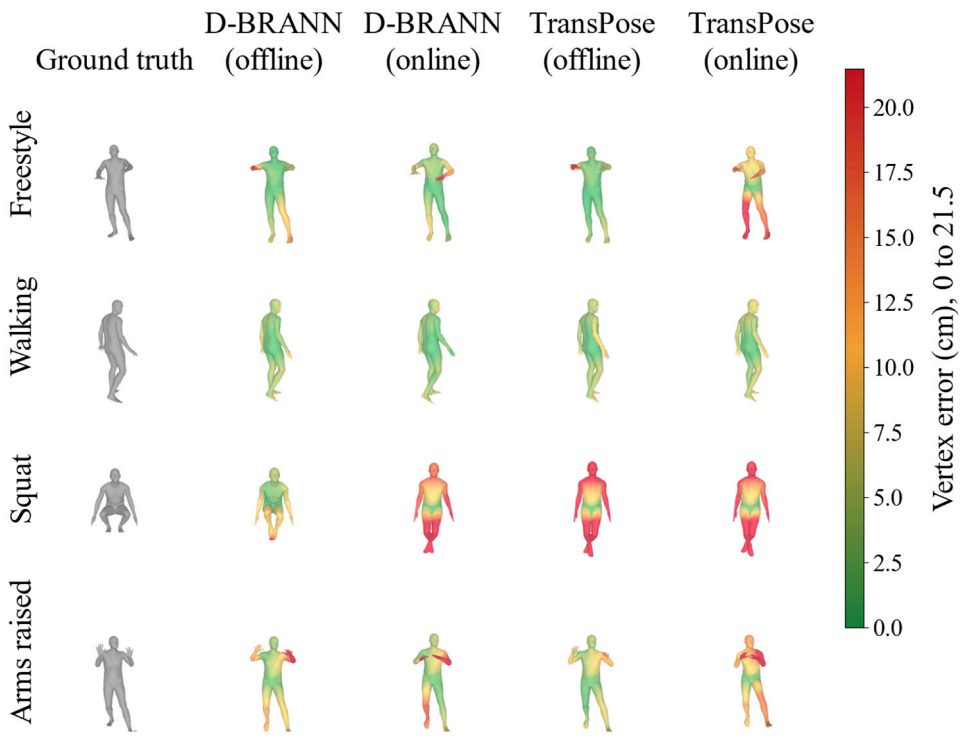


Figure 11. Ground-truth, D-BRANN, and TransPose meshes under offline and fixed-window online inference for four selected TotalCapture poses. Vertex colors represent the Euclidean distance from the root-aligned ground-truth mesh using the common scale defined in Section 2.8; errors above the upper limit of the scale are shown in the saturated color.

3.6. External Evaluation on Xsens–OptiTrack Recordings

Table 14 reports the external evaluation described in Section 2.9, in which both models are applied to the eleven Xsens recordings under fixed-window online inference.

Table 14. Fixed-window online reconstruction performance on the eleven Xsens–OptiTrack recordings using the checkpoints trained only on AMASS and DIP-IMU (mean ± SD across recordings).

Method	SIP Err. (deg)	Ang. Err. (deg)	Pos. Err. (cm)	Mesh Err. (cm)	Jitter (10^2 m/s^3)
TransPose	28.60 ± 8.43	27.52 ± 7.67	12.68 ± 3.31	15.32 ± 4.32	7.14 ± 1.37
D-BRANN	**26.91 ± 8.60**	**26.58 ± 7.53**	**10.66 ± 3.41**	**13.28 ± 4.32**	**3.27 ± 0.62**

Bold indicates the lowest value in each column.

D-BRANN yields lower SIP, positional, and mesh errors and lower jitter than TransPose in all eleven recordings, and lower angular error in ten of them. The absolute errors of both models are higher than on TotalCapture, which is consistent with the change in sensor setup and with the approximate nature of the optical reference, whose neutral body shape and bone-based mapping are common to both models. Because both models are evaluated against the same reference, these results support a relative comparison for the recorded participant rather than an absolute accuracy assessment, and they do not establish population-level generalization or isolate a benefit of anatomical partitioning.

