



Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework

Eric Rapp^{a,1}, Soyong Shin^{a,1}, Wolf Thomsen^a, Reed Ferber^b, Eni Halilaj^{a,c,d,*}^a Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA^b Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada^c Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA^d Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

ARTICLE INFO

Article history:

Accepted 3 January 2021

Keywords:

Kinematics
Inertial measurement units
Neural networks
Wearable sensors
Gait

ABSTRACT

The difficulty of estimating joint kinematics remains a critical barrier toward widespread use of inertial measurement units in biomechanics. Traditional sensor-fusion filters are largely reliant on magnetometer readings, which may be disturbed in uncontrolled environments. Careful sensor-to-segment alignment and calibration strategies are also necessary, which may burden users and lead to further error in uncontrolled settings. We introduce a new framework that combines deep learning and top-down optimization to accurately predict lower extremity joint angles directly from inertial data, without relying on magnetometer readings. We trained deep neural networks on a large set of synthetic inertial data derived from a clinical marker-based motion-tracking database of hundreds of subjects. We used data augmentation techniques and an automated calibration approach to reduce error due to variability in sensor placement and limb alignment. On left-out subjects, lower extremity kinematics could be predicted with a mean (\pm STD) root mean squared error of less than 1.27° ($\pm 0.38^\circ$) in flexion/extension, less than 2.52° ($\pm 0.98^\circ$) in ad/abduction, and less than 3.34° ($\pm 1.02^\circ$) internal/external rotation, across walking and running trials. Errors decreased exponentially with the amount of training data, confirming the need for large datasets when training deep neural networks. While this framework remains to be validated with true inertial measurement unit data, the results presented here are a promising advance toward convenient estimation of gait kinematics in natural environments. Progress in this direction could enable large-scale studies and offer new perspective into disease progression, patient recovery, and sports biomechanics.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The ability to passively estimate movement kinematics in natural environments with inertial measurement units (IMUs) could transform how we monitor, diagnose, and treat mobility limitations. These devices are now unobtrusive, can flex and bend with the skin, and allow for patient monitoring throughout the day (Patel et al., 2012; Shull et al., 2014; Shull and Damian, 2015; Son et al., 2014). Despite recent progress in hardware miniaturization, turning large multimodal data from wearable sensors into meaningful biomechanical outcomes that can be easily interpreted in the context of healthy and pathological movement remains a

key challenge toward their widespread use (Picerno, 2017). Biomechanists and rehabilitation specialists have traditionally characterized gait using joint kinematics and kinetics. While segment inertial data generated by IMUs may also lead to important actionable insights, accurate estimation of joint angles is needed to place future findings in the context of past work.

There are currently no streamlined tools to accurately estimate three-dimensional (3-D) joint kinematics from wearable sensors worn in uncontrolled environments. Strap-down integration of inertial data introduces drift, and sensor fusion algorithms that rely on magnetometer data suffer from ferromagnetic disturbances (de Vries et al., 2009). Solutions that incorporate full-body biomechanical models (Robert-Lachaine et al., 2017a, 2017b, 2020) are currently not portable for anytime, anywhere use, and accuracy over long durations remains to be demonstrated. Additionally, the dependence of most algorithms on accurate sensor-to-segment alignment makes translation difficult for multi-day monitoring

* Corresponding author at: Department of Mechanical Engineering, Carnegie Mellon University, Scaife Hall 317, 4805 Frew St., Pittsburgh, PA 15213, USA.

E-mail address: ehalilaj@andrew.cmu.edu (E. Halilaj).

¹ Equal contribution

outside of the laboratory, given human error in sensor placement. Static and dynamic calibrations (Cutti et al., 2010; Favre et al., 2009; Picerno et al., 2008; Roetenberg et al., 2009) may also add to lack of compliance and increased drop-out rates in remote monitoring studies. Further, static poses or functional calibration trials may not be performed as expected in remote scenarios, especially when users suffer from a mobility-limiting condition.

Deep learning and iterative optimization techniques offer a new opportunity to overcome the limitations of previously proposed approaches for estimating kinematics from IMUs. Deep neural networks are highly efficient in learning non-linear relationships from high-dimensional data, such as dense time series from wearable sensors. A key drawback, however, is that they require large datasets to generate accurate models, and such datasets are scarce in the field of biomechanics. A large dataset that contains both IMU data and ground truth joint kinematics from marker-based motion tracking systems, for example, is currently not available to the research community. In other domains, however, synthetic data have been successfully used to train accurate predictive models (Jaderberg et al., 2014). Additionally, optimization approaches have demonstrated success in improving pose estimation in computer vision applications (von Marcard et al., 2018).

The goal of this study was to build deep learning models for predicting 3-D lower extremity joint kinematics directly from angular velocity and linear acceleration data, circumventing the limitations of magnetometer-dependent algorithms. To generate sufficient data to train such models, we created synthetic inertial data from a marker-based motion capture database of hundreds of subjects collected at a clinical center (Ferber et al., 2014; Osis et al., 2015). We further incorporated data augmentation strategies (Shorten and Khoshgoftaar, 2019) to increase the effective sensor placement variability represented in the data, allowing the developed models to tolerate placement ambiguity without compromising accuracy. Additionally, we used an iterative top-down optimization approach to improve the predictions of the deep learning models.

2. Methods

2.1. Data collection and pre-processing

To train the models, we used marker-based motion capture data that were previously collected at the University of Calgary Running Injury Clinic after receiving approval from the University of Calgary's Conjoint Health Research Ethics Board (Ferber et al., 2016; Jauhainen et al., 2020; Phinyomark et al., 2018; Pohl et al., 2010). Retro-reflective marker trajectories were collected at 200 Hz using eight high-speed infrared video cameras (Vicon Motion Systems Ltd., Oxford, UK). After a static neutral trial, subjects performed 60-second walking and running trials at self-selected speeds on a treadmill, following a 2–5 min acclimation period. The participants were patients who enrolled in either clinical or research activities and gave written informed consent prior to participation. A number of the participants were pain-free, while others were experiencing a lower extremity running-related injury.

