scientific reports

OPEN

# Validation of portable in-clinic video-based gait analysis for prosthesis users

Anthony Cimorelli[1], Ankit Patel[2,3], Tasos Karakostas[1,4] & R. James Cotton[1,4]✉

**Despite the common focus of gait in rehabilitation, there are few tools that allow quantitatively characterizing gait in the clinic. We recently described an algorithm, trained on a large dataset from our clinical gait analysis laboratory, which produces accurate cycle-by-cycle estimates of spatiotemporal gait parameters including step timing and walking velocity. Here, we demonstrate this system generalizes well to clinical care with a validation study on prosthetic users seen in therapy and outpatient clinics. Specifically, estimated walking velocity was similar to annotated 10-m walking velocities, and cadence and foot contact times closely mirrored our wearable sensor measurements. Additionally, we found that a 2D keypoint detector pretrained on largely able-bodied individuals struggles to localize prosthetic joints, particularly for those individuals with more proximal or bilateral amputations, but after training a prosthetic-specific joint detector video-based gait analysis also works on these individuals. Further work is required to validate the other outputs from our algorithm including sagittal plane joint angles and step length. Code for the gait transformer and the trained weights are available at https://github.com/peabody124/GaitTransformer.**

Gait impairments are a common target for rehabilitation. The most widely used outcome measures are the 10-m walk test or 6-min walk test which measure walking speed and endurance[1], but do not capture more detailed walking biomechanics. Alternatively, a motion analysis laboratory uses optical motion capture and force plates to obtain precise estimates of joint kinematics and kinetics and compute temporal and spatiotemporal gait parameters[2]. However, the cost, time, and equipment required for formal gait analysis preclude performing it frequently, during clinical encounters, or outside the laboratory. There is a substantial need for clinically-usable tools that fill the gap between formal gait assessment and performance-based outcome measures to enable routine, quantitative characterization of gait and its associated impairments. These tools would enable more sensitive outcome measures to follow patients' progress with therapy as well as to better power research to improve interventions. It would also enable routine screening of gait parameters in the clinic. This has many potential applications, such as allowing early detection of gait parameters associated with a risk of falling[3], which could then enable earlier interventions with fall prevention strategies.

Given the importance of analyzing gait and mobility, numerous approaches have been explored including wearable sensors and video analysis. Many different types of wearable sensors and algorithms have been described ranging from wrist devices that estimate step counts to placing numerous sensors over the body to estimate more complete kinematics[4–6]. An advantage of wearables is they can enable ubiquitous monitoring of activity throughout the day[7]. However, they can often require extensive and time-consuming calibration, require complicated and often proprietary algorithms to extract the relevant information, and have a wide range of reported accuracies[8].

In recent years, human pose estimation using deep learning has seen an extremely rapid advance yielding numerous approaches that can estimate 3D joint locations and pose[9–11]. However, the performance measure for the majority of this computer vision research is the accuracy of estimating 3D joint locations and not biomechanically motivated kinematics or gait parameters. Using multiple cameras, joint locations can be triangulated in 3D to produce more accurate estimates and these systems have been validated on numerous aspects of gait[12–18]. In general, the need for multiple cameras makes these systems less portable and amenable for use in a clinic, although OpenCap has shown it is possible using only a computer and two smartphones with calibrated positions[19]. Approaches have been developed that can analyze cycle-by-cycle gait from monocular video[20,21], but they have not been validated on data acquired in the clinic or on clinical populations. Alternatively, other

[1]Shirley Ryan AbilityLab, Chicago, USA. [2]Department of Neuroscience, Baylor College of Medicine, Houston, USA. [3]Department of Electrical & Computer Engineering, Rice University, Houston, USA. [4]Department of Physical Medicine and Rehabilitation, Northwestern University, Evanston, USA. ✉email: rcotton@sralab.org

approaches train a neural network to directly map a sequence of 2D keypoints to average gait parameters that have been tested on clinical populations but do not enable analyzing individual gait cycles[22,23].

We developed an algorithm, the Gait Transformer, trained on a large clinical gait laboratory dataset of paired videos and motion capture data[24]. This Gait Transformer decomposes Human Pose Estimation (HPE) trajectories of walking into individual gait cycles to produce accurate estimates of gait event timing and walking velocity, *when tested on that dataset*. However, artificial intelligence (AI) algorithms can be sensitive to changes in the data distribution. In this case, there are numerous differences between videos of gait collected in the real world or clinic and those from the gait laboratory. Thus, validating how this algorithm generalizes to data collected in clinical settings—the primary goal—is critical to enabling its use. The goals of this work include: (1) describe our combined system including smartphone application, wearable sensors, Pose Pipeline[25] and Gait Transformer[24] for clinical gait analysis, (2) validate the performance of this system on data acquired in the clinic, (3) identify under which conditions it performs well and when it is less reliable.

We performed this validation on prosthesis users and selected this population for several reasons. From a technical perspective, this is a challenging test of this system as the limbs of prosthesis users often appear visually different than able-bodied individuals, and it was previously unknown if pretrained HPE algorithms will generalize to prosthesis users. In addition, some prosthesis users walk with significant gait deviations compared to able-bodied individuals[26], which further challenges the Gait Transformer. From a clinical perspective, we expect that routine access to video-based gait analysis would enable better outcomes to monitor improvements in walking with therapy or with adjustments to prosthetic components, but this will ultimately need to be empirically validated through clinical trials. Previous studies have demonstrated that lower limb prosthesis users saw an improvement in various gait parameters including walking speed, distance walked, spatiotemporal measures and biomechanics with specialized therapeutic interventions[27,28]. However, these studies used performance-based outcome measures and those that analyzed biomechanics were limited to laboratory settings over level ground. The ability to perform routine, quantitative gait analysis could identify improvements in the quality of walking during therapy, such as increased symmetry and time spent in single stance on the prosthetic limb. This increased level of detail for analyzing gait may allow for more sensitive outcome measurements to demonstrate how an individual's gait continues to improve when their walking velocity plateaus. This could enable physical therapists to demonstrate to insurance agencies that patients are still making progress and justify additional sessions. Or, more sensitive measures could provide data for prosthetists to show advanced components also improve quality of walking. As prosthetic components continue to advance, it is becoming increasingly difficult to get high-end prosthetic components covered by payers and there is a need to develop systems that can quantify prosthetic gait in real-world settings[29]. Due to the many limitations associated with current gait analysis systems, prosthetists typically rely on observational gait analysis when performing dynamic alignment of a prosthetic device[30]. Therefore, having access to routine quantitative gait analysis could assist prosthetists with dynamic alignment during clinical visits.