4. Discussion

The results show that D-BRANN achieves a substantial reduction in the computational cost of pose estimation while maintaining offline SIP and angular errors close to those of the centralized reference. The pose-estimation network is reduced by approximately

90%, while offline SIP error remains closely aligned with TransPose (14.88° versus 14.95°), with no statistically significant paired difference across the 45 TotalCapture sequences. The non-inferiority analysis further shows that offline SIP and angular errors remain within a 2° margin of the centralized reference, although the paired analysis identifies systematically higher angular, positional, and mesh errors for D-BRANN, together with lower jitter. Under fixed-window online inference, the differences become more noticeable and non-inferiority within this margin is no longer established, reflecting the effect of the reduced temporal context while preserving the overall accuracy–computation trade-off targeted by the proposed architecture.

The ablation experiments help explain this behavior. Removing residual fusion increases SIP, angular, and mesh errors while leaving jitter nearly unchanged, indicating that fusion mainly corrects spatial and rotational inconsistencies between independently estimated branches. The five-branch architecture also yields lower reconstruction errors than the three- and single-branch variants. Because these variants differ in parameter count, however, the experiment does not fully separate the effect of anatomical partitioning from network capacity.

The computational reduction is most relevant for future embedded deployment. An individual branch requires 3.337–3.375 MFLOPs per 26-frame prediction, or approximately 0.20 GFLOP/s at 60 Hz, compared with approximately 1.03 GFLOP/s for the complete pose estimator including residual fusion. With approximately 66–67 k parameters, or about 0.25 MiB of FP32 weights per branch, these requirements place high-performance microcontrollers within a plausible range for per-branch execution. FLOP counts alone, however, do not capture recurrent-layer efficiency, activation memory, communication, or power consumption, so actual feasibility must be established through on-device measurements. On the evaluated GPUs, both D-BRANN and TransPose sustain more than 60 predictions/s, with TransPose running faster on a single shared GPU (Section 3.3). These measurements characterize computation only; the five future frames of the bidirectional window add a fixed 83.33 ms look-ahead at 60 Hz. Removing this look-ahead without retraining increases the reconstruction error of both models, particularly for D-BRANN (Table 9), which motivates the causal architectures discussed in Section 5.

Beyond computation, a physically distributed implementation also defines how data move between nodes. Each anatomical branch would run on a body-worn node, and the hip/root node would broadcast its 12 inertial values per frame to the five branch nodes, corresponding to 2.88 kB/s per receiving node in FP32 at 60 Hz. The branch nodes would execute their three stages locally and transmit their outputs to a central node that executes residual fusion and, when global translation is required, the translation modules, which also receive the raw inertial measurements. In the complete pipeline, the five branches produce 174 output values per frame (15 leaf-position, 69 joint-position, and 90 rotation values). With the current bidirectional window, the outputs of all 26 frames change at every step, so the branch nodes would jointly transmit approximately 1.09 MB/s of payload to the central node, excluding packet overhead; a causal formulation would reduce this to the outputs of the newest frame, approximately 42 kB/s. For joint-angle estimation alone, the translation modules are not required, and the central workload reduces to residual fusion. Communication latency, synchronization between nodes, and energy consumption remain to be characterized on physical hardware.

The robustness experiments show that both failure duration and sensor location matter. Moderate intermittent loss produces only small increases in reconstruction error, whereas sustained disconnection causes greater degradation. The hip/root sensor is the most consequential single-node failure because it provides the shared reference to all five branches, and simultaneous loss of both arm sensors produces the largest degradation among the

tested pairs. A distributed implementation should therefore detect prolonged node failures explicitly rather than rely indefinitely on held sensor values. Under identical corruption, TransPose shows a larger increase in error than D-BRANN in every sustained-disconnection condition and with first- and second-order hold, although its lower baseline keeps its absolute error below that of D-BRANN in most conditions.

The present study therefore establishes an architectural and computational baseline rather than demonstrating a physically distributed wearable system. The five branches are executed in parallel on a single device, and the software-injected faults do not reproduce wireless interference, burst losses, timing jitter, asynchronous packet arrival, synchronization errors, or heterogeneous embedded execution. In addition, all D-BRANN models were trained with a single random seed, so seed-to-seed variability is not quantified. The external Xsens–OptiTrack evaluation involves a single participant and a reference derived from the optical skeleton with a neutral body shape, so it represents a feasibility assessment rather than a full external validation. Residual synchronization uncertainty between the inertial and optical streams is also not quantified.