To perform 3-D kinematic analysis, segment coordinate systems were constructed based on anatomical markers placed on the following landmarks: 1st and 5th metatarsal heads; medial and lateral malleoli; medial and lateral femoral condyles; greater trochanter (bilateral); anterior superior iliac spine (bilateral); iliac crest (bilateral). Tracking marker clusters were placed on the pelvis and bilateral thighs and shanks. A rigid shell with three markers was placed over the sacrum with the two superior markers at the level of the posterior superior iliac spines, while rigid shells

with four markers were attached to the shank and thigh (Fig. 1A). Tracking markers for the feet were placed on the posterior aspect of the shoes: two markers were vertically aligned on the posterior heel counter with a third marker placed laterally. Marker trajectories were filtered with a 10 Hz low-pass 2nd order recursive Butterworth filter. Joint angles were calculated with custom software (Running Injury Clinic Inc., Calgary, Alberta, Canada) using the six degree-of-freedom approach and expressed in terms of joint coordinate systems recommended by the International Society of Biomechanics (Grood and Suntay, 1983; Wu and Cavanagh, 1995). Walking trials from 420 subjects (203 male and 217 female, height of 172.53 ± 8.94 cm, weight of 71.24 ± 12.71 kg) and running trials from 580 subjects (292 male and 288 female, height of 172.70 ± 11.39 cm, weight of 71.30 ± 12.97 kg) were selected for further analysis here, after excluding data with marker-tracking errors.

We generated synthetic inertial data by placing virtual sensors on the tracking marker clusters (Fig. 1B). Seven coordinate systems were created for the lower body (pelvis and both thighs, shanks, and feet) based on the marker clusters, with the origins located at the centroids of the clusters. Inertial data—angular velocity and linear acceleration—were generated through numerical differentiation. Linear acceleration was transformed from a global to a local coordinate system after adding gravity, to match true accelerometer readings. Gaussian noise with a standard deviation that was 15% of the standard deviation of the signal was added to simulate the random noise experienced by true sensors.

2.2. Deep neural networks for segment pose and joint angle prediction

We built deep neural networks that used the synthetic inertial data described above (angular velocities and linear accelerations) from two adjacent segments to predict true kinematics for a particular joint (Fig. 2). The networks were initially trained to predict both segment orientation and joint angles. Each model took the simulated inertial data from the two adjacent segments as input and predicted both adjacent segment orientations and 3-D joint angles. We tested two overarching deep learning architectures with complementary strengths. The first was a one-dimensional convolutional neural network (Conv1D), which makes efficient predictions on a time-window that slides over the entire time series and therefore is more efficient for offline processing. The second was a long short-term memory network (LSTM), which is a specific recurrent neural network that makes predictions one frame at a time, updating an internal state to propagate previous information. LSTMs trade lower computational efficiency during training for higher efficiency in real-time applications.

The learning rate, learning rate decay factor, number of batch iterations, dropout, and the number and size of the hidden layers for both the Conv1D and LSTM models, as well as window size for the Conv1D models and directionality for the LSTM models, were tuned using a Tree of Parzen Estimators optimization algorithm (Bergstra et al., 2011). For each set of hyperparameters, we trained the model by randomly selecting a sequence length of 1 s in stage one and a sequence length of 2 s in stage two. This two-stage scheme allowed the model to update its parameters by first learning the general trends of the gait cycle and then the cyclical nature of the whole time series.

2.3. Data augmentation

To assess if data augmentation approaches could improve sensitivity to sensors placement variability, we performed additional experiments. We initially aligned the virtual sensors with the marker clusters and trained models using only these sensor placement configurations (*baseline model*). The data were then augmented to

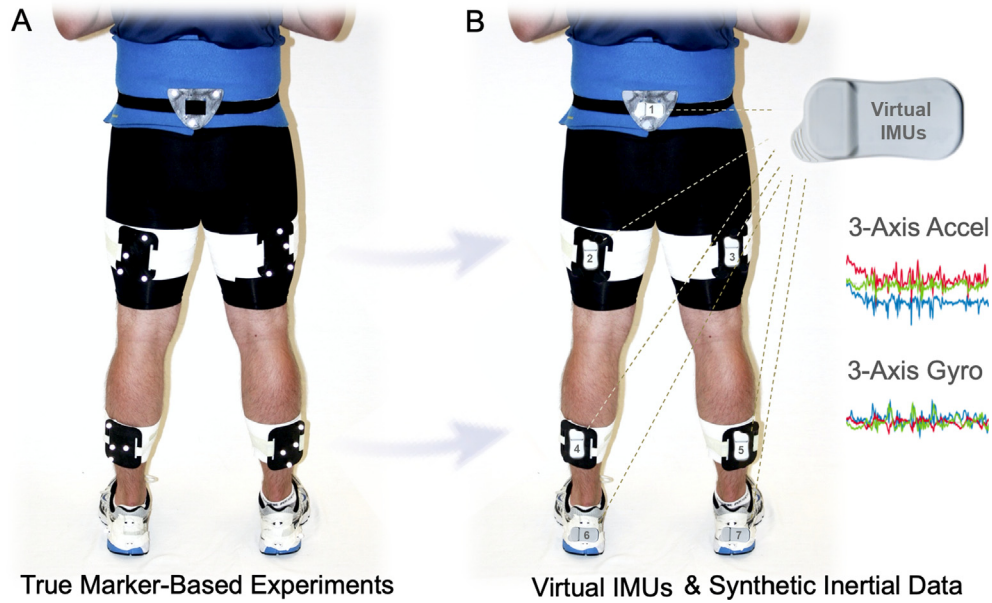


Fig. 1. Experimental Set Up and Data Generation. A) Marker-based motion capture data were previously collected at the Calgary Running Injury Clinic from hundreds of subjects. B) Tracking marker clusters were used to place virtual IMUs in the sacrum, thighs, shanks, and feet (1–7, respectively). Synthetic accelerometry and gyroscope data were generated by taking numerical derivatives and adding Gaussian noise.