## Methods

### Data collection

*Mobile acquisition and wearable sensors*

Video and sensor data is acquired on an Android smartphone using a custom app to synchronize recording from both modalities (Fig. 1). The mobile phone is mounted on a 3-axis gimbal (DJI Osmo Mobile 2) to improve the

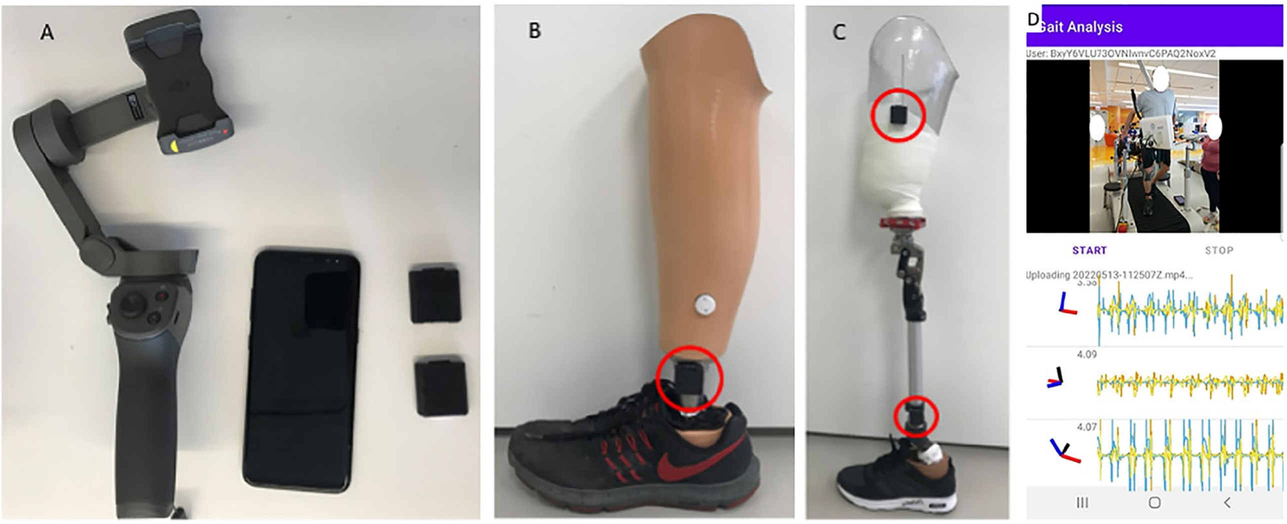
A
B
C
D
Gait Analysis
User: BxyY6VLU73OVNlwnvC6PAQ2NoxV2
START
STOP
Uploading 20220513-112507Z.mp4...
3.58
4.09
4.07

**Figure 1.** (**A**) Android cell phone, two wearable sensors and gimbal. (**B**) One sensor placed on the shank of a definitive transtibial prosthesis with adhesive Velcro®. (**C**) Two sensors placed on the shank/thigh of a diagnostic transfemoral. (**D**) Screenshot from gait analysis app, with video in the top panel and sensor data in the bottom panel.

video stability when following subjects during ambulation. The video was obtained in portrait orientation at 1080 × 1920 resolution at 30 frames per second.

The wearable sensors are a custom design and have previously been described[31,32]. For the purpose of this study, the wearable sensor data was used solely for validation. The sensors stream data from the IMU to the smartphone over Bluetooth Low-Energy (BLE). The IMU is an ICM-20948 and outputs 3-axis accelerometer and gyroscope data at 562.5 Hz. Magnetometer data is available but is not used in this study and we typically do not stream it in order to optimize the BLE bandwidth. The sensors can also acquire two channels of EMG data, but this was not used in these experiments. Prior to experiments, the magnetometer is calibrated in the location where data will be collected by rotating around each of the axes and accounting for hard and soft iron distortions. IMU data is fused on the sensor using a complementary filter to compute the 3D orientation, and the estimated orientation is streamed over BLE. Compared to our prior works using silicone-encapsulated wearable sensors attached to the skin, in this work the sensors were placed in a 3D-printed case with Velcro® on the outside for attachment to the body.

### *Clinical population and data annotation*

This study was approved by the Northwestern University Institutional Review Board. All methods were performed in accordance with the relevant guidelines and regulations. All participants gave written consent. Participants shown in images in this manuscript gave additional consent for the use of their images and videos in scientific publications and their faces were masked. Video of gait and other activities was obtained from a convenience sample of lower-limb prosthesis users seen in an outpatient prosthetics clinic or participating in outpatient physical therapy at Shirley Ryan AbilityLab. For each subject, we recorded their age, height, level, and (bi)laterality of amputation, the etiology of the amputation, the type of prosthetic components, and their Medicare Functional Classification Level (K-level). Walking data was collected in either prosthetic or therapy clinics. In the case of therapy, videos were obtained as participants performed their usual therapy.

Sensors were placed on both the shank and thigh of the prosthetic limb(s), with up to four sensors used in the case of individuals with bilateral amputations. They were attached either with double-sided Velcro® attached to the prosthetic pylon, and socket (for transfemoral amputees) or with a Velcro® strap around the thigh. They were placed laterally, with the IMU Z axis pointed laterally and the X axis pointed down. In this work, only the data from the shank IMU was used to detect the prosthetic limb swinging.

After data collection was complete, videos were annotated with the following categories by the author: the activity being performed (e.g., overground walking, treadmill walking, running, other therapeutic tasks), the view (frontal, sagittal, or a mixture), and the subjective accuracy of keypoint tracking of the prosthetic limb with a 3-point Likert scale ranging from 1 (poor) to 3 (good). Specifically 1 indicated the keypoints do not locate the prosthetic joints and tracking is frequently inaccurate, 2 indicated they locate the joint but are intermittently inaccurate, and 3 indicates the locate the joint well throughout the video. The individual performing the manual annotation was the main author and they were blinded to the keypoint confidence predictions prior to annotation. Whether the prostheses were visible or occluded by clothing was also annotated for each video. Time points when the participant entered and exited 10-m areas indicated by tape on the ground were also recorded to compute ground truth velocity.

### *Velocity annotation*

Tape was placed on the ground at 10-m spacing in locations where subjects would typically walk such as the hallway in the prosthetic clinic and in multiple locations in the therapy gyms. The start and end times when subjects completed a straight overground walk between these markers were retrospectively annotated. This provided both ground truth measurements of their walking velocity and identified specific video segments where participants were walking, as the collected data contained a mixture of activities.

In this work, we focus our analysis on video segments where individuals are performing above-ground walking in a forward direction during the time window where they were walking between the two pieces of tape, which we refer to as timed walking segments. We focus on the 10-m annotation periods both to determine the accuracy of the system when computing gait speed and because these were segments where individuals were known to be walking and were in view of the camera.

## Data processing

### *Pose pipe*

The input to the gait transformer is a series of 3D keypoint locations. To obtain these, the video was processed with PosePipe[25], a human pose estimation pipeline based on DataJoint[33] that simplifies running cutting-edge HPE algorithms on video. The steps used in the pipeline include (1) a tracking algorithm[34] to compute bounding box tracks for all people in the scene followed by (2) manually annotating the bounding box for the subject of interest undergoing gait analysis. (3) Then we perform top-down 2D keypoints detection in each frame using the MMPose toolbox[35], specifically using an HRNet[36] trained on the COCO dataset[37] using distribution aware relative keypoints[38], (4) the 2D keypoint trajectories are then lifted to 3D joint locations[39].

We also used DeepLabCut (DLC)[40] to train a custom 2D keypoint detector for a subset of prosthesis users with videos where keypoint detection was performing poorly. This is completed by manually annotating both intact and prosthetic hip, knee and ankle joints. We computed the 2D keypoints on those same videos with this model and replaced any prosthetic joints computed by MMPose with the estimates from DLC. This corrected set of keypoints was then passed to the lifting step and then to the gait transformer.