5. Conclusions and Future Work

D-BRANN demonstrates that full-body sparse-IMU pose estimation can be decomposed into five smaller anatomical estimation paths while retaining useful reconstruction performance. Relative to the centralized reference, the pose-estimation network is approximately 90% smaller, offline SIP and angular errors remain within a 2° non-inferiority margin, and jitter is lower under both inference protocols. Angular, positional, and mesh errors are nevertheless significantly higher, and under online inference non-inferiority is narrowly missed, as the upper confidence limits of the SIP and angular differences (2.20° and 2.07°, respectively) only slightly exceed the 2° margin. Residual fusion improves the assembled pose, and the five-branch configuration yields the lowest reconstruction errors among the tested partitions. The system tolerates moderate intermittent sensor loss but is more sensitive to sustained failures, especially of the shared hip/root reference. Under sustained disconnection, D-BRANN shows smaller error increases than TransPose, and in a single-participant Xsens–OptiTrack evaluation it yields lower errors than the centralized reference.

Future work will distribute the anatomical branches across wearable processing nodes. At 60 Hz, each branch requires approximately 0.20 GFLOP/s and 66–67 k parameters, which provides a baseline for hardware selection: high-performance microcontrollers such as the STMicroelectronics STM32H7 series (Arm Cortex-M7 core, up to 600 MHz) [34] represent a candidate platform with a tighter resource budget, whereas devices such as the STMicroelectronics STM32N6 series (Arm Cortex-M55 core at 800 MHz with an integrated neural processing unit) [35] offer additional processing and memory headroom for an initial embedded implementation. On-device experiments will measure inference latency, memory use, energy consumption, synchronization, and communication overhead. Pruning and quantization will be investigated together with temporal convolutional and causal architectures, which are more amenable to embedded acceleration and would remove the 83.33 ms look-ahead of the current bidirectional window. Parameter-matched partition variants will be evaluated to separate the effect of anatomical decomposition from model capacity. Finally, the external evaluation will be extended to multiple participants.

Author Contributions: Conceptualization, K.N.-T. and J.F.P.-M.; Methodology, K.N.-T.; Software, K.N.-T.; Validation, K.N.-T. and J.F.P.-M.; Formal analysis, K.N.-T.; Investigation, K.N.-T., J.L.R.-C., L.S.-A. and J.F.P.-M.; Resources, J.F.P.-M.; Data curation, K.N.-T.; Writing—original draft, K.N.-T.; Writing—review and editing, L.S.-A., J.L.R.-C. and J.F.P.-M.; Visualization, K.N.-T.; Supervision, J.F.P.-M.; Funding acquisition, J.F.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received financial support from the NSF CAREER project "Intelligent Biomarker Analysis based on Wearable Distributed Computing" under Grant No. OAC-2439345.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the University of Puerto Rico's Institutional Review Board (IRB), protocol CPHSI/IRB-URPM No. 2024070003 (approved on 27 January 2025).

Data Availability Statement: The Xsens–OptiTrack dataset is available on Zenodo (https://zenodo.org/records/22842322) (accessed on 1 November 2026) [36]. The D-BRANN source code for reproducing the experiments is available at GitHub D-BRANN (https://github.com/WearNET/D-BRANN) (accessed on 1 November 2026), with accompanying documentation at wearnet.github.io/D-BRANN/ (https://wearnet.github.io/D-BRANN/) (accessed on 1 November 2026).

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

Abbreviations

The following abbreviations are used in this manuscript:

IMU	Inertial Measurement Unit
D-BRANN	Distributed Bidirectional Recurrent Neural Network
BLSTM	Bidirectional Long Short-Term Memory
AMASS	Archive of Motion Capture as Surface Shapes
SMPL	Skinned Multi-Person Linear (body model)
FLOPs	Floating-Point Operations
FP32	32-bit Floating-Point Format
ZOH	Zero-Order Hold
FOH	First-Order Hold
SOH	Second-Order Hold
GPU	Graphics Processing Unit
CUDA	Compute Unified Device Architecture
CI	Confidence Interval
IRB	Institutional Review Board