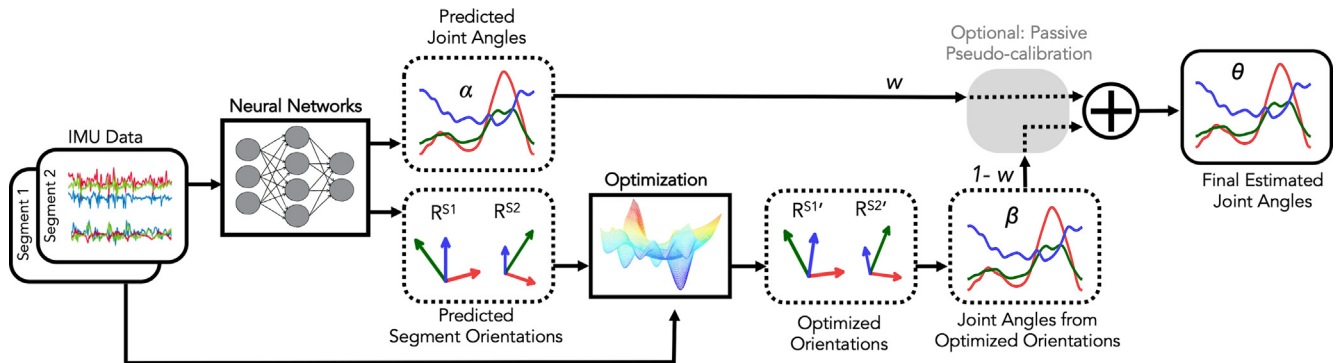


Fig. 2. Hybrid Deep Learning and Optimization Approach. The synthetic sensor data (tri-axial linear acceleration and angular velocity) from adjacent segments were used to predict true segment orientation \mathbf{R}^{S1} , \mathbf{R}^{S2} and joint angles $[\alpha_1, \alpha_2, \alpha_3]$. Segment orientation predictions were updated iteratively (optimization step) until the angular velocity associated with them best matched the virtual sensor data, and updated joint angles were computed from these optimized orientations $[\beta_1, \beta_2, \beta_3]$. The final 3-D joint angles (θ) were then computed using a weighted average of the prediction from the deep learning model (α) and the results of the optimization routine (β).

account for sensor orientation variability by drawing randomly sampled rotations from a normal distribution (STD = 13.5°) centered around the “true” orientation via an online augmentation paradigm (Krizhevsky et al., 2012; Shorten and Khoshgoftaar, 2019). Models were also trained with the augmented dataset (*augmented model*). We tested the performance of both of these models on test data that were not augmented. Additional experiments were run to gauge the ability of each of these models to compensate for variability in sensor placement for the test subjects, which was achieved by augmenting the test data in a similar manner to the training data.

2.4. Segment pose optimization

To further improve accuracy, we implemented a top-down optimization method that fine-tuned the neural network prediction α by comparing the angular velocities (Ω') from the predicted segment orientations \mathbf{R}^S with the angular velocities from the true virtual IMU data (Ω). Each body-segment orientation

was updated iteratively until the angular velocities associated with it best matched the virtual IMU data (Fig. 2). The final 3D joint angles (θ) were then computed using a weighted average of the prediction from the deep learning model (α) and the results of the optimization routine (β). Here, we refer to this as the *optimized model*.

Specifically, the combined deep learning and optimization approach includes two stages: initialization and frame-by-frame optimization. At the first stage, the deep neural networks estimate an initial joint angle (α) and adjacent body segment orientations (\mathbf{R}^{S1} and \mathbf{R}^{S2}). The frame-by-frame optimization stage updates the predicted segment orientations iteratively by minimizing the following error term for each frame, k :

$$e_k = |\Omega_k - \Omega'_k|$$

The optimization was carried out using an unconstrained Broyden-Fletcher-Goldfarb-Shanno optimizer with a function tolerance of 10^{-6} , gradient tolerance of 10^{-5} , and maximum iteration number of 200. To derive the final joint angle estimation θ , we used a weighted average of the neural network prediction α and the

optimization result β , where $\theta = \mathbf{w} \cdot \alpha + (1 - \mathbf{w}) \cdot \beta$. The optimal weight \mathbf{w} was learned by minimizing the error between θ and the ground truth joint angles from optical motion tracking on the training dataset.

2.5. Passive pseudo-calibration

Since the approach is purely data-driven and does not require a biomechanical model with new data, an active calibration routine with either static poses or dynamic functional tasks is not necessary. Data-driven models, however, generalize better to subjects whose neutral limb alignment is most represented in the training dataset. To overcome this limitation, we expressed the time series as a function of the mean joint angle in each degree of freedom. We call this procedure *passive pseudo-calibration*, since it does not engage the user and is not meant to replace a standard calibration. It does, however, ensure that subjects with limb alignments away from the mean have the same chance at accurate prediction of kinematics as subjects who are closer to the mean.

2.6. Performance evaluation

The data were split into training (80%), validation (10%), and test (10%) sets, and separate predictive models were trained for each joint (hip, knee, ankle). The root-mean-squared error (RMSE) between the predicted and ground truth kinematics was computed for each joint and each hyper-parameter configuration. The model with the minimum RMSE on the validation data was selected as the best model for each joint. Final models were then tested on the left-out, test data. Here we report RMSEs for the test data, using the best models trained with non-augmented data (*baseline model*) and augmented data (*augmented model*) before the optimization step, and after incorporation of the optimization step (*optimized model*). These models were evaluated with and without the *passive pseudo-calibration* step. To determine the effect of training data size on model performance, we trained and evaluated models using only 10, 50, 100, 200 subjects, in addition to the standard experiments with 80% of the total subjects. We randomly selected the subsets and repeated the procedure 5 times. Here we report mean (\pm STD) RMSEs from the 5 trials using models trained with augmented data and tested with non-augmented data, without implementing the optimization or pseudo-calibration steps since the goal was to determine the effect of sample size on the neural networks. All the models were tested on the same set of test subjects ($n = 42$ for walking and $n = 58$ for running) per trial. Repeated measures analyses of variance (ANOVAs) were carried out to determine if the RMSE changed significantly with more training data.