We flagged any frames as clipped whenever any of the keypoints of the leg came within 10 pixels of the edge of the screen.

### *Gait transformer*

The sequence of 3D joint location was mapped onto the relative timing of four gait events (right and left foot contact and toe off) as well as the pelvis velocity using the Gait Transformer[24]. This is trained on a large dataset of walking videos with synchronous marker-based motion capture and force plate data from our clinical gait laboratory. For training, data was aligned in the sagittal direction using the medial orientation of the pelvis. It also outputs sagittal plane joint kinematics including foot position relative to the pelvis, foot velocity, and hip and knee angles, all of which we do not focus on in this work. We refer the reader to the[24] for details of the architecture and training of the gait transformer. Compared to that work, we retrained the gait transformer and excluded bilateral elbows and wrists as we found occasionally the use of assistive devices (i.e., canes or crutches) in a less common pattern would trigger false detection of steps.

The Gait Transformer was applied to the lifted 3D joint trajectories. Training samples from the gait laboratory are typically only a few gait cycles long and we found that it did not generalize to inference on much longer sequences. Videos acquired in the clinic were much longer than a few gait cycles, so we applied it on a sliding window of 90 frames (3 s), which covers at least one complete stride for the majority of subjects. For each position of the sliding window, we used the middle output other than the beginning and end where we used the corresponding half of the sequence to pad the output.

### *Sensor processing*

In order to validate the accuracy of the gait transformer event timing of the prosthetic limb, we used gyroscope data from the wearable sensor on the prosthetic shank to detect the prosthetic-limb swing phase. Sensor data is timestamped to the smartphone time. Gyroscope data were sampled at a nominal sampling rate of 562.5 Hz. The Android system time of each Bluetooth packet is also stored and linear regression is used to calibrate the sensor timebase against the Android time, typically with an updated sampling rate of 1–2 Hz different than the nominal value.

We also noted that the video start timestamp showed some latency compared to the sensor timestamps. This has been resolved in more recent versions of our smartphone application with an API that acquires a more precise, per-frame timestamp. We used the hip and knee sagittal plane angles from the gait transformer to adjust for this timing error by finding the offset that minimized the mean squared error between the gyroscope on the shank and the change in that angle computed from the gait transformer outputs. This was typically around 170 ms.

We detected a prosthetic limb swing from the gyroscope on the prosthetic shank. With sensors placed on the lateral side of the shank, the z-axis is roughly aligned to be perpendicular to the sagittal plane. We applied an 8th-order low-pass Chebyshev filter to the gyroscope with a cutoff of 35 Hz. All negative values were zeroed out and a median filter of 360 ms was applied to deglitch a few strides where participants caught their toe and a brief reversal was seen in the gyroscope midtrace. These deglitched positive segments were identified as swings with the time the sign became negative identified as the start and end of swing periods.

### *Sensor versus video cadence*

We compared the cadence estimated with the gait transformer over 10-m walking segments to that computed from the sensors. We computed the cadence from video by averaging $\dot{\omega}$ over the timed walking segment and converting from strides in rad/s to steps/min (i.e. $c_v = \frac{120}{2\pi(t_e - t_s)} \int_{t_s}^{t_e} \dot{\omega}(t)\,\partial t$). We computed cadence from the sensors using the average stride time of the prosthetic limb side over the steps in the time window: $c_s = 120/\frac{s_j - s_i}{j-i}$, where $i$ and $j$ index the first and last sensor swing time, respective, that fall within the timed walking period.

### *Matching video and sensor foot contacts*

We also compared the time when the end of swing from the gyroscopes to the foot contact time from the gait transformer. This analysis only reflects a bound, because detecting when the prosthetic shank stops rotating forward in swing (i.e., when the gyroscope swaps signs) approximates the end of swing time but is not the actual time the foot contacts the ground and will typically occur slightly before true foot contact (Fig. 2). For each

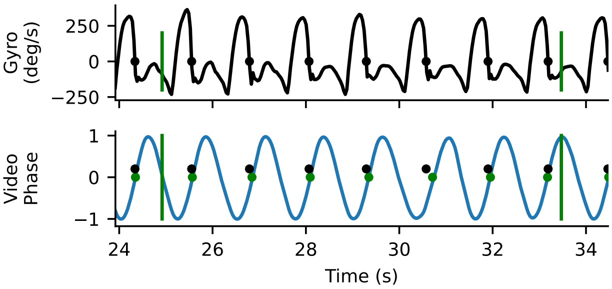

**Figure 2.** Example gyro and video timing information. Top trace is the z-component of the gyroscope mounted on the tibial shank. The time point where the gyroscope goes from positive to negative (where heel stops rotating forward relative to knee) is identified as a close proxy for foot contact. Bottom plot shows the sin component of the quadrature output for heel strike, with the positively directed zero crossing marker. There is a close correspondence between detected foot contacts from the gyro (blue) and the video (yellow). Vertical bars correspond to the annotated boundaries of the 10-m walk test.

walking segment, we computed the offset that produced the closest matches between the end of prosthetic leg swings detected by the sensors and the time of prosthetic foot contact estimated from the Gait Transformer and Kalman filter (typically about 100ms). After this we measured the foot contact detection accuracy as the residual timing error between the offset sensor times and the detected video times. We also measured the fraction of events that were detected with a window of 500 ms.

## Results

### Participant demographics

From 19 participants, we annotated 231 timed walking segments during level walking, with 79 in the frontal plane, 67 in the sagittal plane and the remainder either oblique or changing. When restricted to only the frontal plane, timed walking segments were obtained from a total of 16 participants (Table 1).

### Usability of the system

Our system made it easy to collect gait data in clinical situations, including physical therapy and prosthetic appointments. Wearable sensors took less than 30 seconds to apply and remove and would require no setup time if not using sensors for validation studies. The sensors connect to the cell phone through our Android app, so acquiring data is as easy as pressing the video record button in our app. Currently, running the analysis pipeline requires technical skills, but we hope to fully automate this in the future. Data was easily and routinely obtained from subjects in both therapy and prosthetic appointments. This was particularly true for the frontal view, but obtaining clear sagittal views in hallways was often challenging due to space limitations. No subjects withdrew from the study. Example visualizations from the system are shown in Figs. 3 and 4.