References

1. Iosa, M.; Picerno, P.; Paolucci, S.; Morone, G. Wearable inertial sensors for human movement analysis. *Expert Rev. Med. Devices* **2016**, *13*, 641–659. [CrossRef] [PubMed]
2. Fong, D.T.-P.; Chan, Y.-Y. The Use of Wearable Inertial Motion Sensors in Human Lower Limb Biomechanics Studies: A Systematic Review. *Sensors* **2010**, *10*, 11556–11565. [CrossRef] [PubMed]
3. Conconi, M.; Pompili, A.; Sancisi, N.; Parenti-Castelli, V. Quantification of the errors associated with marker occlusion in stereophotogrammetric systems and implications on gait analysis. *J. Biomech.* **2021**, *114*, 110162. [CrossRef] [PubMed]
4. Aurand, A.M.; Dufour, J.S.; Marras, W.S. Accuracy map of an optical motion capture system with 42 or 21 cameras in a large measurement volume. *J. Biomech.* **2017**, *58*, 237–240. [CrossRef] [PubMed]
5. Lopez-Nava, I.H.; Munoz-Melendez, A. Wearable Inertial Sensors for Human Motion Analysis: A Review. *IEEE Sens. J.* **2016**, *16*, 7821–7834. [CrossRef]
6. Seel, T.; Raisch, J.; Schauer, T. IMU-Based Joint Angle Measurement for Gait Analysis. *Sensors* **2014**, *14*, 6891–6909. [CrossRef] [PubMed]
7. Pacher, L.; Chatellier, C.; Vauzelle, R.; Fradet, L. Sensor-to-Segment Calibration Methodologies for Lower-Body Kinematic Analysis with Inertial Sensors: A Systematic Review. *Sensors* **2020**, *20*, 3322. [CrossRef] [PubMed]
8. von Marcard, T.; Rosenhahn, B.; Black, M.J.; Pons-Moll, G. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Comput. Graph. Forum* **2017**, *36*, 349–360. [CrossRef]
9. Huang, Y.; Kaufmann, M.; Aksan, E.; Black, M.J.; Hilliges, O.; Pons-Moll, G. Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time. *ACM Trans. Graph.* **2018**, *37*, 185. [CrossRef]
10. Niño-Tejada, K.; Saldaña-Aristizábal, L.; Rivas-Caicedo, J.L.; Patarroyo-Montenegro, J.F. Upper Limb Joint Angle Estimation Using a Reduced Number of IMU Sensors and Recurrent Neural Networks. *Electronics* **2025**, *14*, 3039. [CrossRef]