3. Results

Walking kinematics could be predicted with a mean (\pm STD) RMSE of less than $2.75^\circ (\pm 0.66^\circ)$, while running kinematics could be predicted with a mean RMSE of less than $3.34 (\pm 1.02^\circ)$ using

the optimized model, along with pseudo-calibration. During walking, flexion/extension was the most accurate degree of freedom, with a mean RMSE of less than $0.97^\circ (\pm 0.38^\circ)$ across the ankle, knee, and hip joints, followed by ab/adduction with a mean RMSE of less than $2.16^\circ (\pm 0.85^\circ)$ and internal/external rotation with a mean RMSE of less than $2.75^\circ (\pm 0.66^\circ)$ (Table 1; Fig. 3). Similarly, during running, flexion/extension was the most accurate degree of freedom, with a mean RMSE of less than $1.27^\circ (\pm 0.38^\circ)$ across the ankle, knee, and hip joints, followed by ab/adduction with a mean RMSE of less than $2.52^\circ (\pm 0.98^\circ)$, and internal/external rotation with a mean RMSE of less than $3.34^\circ (\pm 1.02^\circ)$ (Table 2; Fig. 3). Generally, Conv1D models with 2–3 hidden layers, more than 50 nodes per layer, a window size greater than 55 ms, learning rate greater than 0.001, and 5000 iterations performed the best.

The final predictive performance reported above improved from the baseline neural network predictions through the use of iterative optimization, mainly for flexion/extension angles, but not the other degrees of freedom. Optimization improved prediction of knee flexion by over 15%, hip flexion by over 10%, and ankle flexion by over 15% (Tables 1 and 2). When passive pseudo-calibration was also applied, optimization had an even larger effect on final kinematics. In this case, prediction of knee flexion improved by over 50%, hip flexion by over 65%, ankle flexion by over 45%, and internal/external rotation of the hip and ankle by about 30% compared to the neural network outputs. For other degrees of freedom, optimal weight priors (\mathbf{w}) were close to 1, indicating that the prediction was predominantly based on the neural network. Augmenting the training dataset to account for sensor application variability resulted in lower RMSEs when the test data were also augmented (Table 3), but not when the test data were not augmented (Table 2).

Increasing the amount of training data led to lower mean RMSEs when predicting kinematics in new subjects, during both walking and running (Fig. 4; $p < .0001$). Results from models using the maximum number of training subjects are different from those reported earlier (Tables 1 and 2) since here we averaged across 5 trials. Also, the reported RMSEs do not include optimization and pseudo-calibration steps. Overall, fitted learning curves indicated improvements of over 1° for flexion/extension. On a linear scale, the learning curves also started to, but did not completely, plateau as the training data increased from tens to hundreds of subjects, indicating that inclusion of more data in the future could continue to reduce the RMSE. Projections indicate that thousands of subjects are needed to improve the models by another degree (Fig. 4).

4. Discussion

The goal of this study was to introduce a new framework that combines deep neural networks with top-down optimization for prediction of lower extremity kinematics from inertial sensing data. Using this hybrid approach and synthetic inertial data from a large clinical motion capture database, we could predict walking and running kinematics with accuracies that are similar to the reli-

Table 1
Predictive accuracy (mean RMSE \pm STD in degrees) for walking kinematics at different steps of the pipeline.

RMSE (mean \pm std)	Knee			Hip			Ankle			
	Flex	Add	Rot	Flex	Add	Rot	Flex	Add	Rot	
Baseline	3.73 \pm 1.64	3.20 \pm 1.25	6.73 \pm 3.50	4.97 \pm 2.40	2.78 \pm 1.21	5.71 \pm 3.06	2.15 \pm 0.77	2.76 \pm 1.42	4.81 \pm 2.70	
+ Augmentation	3.80 \pm 1.68	3.22 \pm 1.36	6.58 \pm 3.79	5.02 \pm 2.49	2.86 \pm 1.30	5.75 \pm 3.13	2.20 \pm 0.77	2.90 \pm 1.45	4.62 \pm 2.72	
+ Optimization	3.12 \pm 1.91	3.18 \pm 1.36	6.35 \pm 3.88	4.34 \pm 2.77	2.70 \pm 1.33	5.16 \pm 3.37	1.77 \pm 0.91	2.75 \pm 1.50	4.28 \pm 2.87	
+ Passivepseudo-calibration	Baseline	2.27 \pm 0.47	2.23 \pm 0.76	3.53 \pm 0.82	2.37 \pm 0.59	2.02 \pm 0.74	2.79 \pm 0.73	1.37 \pm 0.30	1.56 \pm 0.38	2.40 \pm 0.50
	+ Augmentation	2.17 \pm 0.57	2.20 \pm 0.86	3.12 \pm 0.73	2.27 \pm 0.56	2.10 \pm 0.83	2.61 \pm 0.73	1.40 \pm 0.31	1.57 \pm 0.46	2.02 \pm 0.49
	+ Optimization	0.97 \pm 0.38	2.16 \pm 0.85	2.75 \pm 0.66	0.67 \pm 0.37	1.88 \pm 0.89	1.42 \pm 0.71	0.72 \pm 0.28	1.32 \pm 0.51	1.30 \pm 0.62