### 2D keypoint accuracy

Our 2D keypoint detection used an MMPose model, which is trained on the COCO dataset and primarily contains able-bodied individuals. We found that it frequently generalized to prosthesis users, but not always, and that quality of 2D keypoint tracking varied across subjects. In some cases, the detector would eschew localizing the prosthetic joint of the subject of interest and track the corresponding joint of a nearby therapist. While 2D keypoint detectors, including MMPose, output a per-joint confidence estimate, this has not been systematically tested for prosthesis users. For each video, we manually annotated our subjective estimate of tracking quality from 1 (poor) to 3 (good). For each timed walking segment, we also computed the average confidence that MMPose reported for the prosthetic ankle. As the 2D keypoint algorithms have primarily been trained on able-bodied individuals, we first wanted to analyze the accuracy of the keypoint confidence values on prosthetic limbs compared to manual annotation. Figure 5a shows the histogram of ankle keypoint qualities, which demonstrates that the confidence estimates from the keypoint detector align with our manual annotation. In fact, the interquartile ranges of ankle confidence conditioned on each of our annotated qualities were nearly non-overlapping (IQR for keypoint quality 1: [0.44,0.60]; 2: [0.59,0.80], 3: [0.77,0.87]).

| Age | Gender | Height (cm) | Weight (kgs) | Level | Side | Etilogy | K-Level | Frontal |
|---|---|---|---|---|---|---|---|---|
| 50 | Male | 183 | 101 | TT | Left | Trauma | 3 | + |
| 45 | Female | 167 | 71 | TT | Right | Infection | 3 | + |
| 40 | Male | 186 | 102 | TT | Left | Infection | 3 | + |
| 60 | Male | 178 | 84 | TT | Left | Infection | 3 | + |
| 65 | Male | 176 | 120 | TF | Bilateral | Trauma | 3 | + |
| 55 | Female | 162 | 68 | TT | Right | Vascular | x | + |
| 45 | Male | 180 | 96 | TT | Left | Trauma | 3 | + |
| 25 | Male | 172 | 70 | TF | Left | Sarcoma | 3 | + |
| 75 | Female | 152 | 79 | TT | Left | Vascular | 3 | + |
| 75 | Male | 184 | 95 | TF | Right | Sarcoma | 3 | + |
| 60 | Male | 173 | 91 | TT | Right | Sarcoma | 3 | + |
| 35 | Female | 158 | 74 | HD | Right | Sarcoma | 3 | – |
| 60 | Male | 167 | 79 | TF | Right | Trauma | 3 | – |
| 65 | Male | 177 | 55 | TT | Left | Infection | 3 | + |
| 70 | Female | 162 | 85 | TT | Bilateral | Vascular | 2 | – |
| 20 | Male | 185 | 75 | KD | Left | Trauma | 3 | + |
| 75 | Male | 177 | 95 | TF | Left | Infection | 2 | + |
| 50 | Male | 175 | 118 | TT | Right | Infection | 3 | + |
| 65 | Male | 190 | 90 | TT | Bilateral | Vascular | 3 | + |

**Table 1.** Demographic information of subjects, type and etiology of amputation, laterality and K-level. Age is rounded to nearest 5 years. *TT* transtibial, *TF* transfemoral, *HD* hip-disarticulation, *KD* Knee-disarticulation. The K-level marked ‘x’ was an individual referred from an external site without a documented K-level. The last column indicates the subjects without frontal views of their 10mWT.

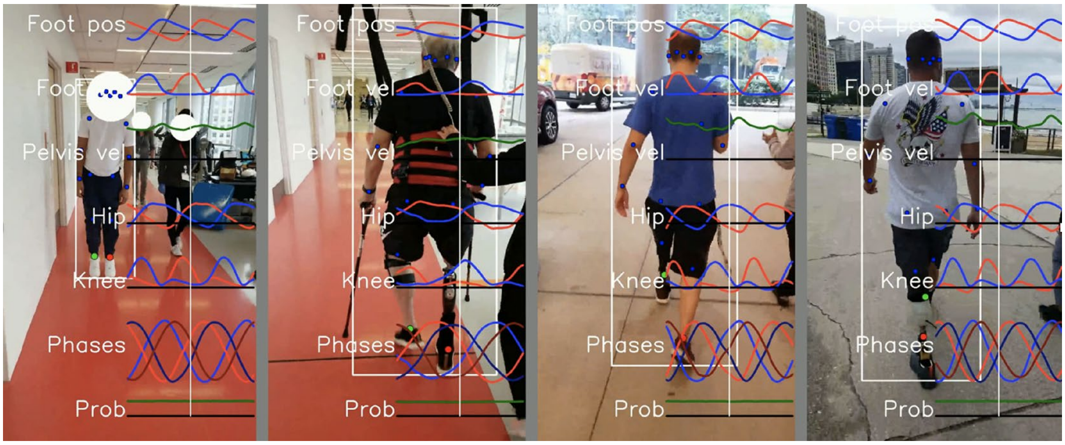

**Figure 3.** Example visualizations from the gait transformer. Keypoints on the ankle become colored red when in contact with the ground. Overlaid traces show the foot position, foot velocity, pelvis velocity, hip and knee angles, and the quadrature encoded gait timing.

After validating that the ankle joint confidence is valid in prosthesis users, we then conditioned on the prosthetic level and whether clothing was covering the prosthesis or not, which revealed two trends (Fig. 5b). First, ankle tracking was worse for bilateral prosthesis users compared to unilateral prosthesis users. Secondly, among unilateral prosthesis users, ankle tracking was worse for those with higher levels of amputation. Ankle tracking was particularly poor in the case of a prosthesis user participant with a bilateral transfemoral amputation and another with a hip disarticulation.

In comparison to the MMPose model trained on COCO[37], our DLC model that was specifically trained on our prosthetic user was able to perform much better, Fig. 6. Note that we did not test the generalization of this model to new users, and all prosthesis users we analyzed with our DLC model had 20 frames manually annotated and were included in the training dataset.

## Viewpoint sensitivity and clipping

We also classified segments as clipped or not if any of the 2D keypoints of the leg hit the edge of view on more than 1% of the frames, because we found the transformer was sensitive to the errors this produced. This occurred in none of the 79 frontal views, in 16 of the 67 sagittal views, and in 7 of the 85 mixed or oblique views. This was due to the somewhat limited space in clinical settings, where in hallways it can be hard to track far enough to the side of a person of interest to frame them with room for error.

## Velocity and cadence accuracy

We compared the velocity estimated from the video acquired in the frontal plane with the gait transformer to the velocity computed from the manually annotated times as participants walked over a 10-meter interval between the tape on the ground. We found the gait transformer velocity for videos acquired in the frontal plane was quite accurate compared to ground truth, with a correlation (r) of 0.95 and a mean absolute error (MAE) of 0.14 m/s (Fig. 7a). We also found that the cadence from the gait transformer was a close fit to the sensor data with an MAE of 4.8 steps/min, with most of this error coming from outliers (Fig. 7e).

We repeated this analysis, excluding segments where the average prosthetic ankle quality was less than 0.7 (Fig. 7b,f). Notably, we found this removed almost all the error estimating cadence (r = 0.998, MAE = 1.0). The accuracy of the estimated velocity also improved slightly (MAE = 0.13 m/s). We then used our custom prosthetic keypoint detector trained in DLC to replace the prosthetic joints in the excluded segments (Fig. 7c,g). We found this improved cadence detection for most sessions, but in particular, there were still several outliers for one individual with a slower cadence compared to other individuals.

We found the performance was quite poor when tested on video acquired in the sagittal plane (Fig. 7d,h). This was relatively unsurprising since the gait transformer is trained on video acquired in the frontal plane. We attempted to improve the viewpoint invariance by augmenting the training process by randomly rotating the lifted 3D keypoints. We found that while it resulted in a tighter correlation with the outputs, both in the gait laboratory validation dataset and on the prosthetic data it resulted in a bias that underestimated velocity at faster speeds.

## Step time accuracy

Over the frontal view timed walking segments, we detected and matched 889 prosthetic foot contact events and were unable to detect only 12 of these events. For each walking segment, we also computed the mean absolute error of the residuals. The average of this over all sessions was 72 ms. For sessions with good ankle tracking, we detected 490 foot contact events and missed only 1 event, with an average error over sessions of 45 ms.