11. Mollyn, V.; Arakawa, R.; Goel, M.; Harrison, C.; Ahuja, K. IMUPoser: Full-Body Pose Estimation using IMUs in Phones, Watches, Earbuds. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023. [CrossRef]
12. Fürst-Walter, I.; Nappi, A.; Harbaum, T.; Becker, J. Design Space Exploration on Efficient and Accurate Human Pose Estimation from Sparse IMU-Sensing. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023. [CrossRef]
13. Yi, X.; Zhou, Y.; Habermann, M.; Shimada, S.; Golyanik, V.; Theobalt, C.; Xu, F. Physical Inertial Poser (PIP): Physics-Aware Real-Time Human Motion Tracking from Sparse Inertial Sensors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 13167–13178.
14. Yi, X.; Zhou, Y.; Xu, F. Physical Non-inertial Poser (PNP): Modeling Non-inertial Effects in Sparse-inertial Human Motion Capture. In Proceedings of the SIGGRAPH 2024 Conference Papers, Denver, CO, USA, 28 July–1 August 2024. [CrossRef]
15. Jiang, Y.; Ye, Y.; Gopinath, D.; Won, J.; Winkler, A.W.; Liu, C.K. Transformer Inertial Poser: Real-time Human Motion Reconstruction from Sparse IMUs with Simultaneous Terrain Generation. In Proceedings of the SIGGRAPH Asia 2022, New York, NY, USA, 6–9 December 2022. [CrossRef]
16. Zhang, Y.; Xia, S.; Chu, L.; Yang, J.; Wu, Q.; Pei, L. Dynamic Inertial Poser (DynaIP): Part-Based Motion Dynamics Learning for Enhanced Human Pose Estimation with Sparse Inertial Sensors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 1889–1899. [CrossRef]
17. Liao, G.; Zheng, C.; Cheng, L.; Xie, H.; Huang, S.; Liao, J.; Li, H.; Liu, L. HiPoser: 3D Human Pose Estimation with Hierarchical Shared Learning at Parts-Level Using Inertial Measurement Units. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 14309–14317. [CrossRef]
18. Xia, D.; Zhu, Y.; Zhang, H. Faster Deep Inertial Pose Estimation with Six Inertial Sensors. *Sensors* **2022**, *22*, 7144. [CrossRef] [PubMed]
19. Sze, V.; Chen, Y.-H.; Yang, T.-J.; Emer, J.S. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proc. IEEE* **2017**, *105*, 2295–2329. [CrossRef]
20. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge Computing: Vision and Challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646. [CrossRef]
21. Rivas-Caicedo, J.L.; Niño-Tejada, K.; Saldaña-Aristizábal, L.; Patarroyo-Montenegro, J.F. A Distributed Wearable Computing Framework for Human Activity Classification. *Electronics* **2025**, *14*, 3203. [CrossRef]
22. Baghezza, R.; Bouchard, K.; Bouzouane, A.; Gouin-Vallerand, C. From Offline to Real-Time Distributed Activity Recognition in Wireless Sensor Networks for Healthcare: A Review. *Sensors* **2021**, *21*, 2786. [CrossRef] [PubMed]
23. Rivas-Caicedo, J.L.; Saldaña-Aristizábal, L.; Niño-Tejada, K.; Patarroyo-Montenegro, J.F. A Real-Time Communication Framework for Distributed Wearable Human Activity Recognition. *Electronics* **2026**, *15*, 3714. [CrossRef]
24. Yi, X.; Zhou, Y.; Xu, F. TransPose: Real-time 3D human translation and pose estimation with six inertial sensors. *ACM Trans. Graph.* **2021**, *40*, 86. [CrossRef]
25. Zhou, Y.; Barnes, C.; Lu, J.; Yang, J.; Li, H. On the Continuity of Rotation Representations in Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 5745–5753. [CrossRef]
26. Leon, S.J.; Björck, Å.; Gander, W. Gram–Schmidt orthogonalization: 100 years and more. *Numer. Linear Algebra Appl.* **2013**, *20*, 492–532. [CrossRef]
27. Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M.J. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graph.* **2015**, *34*, 248. [CrossRef]
28. Mahmood, N.; Ghorbani, N.; Troje, N.F.; Pons-Moll, G.; Black, M.J. AMASS: Archive of Motion Capture as Surface Shapes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019. [CrossRef]
29. Trumble, M.; Gilbert, A.; Malleson, C.; Hilton, A.; Collomosse, J. Total Capture: 3D Human Pose Estimation Fusing Video and Inertial Sensors. In Proceedings of the 2017 British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017.
30. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80–83. [CrossRef]
31. Meiraz, O.; Laufer, S.; Spector, R.; Or, I.; Bolotin, G.; Friedman, T. Enhancing open-surgery gesture recognition using 3D pose estimation. *Int. J. Comput. Assist. Radiol. Surg.* **2026**, *21*, 805–813. [CrossRef] [PubMed]
32. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
33. McGinley, J.L.; Baker, R.; Wolfe, R.; Morris, M.E. The reliability of three-dimensional kinematic gait measurements: A systematic review. *Gait Posture* **2009**, *29*, 360–369. [CrossRef] [PubMed]

34. STMicroelectronics. STM32H7 Series: High-Performance Arm Cortex-M7 Microcontrollers. Available online: https://www.st.com/en/microcontrollers-microprocessors/stm32h7-series.html (accessed on 4 October 2026).
35. STMicroelectronics. STM32N6 Series: High-Performance Arm Cortex-M55 Microcontrollers with Neural-ART Accelerator. Available online: https://www.st.com/en/microcontrollers-microprocessors/stm32n6-series.html (accessed on 4 October 2026).
36. Niño-Tejada, K.; Rivas-Caicedo, J.L.; Saldaña-Aristizábal, L.; Patarroyo-Montenegro, J.F. *D-BRANN Custom OptiTrack+Xsens Motion Capture Dataset*; Zenodo: Geneva, Switzerland, 2026. [CrossRef]

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.