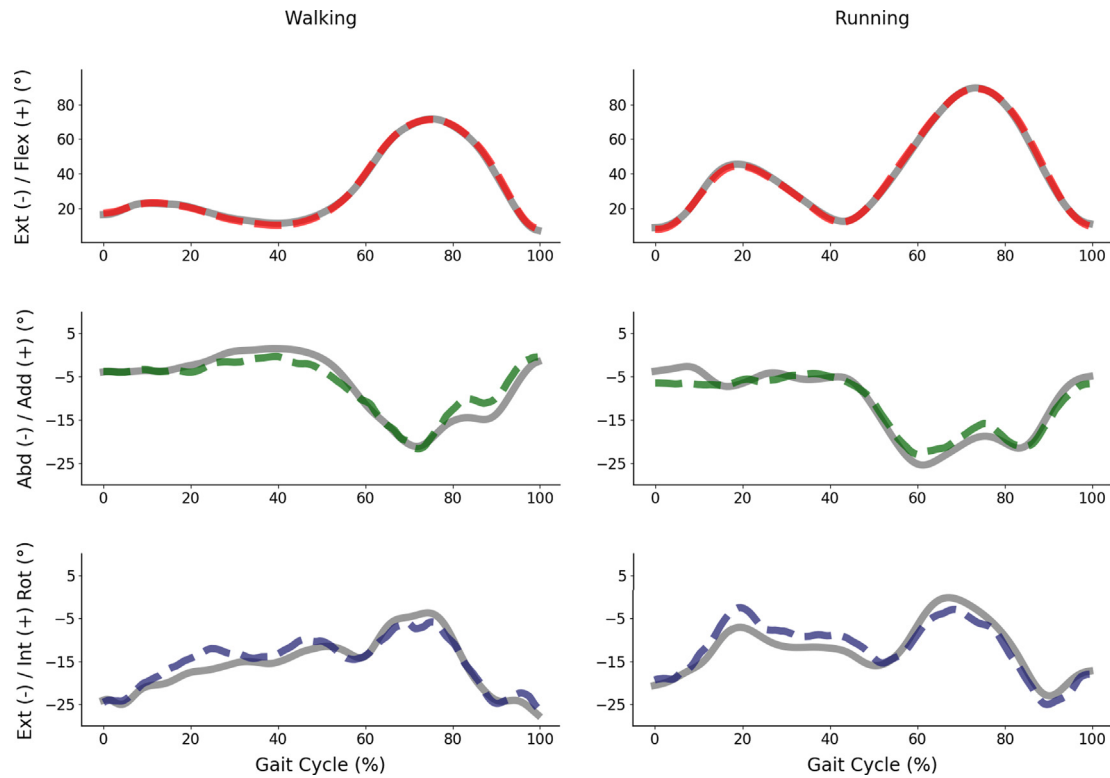


Fig. 3. Predicted and Marker-Based Kinematics. A representative gait cycle from a subject whose RMSEs of the predicted versus marker-based kinematics were close to the mean across all subjects. Predictions from the combined deep learning and optimization framework (dashed) closely match marker-based kinematics (solid gray), especially in flexion/extension (red), followed by add/abduction (green) and internal/external rotation (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Predictive accuracy (mean RMSE ± STD in degrees) for running kinematics at different steps of the pipeline.

RMSE (mean ± std)	Knee			Hip			Ankle			
	Flex	Add	Rot	Flex	Add	Rot	Flex	Add	Rot	
Baseline	4.25 ± 1.76	4.12 ± 2.01	6.51 ± 3.90	5.08 ± 2.72	3.33 ± 1.46	5.36 ± 2.57	2.75 ± 1.16	2.92 ± 1.69	4.62 ± 2.70	
+ Augmentation	4.08 ± 1.83	4.22 ± 2.17	6.62 ± 3.84	5.09 ± 2.76	3.24 ± 1.46	5.17 ± 2.70	2.67 ± 1.11	3.15 ± 1.83	4.53 ± 2.65	
+ Optimization	3.21 ± 2.10	4.16 ± 2.20	6.55 ± 3.88	4.31 ± 3.08	2.98 ± 1.46	4.59 ± 2.90	2.13 ± 1.28	3.00 ± 1.90	4.51 ± 2.73	
+ Passivepseudo-calibration	Baseline	2.84 ± 0.69	2.63 ± 0.91	3.43 ± 0.96	2.63 ± 0.54	2.31 ± 0.89	2.86 ± 0.77	1.91 ± 0.55	1.66 ± 0.44	2.20 ± 0.56
+ Augmentation	2.60 ± 0.58	2.59 ± 0.96	3.46 ± 0.93	2.42 ± 0.55	2.25 ± 0.97	2.54 ± 0.69	1.76 ± 0.45	1.85 ± 0.52	1.92 ± 0.55	
+ Optimization	1.27 ± 0.38	2.52 ± 0.98	3.34 ± 1.02	0.79 ± 0.32	1.87 ± 0.94	1.32 ± 0.78	0.94 ± 0.34	1.60 ± 0.67	1.78 ± 1.07	

Table 3

Predictive accuracy (mean RMSE ± STD in degrees) of the baseline and augmented models. No optimization or passive pseudo-calibration was applied.

RMSE (mean ± std)		Knee			Hip			Ankle		
		Flex	Add	Rot	Flex	Add	Rot	Flex	Add	Rot
Walking	Baseline	8.27 ± 4.38	4.87 ± 2.01	9.38 ± 4.72	9.57 ± 5.70	4.33 ± 2.07	8.76 ± 4.92	5.79 ± 2.57	4.82 ± 2.34	7.12 ± 3.42
	+ Augmentation	3.87 ± 1.69	3.35 ± 1.43	6.77 ± 3.85	5.13 ± 2.41	2.82 ± 1.26	5.91 ± 3.25	2.32 ± 0.78	2.98 ± 1.48	4.70 ± 2.66
Running	Baseline	7.56 ± 4.03	5.55 ± 2.72	8.54 ± 3.90	7.91 ± 4.29	4.45 ± 1.79	6.94 ± 3.15	6.64 ± 3.75	5.42 ± 2.50	7.09 ± 3.02
	+ Augmentation	4.23 ± 1.78	4.26 ± 2.18	6.59 ± 3.63	5.14 ± 2.72	3.37 ± 1.56	5.39 ± 2.69	2.82 ± 1.24	3.28 ± 1.88	4.60 ± 2.67

ability of marker-based motion tracking. We also demonstrated that augmentation techniques that increase effective sensor placement variability can improve model generalizability to sensor placement inconsistencies. Specifically, kinematics could be predicted with a mean RMSE of less than 1.27° (±0.38°) in flexion/extension, less than 2.52° (±0.98°) in ad/abduction, and less than 3.34° (±1.02°) internal/external rotation. Predictive accuracy was consistent across male and female subjects and did not vary with subject weight and height (analysis not presented here). The higher RMSE in internal/external rotation may be explained by

the lack of magnetometer data, which, when undisturbed, contribute toward heading estimation. Further, optical motion tracking, which was used as ground truth here, exhibits similar accuracies across degrees of freedom (Miranda et al., 2013; Tsai et al., 2011).