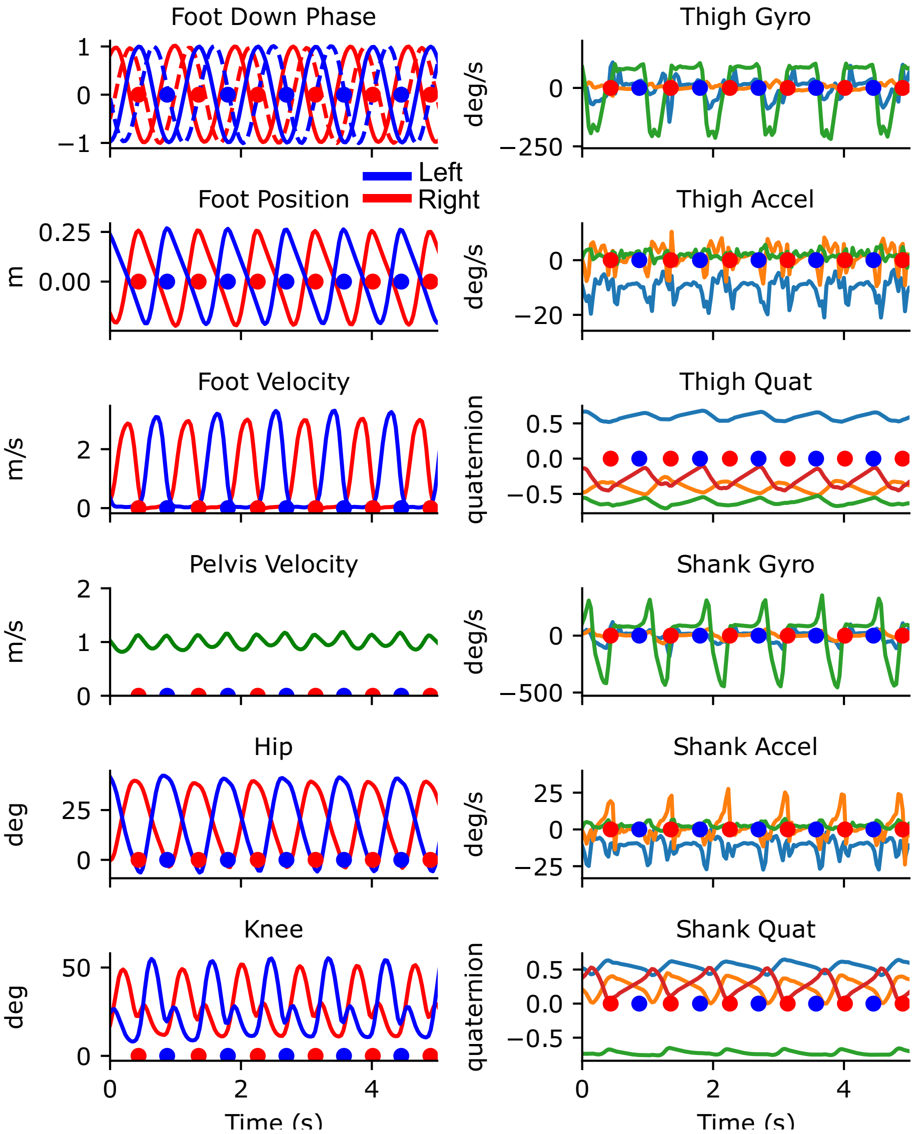

**Figure 4.** Traces from 5 s of walking from the portable system. The left column shows the outputs from the Gait Transformer. Traces from the left leg are shown in blue and from the right leg are shown in red. The additional dashed traces in the foot-down plot correspond to the phase-shifted quadrature encoding. The right column shows the raw sensor data from the thigh and shank. For gyro and accel plots, the three colors correspond to the x, y, and z IMU axis. For the quaternion plots, there are four colors for the individual quaternion components. Blue and red dots indicate the foot down events on the left and right, respectively.

## Discussion

Previously, we trained an algorithm for video-based gait analysis on a large clinical gait laboratory dataset of paired videos and motion capture data[24]. While the algorithm validated well on this dataset, there are differences between the training data and the intended application of the algorithm that could result in poor generalization. These include recording portrait videos while walking with the patient through a therapy clinic where other people are present. Our training dataset also had a predominance of children, who are most commonly analyzed in these laboratories. For any clinical application of AI, it is critical to evaluate the external validity (or out-of-domain generalization in machine learning parlance) for the intended use case.

In this study, we evaluated the performance of our gait transformer when tested on prosthesis users walking in therapy or outpatient clinic. This was a powerful stress test for our algorithm, as there are several properties of prosthesis users that might cause a failure to generalize including the visual appearance of prostheses and prosthetic gait patterns. This study also highlighted the power of an interpretable pipeline, with understandable features at multiple stages, such as 2D keypoint detection accuracy or failures to detect a single step. In development, this aspect of the pipeline was extremely important as it enables identifying and alleviating points of failure throughout the different stages of the pipeline to ensure we can trust the outputs from the transformer and make adjustments when necessary to improve accuracy.

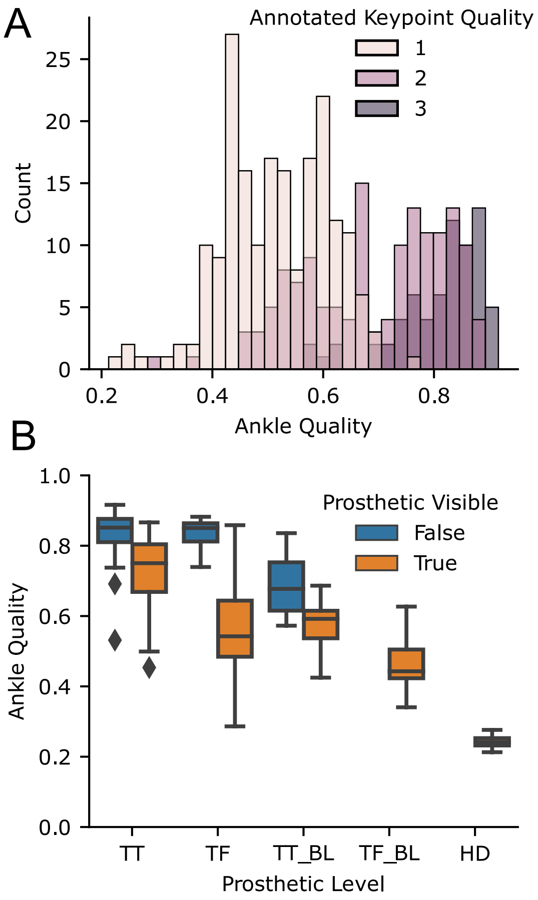

**Figure 5.** Accuracy of ankle keypoints for prosthesis users with off-the-shelf 2D keypoint detection algorithm. (**A**) shows the histogram of average ankle qualities for videos stratified by manual annotation of quality (with 3 being the best), showing that the estimated quality corresponded with our annotation. The (**B**) shows the average ankle quality stratified by prosthetic level and whether the prostheses were covered by clothing. *TT* transtibial, *TF* transfemoral, *HD* hip-disarticulation, *BL* bilateral.

Large datasets are a critical driving force of AI algorithms and have enabled impressive advances in HPE in recent years. However, public HPE datasets, like COCO[37], primarily contain able-bodied individuals. Perhaps unsurprisingly, we found that a 2D keypoint detector trained on COCO did a poor job detecting the ankles of some prosthesis users. This was most pronounced with more proximal or bilateral amputations. This is most likely due the greater visual difference between these limbs and joints compared to the able-bodied individuals in the dataset, resulting in worse out-of-domain generalization of the algorithms. In these cases, tools like DLC[40] make it relatively easy to train a custom keypoint detector. However, using DLC is still labor intensive as it requires manual annotation of each video. Future work will look to improve 2D keypoint detection for prosthetic limbs using self-supervised-learning. Towards this goal, we recently developed a markerless motion capture system and validated it on prosthesis users[41,42] Importantly, this work also speaks to the need for more work on AI fairness for people with disabilities[43,44].

When we removed the videos with poor ankle tracking quality and only included videos where the 2D keypoints were accurately localized, we found that our algorithm performed well on videos of prosthesis users for video acquired in the frontal plane. Specifically, we had a mean absolute error for the velocity of 0.13 m/s and for the cadence of 1 step/min. Whether this is sufficiently accurate ultimately depends on the clinical question. One study found the minimally clinically important difference for walking velocity of prosthesis users as 0.21 m/s[45], which is greater than our algorithm. However, for older adults, it has been suggested a small meaningful change in walking speed is 0.05 m/s[46]. Our results also do not indicate whether analyzing longer segments of walking would reduce the error, or whether this arises from a bias for given individuals that would persist over longer recordings.

The requirement to record video in the frontal plane to obtain accurate results is a limitation of our current approach. However, given the difficulties obtaining good videos in sagittal planes while walking in hallways and therapy gyms with other people present, it is also the most convenient approach for its intended use setting. An argument for sagittal videos is that they should enable more accurate estimates of many important sagittal plane kinematics. Our system outputs many of these and performs well on the training data but testing the external validity of these outputs is important future work. Specific to prosthesis users, testing the external validity of

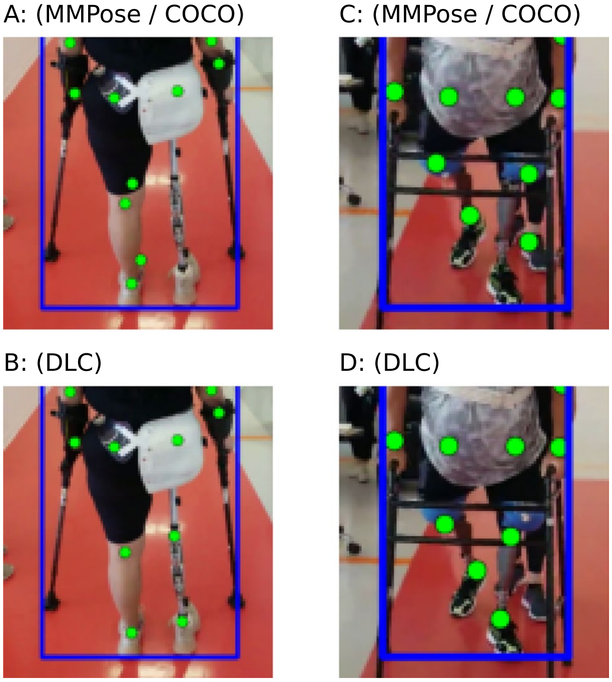

**Figure 6.** Examples of poor keypoint detection using pretrained algorithm (top row) that were corrected by training with DLC (bottom row).

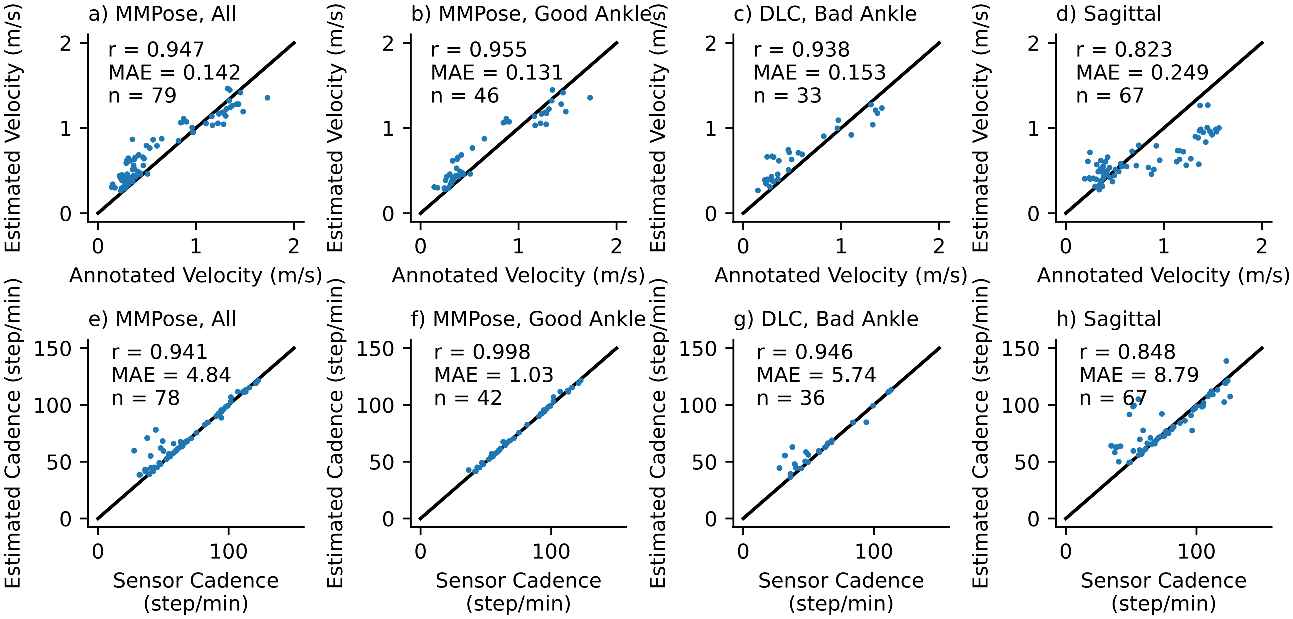

**Figure 7.** Accuracy of gait transformer for velocity (top row) and cadence (bottom row) under different conditions (columns). Text insets note the correlation coefficient (r), mean absolute error (MAE), and number of walking segments (n). The first column is all videos in the frontal plane. The second column is those where the average ankle detection quality was greater than 0.7. The third column is the excluded segments reprocessed with DLC. The last are videos acquired in the sagittal plane.

joint kinematics will be an area of challenge as prosthetic components attempt to mimic anatomic motion but do not move in an exact manner as able-bodied joints. As the gait transformer was primarily trained on individuals with intact limbs, this may affect the ability to accurately estimate prosthetic joint kinematics.In this study we validated the accuracy of the gait transformer's velocity, cadence and step timing measures as these were measures we could easily validate with wearable sensors in the clinic. However, our long-term goal is to develop a system that can accurately output several spatiotemporal and kinematic gait parameters to allow clinicians to further quantify the quality of gait. Further validation to determine the external validity of our portable gait analysis system to measure other spatiotemporal and kinematic gait parameters in prosthesis users is a high priority, which will leverage our recently developed markerless motion capture system[41,42].