These promising results should be interpreted in the context of several limitations associated with the study. First, while we used data from hundreds of subjects, which contributed to improved model generalizability, the data were not acquired from true inertial sensors. Synthetic inertial data were generated from virtual

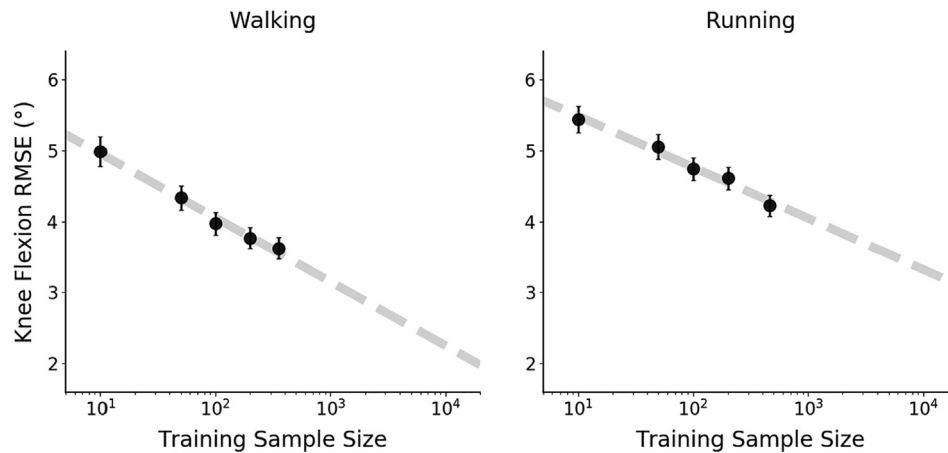


Fig. 4. Neural Network Learning Curves. Models trained using a training data of different sizes (10, 50, 100, 200, and 80% of all the subjects) demonstrate that performance improves with larger datasets. Mean ($\pm 95\%$ CI) RMSE on the same set of left-out test data significantly decreased with models trained on larger data. Fitted lines (dashed gray) estimate the amount of training data needed to improve performance further.

sensors placed in accordance with optical motion tracking markers. We modeled the white noise associated with inertial sensors synthetically, but several other sources of noise, including zero-offset bias and soft tissue movement, were not modeled. While we expect the bias not to significantly affect the performance of these models, soft-tissue artifact remains a challenge. A second limitation is that, although moderate placement errors are surmountable given our data augmentation approach, the models developed here can only lead to plausible predictions if the sensors are placed on the same location as the location that we placed the virtual IMUs, which coincided with the location of the tracking marker clusters. Future work can focus on training more generalizable models that include virtual IMUs placed on several locations across each body segment. Another limitation is that the models were trained using treadmill walking and running data, while their intended use is for over-ground activities. Previous studies have documented little to no differences between treadmill and level ground walking and running (Riley et al., 2007). While we expect the models to generalize well to level walking and running, rugged terrain may introduce challenges. It is also important to note that for the proposed algorithms to be successful at predicting kinematics, walking and running should be segmented out of time series collected in free-living environments, which may include other activities and turns. Activity classification algorithms can predict walking and running with high accuracy given a few sensors worn on lower extremity segments (Mannini and Sabatini, 2010).

While validation with true sensor data is a necessary follow-up step, the accuracy of the presented framework is a promising advance toward prediction of joint kinematics from IMUs without relying on drift-prone algorithms and error-prone calibration techniques. Previously proposed approaches have reported RMSEs of more than 3° over limited durations (Dorschky et al., 2019; Karatsidis et al., 2018; Robert-Lachaine et al., 2020, 2017b). Because many rely on the use of magnetometers and filtering approaches, long-term reliability remains a limitation for natural environment applications. Algorithms that incorporate resetting to reduce drift (e.g., zero velocity update) may not generalize well to clinical populations with unpredictable gait patterns (Yang et al., 2013). In addition to not experiencing drift, our models are also well-suited for real-time applications that integrate monitoring with biofeedback. While the optimization step may introduce delays for real-time estimation and be better suited for post-hoc analyses, deep neural networks without the optimization step demonstrate mean RMSEs of less than $3.46^\circ (\pm 0.93)$.

Reducing model sensitivity to sensor placement variability is desired for translation of wearable sensors out of the laboratory and clinic, where patients will have to apply them on their own, and likely more than once. While previous approaches are reliant on careful sensor-to-segment alignment and calibration trials, including functional (Favre et al., 2009), positional (Robert-Lachaine et al., 2017b), and constraint-based (Seel et al., 2014) calibrations, our approach foregoes any active calibration and can overcome the challenges posed by sensor placement variability or need for supervision by trained research or clinical staff. We instead used data augmentation and passive pseudo-calibration to overcome sensor placement and natural limb alignment variability. Data augmentation is an advantageous technique for increasing the amount of effective training data, and here we demonstrated that it can reduce the effect of sensor placement variability. Further, the predictive accuracy of our algorithms can be tuned passively after a few gait cycles (Fig. 5), without requiring attention from the user as is the case with active calibration routines.

The accuracy of the models presented here is in part due to the large motion capture dataset that we used, which enabled the generation of a synthetic IMU database with hundreds of subjects. Large datasets are a requisite for training generalizable deep neural networks, but they are often not available in movement biomechanics. While the use of machine learning is growing, often deep neural networks are trained on limited data (Halilaj et al., 2018). Our experiments demonstrated that models trained on data from a larger number of subjects (i.e., hundreds) exhibit lower RMSEs compared to models trained on smaller data (i.e., tens), suggesting that they are more likely to generalize well to new, independent data. Learning curves (Fig. 4) are useful tools for determining how much data may be sufficient for a particular application, allowing one to weigh improvement in accuracy against efforts associated with acquiring more data. For studies designed with a machine learning approach in mind (instead of traditional hypothesis testing), they can replace power analyses in the planning phase.

Overall, this study introduced an integrated deep learning and optimization framework for the prediction of joint kinematics from inertial sensing data and presented a series of experiments that enable evaluation of the critical steps that contributed to the accuracy of these data-driven models. The presented techniques can also be extended to estimation of joint kinetics in the future (Johnson et al., 2020). While further validation of the models with