Several other groups have looked at the ability to estimate gait parameters from monocular video. Stenum and colleagues analyzed sagittal videos from a dataset[47] of 32 healthy individuals walking down a walkway a fixed distance from a mounted camera using OpenPose, and showed they could accurately estimate gait event timing, sagittal plane joint angles, step length and walking velocity[20]. A pipeline similar to ours using 2D keypoints followed by lifted 3D joints was used with height-informed skeletal refinement prior to extracting gait parameters of healthy individuals walked towards a stationary camera and produced accurate estimates of step timing, length, and walking velocity[21]. Similarly to our work, both of these computed features on individual gait cycles, but in comparison, they were not externally validated on clinical populations or on video acquired in the community or clinic. A neural network can also be trained to directly map 2D keypoint trajectories from the video onto average gait parameters, and such an approach has been used on gait laboratory data from children[22] and subsequently we showed a similar approach with stroke survivors[23]. However, this approach does not allow for examining the cycle-by-cycle variability of gait parameters. Finally, by triangulating keypoints detected from multiple cameras it is possible to estimate more accurate 3D keypoint locations and perform inverse kinematics to fit biomechanical models to them[16,17,41,42]. OpenCap is a particularly portable version of this that only requires two iPhones[19].

There are several future directions that we anticipate will improve this system. One is better fusion with additional modalities of data, including the camera depth channel and inertial measurements from the wearable sensors. In this work, we have focused purely on monocular video as this is the most widely available modality. We are also enthusiastic about integrating physics-based modeling in the inference process[48–51] and ways to combine this with self-supervised learning[52]. It is also important to make this system easier to use in a higher-throughput manner. One need is to automate the annotation step to robustly identify the subject of interest. A potential solution is through the use of QR codes placed on the participant during data collection which would allow the computer vision to identify the person of interest without requiring manual annotation. In this work, we used the periods where 10-meter walk tests were annotated to select what to analyze, but this is only a small fraction of the data we acquired. Utilizing activity recognition to identify when the subject is walking, using the detected pose to identify viewpoint, and the Kalman error to determine when walking is being reliably tracked could help automate analysis of the remainder of the data. We recently developed a 3D lifting algorithm that produces calibrated confidence estimates of the joint locations, and integrating this could also help determine the trustworthiness of outputs[53]. Finally, there are many other clinically meaningful gait parameters available in our dataset that we could train the gait transformer to output and validate, including step width and center of mass.

## Data availability

The raw video datasets generated and/or analysed during the current study are not publicly available due to videos from clinical settings with faces visible. The 2D and 3D keypoint trajectories and the annotated velocities and cadences will also be made available from the corresponding author on reasonable request. Code for the gait transformer and the trained weights are available at https://github.com/peabody124/GaitTransformer.

Received: 17 January 2023; Accepted: 30 January 2024
Published online: 15 February 2024

## References

1. Moore, J. L. *et al.* A core set of outcome measures for adults with neurologic conditions undergoing rehabilitation: A clinical practice guideline. *J. Neurol. Phys. Ther.* **42**, 174–220. https://doi.org/10.1097/NPT.0000000000000229 (2018).
2. Richard, J., Levine, D. & Whittle, M. *Whittle's Gait Analysis* 5th edn. (Elsevier, 2012).
3. Hamacher, D., Singh, N. B., Van Dieën, J. H., Heller, M. O. & Taylor, W. R. Kinematic measures for assessing gait stability in elderly individuals: A systematic review. *J. R. Soc.* **8**, 1682–1698. https://doi.org/10.1098/rsif.2011.0416 (2011).
4. Prasanth, H. *et al.* Wearable sensor-based real-time gait detection: A systematic review. *Sensors* **21**, 2727. https://doi.org/10.3390/s21082727 (2021).
5. Weygers, I. *et al.* Inertial sensor-based lower limb joint kinematics: A methodological systematic review. *Sensors* **20**, 673. https://doi.org/10.3390/s20030673 (2020).
6. Picerno, P. 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 (2017).
7. Rast, F. M. & Labruyère, R. Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments. *J. NeuroEng. Rehabil.* **17**, 148. https://doi.org/10.1186/s12984-020-00779-y (2020).
8. Klöpfer-Krämer, I. *et al.* Gait analysis: Available platforms for outcome assessment. *Injury* **51**(Suppl 2), S90–S96 (2020).
9. Zheng, C. *et al. Deep Learning-Based Human Pose Estimation: A Survey* (2020).
10. Liu, W. & Mei, T. Recent advances of monocular 2D and 3D human pose estimation: A deep learning perspective. *ACM Comput. Surv.* **2022**, 3524497. https://doi.org/10.1145/3524497 (2022).
11. Muhammad, Z.-U.-D., Huang, Z. & Khan, R. A review of 3D human body pose estimation and mesh recovery. *Digit. Signal Process.* **128**, 103628. https://doi.org/10.1016/j.dsp.2022.103628 (2022).
12. Nakano, N. *et al.* Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. *Front. Sports Active Living* **2**, 50. https://doi.org/10.3389/fspor.2020.00050 (2020).
13. Pagnon, D., Domalain, M. & Reveret, L. Pose2Sim: An end-to-end workflow for 3D markerless sports kinematics—Part 1: Robustness. *Sensors* **21**, 6530. https://doi.org/10.3390/s21196530 (2021).
14. Needham, L. *et al.* The development and evaluation of a fully automated markerless motion capture workflow. *J. Biomech.* **144**, 111338. https://doi.org/10.1016/j.jbiomech.2022.111338 (2022).
15. Kanko, R. M., Laende, E., Selbie, W. S. & Deluzio, K. J. Inter-session repeatability of markerless motion capture gait kinematics. *J. Biomech.* **121**, 110422. https://doi.org/10.1016/j.jbiomech.2021.110422 (2021).
16. Kanko, R. M. *et al.* Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. *J. Biomech.* **122**, 110414. https://doi.org/10.1016/j.jbiomech.2021.110414 (2021).

17. Kanko, R., Laende, E., Davis, E., Selbie, W. S. & Deluzio, K. J. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *J. Biomech.*https://doi.org/10.1101/2020.12.10.420075 *(2020).*
18. McGuirk, T. E., Perry, E. S., Sihanath, W. B., Riazati, S. & Patten, C. Feasibility of Markerless motion capture for three-dimensional gait assessment in community settings. *Front. Hum. Neurosci.* **16**, 867485 (2022).
19. Uhlrich, S. D. *et al.* OpenCap: 3D human movement dynamics from smartphone videos. *PLoS Comput. Biol.* **19**, e1011462. https://doi.org/10.1101/2022.07.07.499061 (2022).
20. Stenum, J., Rossi, C. & Roemmich, R. T. Two-dimensional video-based analysis of human gait using pose estimation. *PLOS Comput. Biol.* **17**, e1008935. https://doi.org/10.1371/journal.pcbi.1008935 (2021).
21. Azhand, A., Rabe, S., Müller, S., Sattler, I. & Heimann-Steinert, A. Algorithm based on one monocular video delivers highly valid and reliable gait parameters. *Sci. Rep.* **11**, 14065. https://doi.org/10.1038/s41598-021-93530-z (2021).
22. Kidziński, Ł *et al.* Deep neural networks enable quantitative movement analysis using single-camera videos. *Nat. Commun.* **11**, 1–10. https://doi.org/10.1038/s41467-020-17807-z (2020).
23. Lonini, L. *et al.* Video-based pose estimation for gait analysis in stroke survivors during clinical assessments: A proof-of-concept study. *Digit. Biomark.* **6**, 9–18. https://doi.org/10.1159/000520732 (2022).
24. Cotton, R. J., McClerklin, E., Cimorelli, A., Patel, A. & Karakostas, T. Transforming Gait: Video-Based Spatiotemporal Gait Analysis. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 115–120. https://doi.org/10.1109/EMBC48229.2022.9871036 (2022).
25. Cotton, R. J. PosePipe: Open-Source Human Pose Estimation Pipeline for Clinical Research. arXiv:2203.08792 [cs, q-bio] (2022).
26. Krajbich, J. I., Pinzur, M. S., Potter, B. K. & Stevens, P. M. *Atlas of Amputations and Limb Deficiencies: Surgical, Prosthetic, and Rehabilitation Principles* (American Academy of Orthopaedic Surgeons, 2016). arXiv:NkN2xwEACAAJ.
27. Wong, C. K. *et al.* Exercise programs to improve gait performance in people with lower limb amputation: A systematic review. *Prosthet. Orthot. Int.* **40**, 8–17 (2016).
28. Highsmith, M. J. *et al.* Gait training interventions for lower extremity amputees: A systematic literature review. *Technol. Innov.* **18**, 99–113 (2016).
29. Gard, S. A. Use of quantitative gait analysis for the evaluation of prosthetic walking performance. *J. Prosthet. Orthot.*https://doi.org/10.1097/00008526-200601001-00011 *(2006).*
30. Brinkmann, P. M. & Stevens, J. T. Clinical Considerations of Observational Gait Analysis. In *Atlas of Amputations and Limb Deficiencies Surgical, Prosthetic, and Rehabilitation Principles* 81–95 (2016).
31. Cotton, R. J. & Rogers, J. Wearable Monitoring of Joint Angle and Muscle Activity. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)* 258–263. https://doi.org/10.1109/ICORR.2019.8779538 (IEEE, 2019).
32. Cotton, R. J. Kinematic Tracking of rehabilitation patients with markerless pose estimation fused with wearable inertial sensors. *IEEE 15th International Conference on Automatic Face & Gesture Recognition* (2020).
33. Yatsenko, D. *et al. DataJoint: Managing big scientific data using MATLAB or Python*. https://doi.org/10.1101/031658 (2015).
34. Wojke, N., Bewley, A. & Paulus, D. *Simple Online and Realtime Tracking with a Deep Association Metric*. arXiv:1703.07402 [cs]. arXiv:1703.07402 (2017).
35. Contributors, M. *OpenMMLab Pose Estimation Toolbox and Benchmark* (2020).
36. Sun, K., Xiao, B., Liu, D. & Wang, J. Deep high-resolution representation learning for human pose estimation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 5686–5696. https://doi.org/10.1109/CVPR.2019.00584 (2019).
37. Lin, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. In *Computer Vision - ECCV 2014 Lecture Notes in Computer Science* (eds Fleet, D. *et al.*) 740–755 (Springer International Publishing, 2014).
38. Zhang, F., Zhu, X., Dai, H., Ye, M. & Zhu, C. Distribution-Aware Coordinate Representation for Human Pose Estimation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 7091–7100. https://doi.org/10.1109/CVPR42600.2020.00712 (2020).
39. Liu, J., Rojas, J., Liang, Z., Li, Y. & Guan, Y. A Graph Attention Spatio-temporal Convolutional Network for 3D Human Pose Estimation in Video (2020). arXiv:2003.14179.
40. Mathis, A. *et al.* DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, 1. https://doi.org/10.1038/s41593-018-0209-y (2018).
41. Cotton, R. J. *et al.* Markerless Motion Capture and Biomechanical Analysis Pipeline. In *IEEE International Consortium for Rehabilitation Robotics*, (arXiv). arXiv:2303.10654.
42. Cotton, R. J. *et al.* Improved Trajectory Reconstruction for Markerless Pose Estimation. In *45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. arXiv:2303.02413.
43. Guo, A., Kamar, E., Vaughan, J. W., Wallach, H. & Morris, M. R. *Toward Fairness in AI for People with Disabilities: A Research Roadmap*. *arXiv* (2019).
44. Trewin, S. *AI Fairness for People with Disabilities: Point of View*. arXiv:1811.10670 [cs] (2018).
45. Carse, B., Scott, H., Davie-Smith, F., Brady, L. & Colvin, J. Minimal clinically important difference in walking velocity, gait profile score and two minute walk test for individuals with lower limb amputation. *Gait Posture* **88**, 221–224. https://doi.org/10.1016/j.gaitpost.2021.06.001 (2021).
46. Perera, S., Mody, S. H., Woodman, R. C. & Studenski, S. A. Meaningful change and responsiveness in common physical performance measures in older adults. *J. Am. Geriatr. Soc.* **54**, 743–749. https://doi.org/10.1111/j.1532-5415.2006.00701.x (2006).
47. Kwolek, B. *et al.* Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition. *Multimed. Tools Appl.* **78**, 32437–32465. https://doi.org/10.1007/s11042-019-07945-y (2019).
48. Yuan, Y., Wei, S.-E., Simon, T., Kitani, K. & Saragih, J. SimPoE: Simulated character control for 3D human pose estimation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 7155–7165, https://doi.org/10.1109/CVPR46437.2021.00708 (2021).
49. Xie, K. *et al. Physics-based Human Motion Estimation and Synthesis from Videos*. arXiv:2109.09913 [cs] (2021).
50. Shimada, S., Golyanik, V., Xu, W. & Theobalt, C. *PhysCap: Physically Plausible Monocular 3D Motion Capture in Real Time*. arXiv:2008.08880 [cs] (2020).
51. Shi, M. *et al.* MotioNet: 3D Human motion reconstruction from monocular video with skeleton consistency. *ACM Trans. Graph.* **40**, 1–15 (2020).
52. Gong, K. *et al. PoseTriplet: Co-evolving 3D Human Pose Estimation, Imitation, and Hallucination under Self-supervision*. arXiv:2203.15625 (2022).
53. Pierzchlewicz, P. A., Cotton, R. J., Bashiri, M. & Sinz, F. H. *Multi-hypothesis 3D human pose estimation metrics favor miscalibrated distributions*. arXiv:2210.11179 (2022).

## Acknowledgements

This work was generously supported by the Research Accelerator Program of the Shirley Ryan AbilityLab. We would like to thank Emoonah McClerklin for help with data collection.

## Author contributions

R.J.C. and A.C. designed the study, collected the data and analyzed the data. T.K. and A.P. provided guidance on the development and validation on the gait transformer algorithm. T.K. provided the dataset for training the gait tranformer algorithm and R.J.C. trained it on this data. All authors were involved in the preparation of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-53217-7.

**Correspondence** and requests for materials should be addressed to R.J.C.