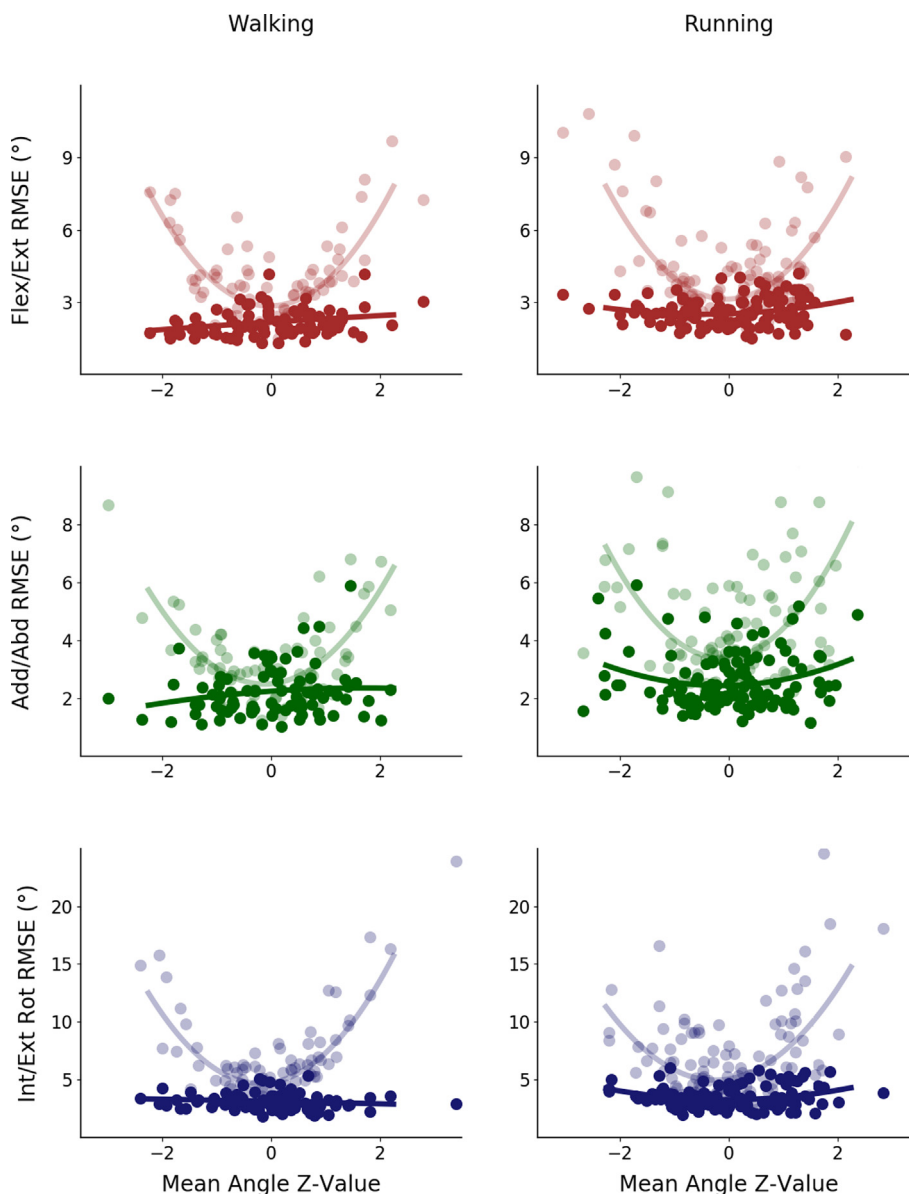


Fig. 5. Effect of Passive Pseudo-Calibration. Prediction error for all three degrees of freedom was higher in subjects with overall kinematics that were further away from the mean, likely due to differences in limb alignment and sensor placement (light red, green, blue). Using the mean joint angles as a neutral pose, we passively calibrated joint angle predictions, which rendered the error relatively similar across subjects, despite their deviation from the mean (dark red, green, blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

true IMU data is a necessary next step, the promising results presented here, along with the publicly available tools, are encouraging steps toward addressing one of the most pressing technical challenges in modern-day biomechanics.

Acknowledgement

The authors would like to thank Allan Brett for his assistance with the data transfer and for patiently answering questions as the project progressed.

References

Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems.*, 2546–2554 <https://doi.org/10.1145/3065386>.
 Cutti, A.G., Ferrari, Alberto, Garofalo, P., Raggi, M., Cappello, A., Ferrari, Adriano, 2010. ‘Outwalk’: a protocol for clinical gait analysis based on inertial and

magnetic sensors. *Med. Biol. Eng. Comput.* 48, 17–25. <https://doi.org/10.1007/s11517-009-0545-x>.
 de Vries, W.H.K., Veeger, H.E.J., Baten, C.T.M., van der Helm, F.C.T., 2009. Magnetic distortion in motion labs, implications for validating inertial magnetic sensors. *Gait Posture* 29, 535–541. <https://doi.org/10.1016/j.gaitpost.2008.12.004>.
 Dorschky, E., Nitschke, M., Seifer, A.-K., van den Bogert, A.J., Eskofier, B.M., 2019. Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models. *J. Biomech.* 95., <https://doi.org/10.1016/j.jbiomech.2019.07.022> 109278.
 Favre, J., Aïssaoui, R., Jolles, B.M., de Guise, J.A., Aminian, K., 2009. Functional calibration procedure for 3D knee joint angle description using inertial sensors. *J. Biomech.* 42, 2330–2335. <https://doi.org/10.1016/j.jbiomech.2009.06.025>.
 Ferber, R., Bolgla, L., Earl-Boehm, J.E., Emery, C., Hamstra-Wright, K., 2014. Strengthening of the Hip and Core Versus Knee Muscles for the Treatment of Patellofemoral Pain: A Multicenter Randomized Controlled Trial. *J. Athl. Train.* 50, 366–377. <https://doi.org/10.4085/1062-6050-49.3.70>.
 Ferber, R., Osis, S.T., Hicks, J.L., Delp, S.L., 2016. Gait Biomechanics in the Era of Data Science. *J. Biomech.* 49, 3759–3761. <https://doi.org/10.1016/j.jbiomech.2016.10.033>.
 Grood, E.S., Suntay, W.J., 1983. A Joint Coordinate System for the Clinical Description of Three-Dimensional Motions: Application to the Knee. *J. Biomech. Eng.* 105, 136–144. <https://doi.org/10.1115/1.3138397>.

- Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J., Delp, S.L., 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J. Biomech.* 81, 1–11. <https://doi.org/10.1016/j.jbiomech.2018.09.009>.
- Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Synthetic data and artificial neural networks for natural scene text recognition. *ArXiv Prepr. ArXiv14062227*.
- Jauhiainen, S., Pohl, A.J., Äyrämö, S., Kauppi, J.-P., Ferber, R., 2020. A hierarchical cluster analysis to determine whether injured runners exhibit similar kinematic gait patterns. *Scand. J. Med. Sci. Sports* 30, 732–740. <https://doi.org/10.1111/sms.13624>.
- Johnson, W.R., Mian, A., Robinson, M.A., Verheul, J., Lloyd, D.G., Alderson, J., 2020. Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. *IEEE Trans. Biomed. Eng.* 1–1. <https://doi.org/10.1109/TBME.2020.3006158>.
- Karatsidis, A., Jung, M., Schepers, H.M., Bellusci, G., de Zee, M., Veltink, P.H., Andersen, M.S., 2018. Predicting kinetics using musculoskeletal modeling and inertial motion capture. *ArXiv180101668 Phys*.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Adv. Neural Info. Process. Syst.*, 1097–1105.
- Mannini, A., Sabatini, A.M., 2010. Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. *Sensors* 10, 1154–1175. <https://doi.org/10.3390/s100201154>.
- Miranda, D.L., Rainbow, M.J., Crisco, J.J., Fleming, B.C., 2013. Kinematic differences between optical motion capture and biplanar videoradiography during a jumpcut maneuver. *J. Biomech.* 46, 567–573. <https://doi.org/10.1016/j.jbiomech.2012.09.023>.
- Osis, S.T., Hettinga, B.A., Macdonald, S.L., Ferber, R., 2015. A novel method to evaluate error in anatomical marker placement using a modified generalized Procrustes analysis. *Comput. Methods Biomech. Biomed. Engin.* 18, 1108–1116.
- Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M., 2012. A review of wearable sensors and systems with application in rehabilitation. *J. NeuroEngineering Rehabil.* 9, 21. <https://doi.org/10.1186/1743-0003-9-21>.
- Phinyomark, A., Petri, G., Ibáñez-Marcelo, E., Osis, S.T., Ferber, R., 2018. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *J. Med. Biol. Eng.* 38, 244–260. <https://doi.org/10.1007/s40846-017-0297-2>.
- Picerno, P., 2017. 25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches. *Gait Posture* 51, 239–246. <https://doi.org/10.1016/j.gaitpost.2016.11.008>.
- Picerno, P., Cereatti, A., Cappozzo, A., 2008. Joint kinematics estimate using wearable inertial and magnetic sensing modules. *Gait Posture* 28, 588–595. <https://doi.org/10.1016/j.gaitpost.2008.04.003>.
- Pohl, M.B., Lloyd, C., Ferber, R., 2010. Can the reliability of three-dimensional running kinematics be improved using functional joint methodology?. *Gait Posture* 32, 559–563. <https://doi.org/10.1016/j.gaitpost.2010.07.020>.
- Riley, P.O., Paolini, G., Della Croce, U., Paylo, K.W., Kerrigan, D.C., 2007. A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects. *Gait Posture* 26, 17–24. <https://doi.org/10.1016/j.gaitpost.2006.07.003>.
- Robert-Lachaine, X., Mecheri, H., Larue, C., Plamondon, A., 2017a. Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Med. Biol. Eng. Comput.* 55, 609–619. <https://doi.org/10.1007/s11517-016-1537-2>.
- Robert-Lachaine, X., Mecheri, H., Larue, C., Plamondon, A., 2017b. Accuracy and repeatability of single-pose calibration of inertial measurement units for whole-body motion analysis. *Gait Posture* 54, 80–86. <https://doi.org/10.1016/j.gaitpost.2017.02.029>.
- Robert-Lachaine, X., Parent, G., Fuentes, A., Hagemester, N., Aissaoui, R., 2020. Inertial motion capture validation of 3D knee kinematics at various gait speed on the treadmill with a double-pose calibration. *Gait Posture* 77, 132–137. <https://doi.org/10.1016/j.gaitpost.2020.01.029>.
- Roetenberg, D., Luinge, H., Slycke, P., 2009. Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors. *Xsens Motion Technol. BV Tech. Rep.*, 1.
- Seel, T., Raisch, J., Schauer, T., 2014. IMU-Based Joint Angle Measurement for Gait Analysis. *Sensors* 14, 6891–6909. <https://doi.org/10.3390/s140406891>.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>.
- Shull, P.B., Damian, D.D., 2015. Haptic wearables as sensory replacement, sensory augmentation and trainer – a review. *J. NeuroEngineering Rehabil.* 12, 59. <https://doi.org/10.1186/s12984-015-0055-z>.
- Shull, P.B., Jirattigalachote, W., Hunt, M.A., Cutkosky, M.R., Delp, S.L., 2014. Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. *Gait Posture* 40, 11–19. <https://doi.org/10.1016/j.gaitpost.2014.03.189>.
- Son, D., Lee, J., Qiao, S., Ghaffari, R., Kim, J., Lee, J.E., Song, C., Kim, S.J., Lee, D.J., Jun, S. W., Yang, S., Park, M., Shin, J., Do, K., Lee, M., Kang, K., Hwang, C.S., Lu, N., Hyeon, T., Kim, D.-H., 2014. Multifunctional wearable devices for diagnosis and therapy of movement disorders. *Nat. Nanotechnol.* 9, 397–404. <https://doi.org/10.1038/nnano.2014.38>.
- Tsai, T.-Y., Lu, T.-W., Kuo, M.-Y., Lin, C.-C., 2011. Effects of soft tissue artifacts on the calculated kinematics and kinetics of the knee during stair-ascent. *J. Biomech.* 44, 1182–1188. <https://doi.org/10.1016/j.jbiomech.2011.01.009>.
- von Marcard, T., Henschel, R., Black, M.J., Rosenhahn, B., Pons-Moll, G., 2018. Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. Presented at the Proceedings of the European Conference on Computer Vision (ECCV), pp. 601–617. [10.1007/978-3-030-01249-6_37](https://doi.org/10.1007/978-3-030-01249-6_37).
- Wu, G., Cavanagh, P.R., 1995. ISB recommendations for standardization in the reporting of kinematic data. *J. Biomech.* 28, 1257–1261. [https://doi.org/10.1016/0021-9290\(95\)00017-c](https://doi.org/10.1016/0021-9290(95)00017-c).
- Yang, S., Zhang, J.-T., Novak, A.C., Brouwer, B., Li, Q., 2013. Estimation of spatio-temporal parameters for post-stroke hemiparetic gait using inertial sensors. *Gait Posture* 37, 354–358. <https://doi.org/10.1016/j.gaitpost.2012.07.032>.