The Most Discriminant Components of Force Platform Data for Gait Based Person Re-identification

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-NC-SA 4.0

SUBMISSION DATE / POSTED DATE

27-09-2021 / 28-09-2021

CITATION

Duncanson, Kayne; Thwaites, Simon; Booth, David; Abbasnejad, Ehsan; Robertson, William; Thewlis, Dominic (2021): The Most Discriminant Components of Force Platform Data for Gait Based Person Re-identification. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16683229.v1>

DOI

[10.36227/techrxiv.16683229.v1](https://doi.org/10.36227/techrxiv.16683229.v1)

The Most Discriminant Components of Force Platform Data for Gait Based Person Re-identification

Kayne Duncanson, Simon Thwaites, David Booth, Ehsan Abbasnejad, William S.P. Robertson, and Dominic Thewlis

Abstract—Walking gait data measured using force platforms is a promising means for person re-identification in authentication and surveillance scenarios. We aimed to determine the most discriminant components of force platform data using a two-stream Convolutional Recurrent Neural Network (KineticNet). Each network in the two-stream architecture extracts features pertaining to a single stance phase and then these features are fused to represent the entire gait cycle. Over two sessions, ground reaction forces (F_x , F_y , F_z), moments (M_x , M_y , M_z), and center of pressure coordinates (C_x , C_y) were acquired from 118 participants as they walked our laboratory five times at preferred speed. For each participant, one sample per session was reserved for network validation, one sample per session was reserved for network testing, and the rest were reserved for network training. KineticNet’s performance was evaluated using both individual component and multi-component inputs before ablation studies were conducted on its architecture. F_z was the most discriminant individual component, and re-identification using F_z , F_y , and C_y together was the most accurate overall at 96.02%. These results warrant further investigation into the utility of force platforms as an accessory or alternative to video cameras for gait based person re-identification.

Index Terms—Deep neural network, recognition, deep learning, biometric, gait analysis, data fusion, ground reaction force, moment, center of pressure, force platforms, force plates, signal processing, locomotion.

I. INTRODUCTION

GAIT is the characteristic pattern of movements employed during locomotion [1]. Bipedal walking, the most common mode of gait in humans, can be observed in discrete cycles that each contain functional phases and sub-phases [2] (Fig. 1). Based on the observation that each gait cycle contains identifiable features, walking gait has been proposed as a means for person re-identification (*re-ID*) in authentication and surveillance scenarios [3, 4, 5]. The *re-ID* task involves matching features of gait from a known identity to features of gait from an unknown identity. The main advantage of using gait features over traditional structural features (e.g. face, iris, or fingerprint) is that gait can be measured using unobtrusive and discreet sensors [6]. For a gait

based person *re-ID* system to be effective, it must represent features of the walking pattern that differ between individuals yet remain consistent within individuals over time [7]. This is particularly challenging because gait is an emergent rather than a prescribed property; that is, its manifestation depends on numerous personal, environmental, and task constraints [8].

Many gait based person *re-ID* systems have been developed in the last twenty years, though few have been adopted for widespread use [9]. Perhaps this is due to an excessive reliance on video cameras for gait measurement [10]. These devices capture the motion of the body’s outer surface, from which the motion of the limbs must be inferred. This means that the human body and its anatomical landmarks must be clearly visible throughout the gait cycle (or at least key phases therein). The most popular public video datasets contain samples collected in controlled settings with adequate lighting and no occlusions or background movement [11, 12, 13]. In many surveillance and forensics tasks, however, individuals must be re-identified in dynamic outdoor environments where these conditions cannot be satisfied [14]. Even if the body is sufficiently visible, kinematic measures cannot be reliably extracted with changes in camera placement between measurement instances [15]. In fact, the impact of changes in camera view angle between measurement instances is currently the biggest limitation to gait based person *re-ID* [16].

Force platforms could augment video cameras for gait based person *re-ID* because they measure kinetics rather than appearance. These instruments can be embedded into a ground surface to measure ground reaction forces (*GRFs*) along three orthogonal axes as well as corresponding moments about each axis. The origin of the *GRF* vector, referred to as the center of pressure (*COP*), can also be calculated from components of the *GRF* and moment [17]. If multiple force platforms are arranged in series, *GRFs*, moments, and *COP* coordinates can be measured during both left and right footsteps (i.e. stance phases) to represent the entire gait cycle (Fig 1). There are several hypothesized advantages of force platforms over video cameras for gait based person *re-ID*, namely:

- Load cells (contained in force platforms) are tactile sensors, so their measures do not depend on visibility;
- Force platform data represents 3D kinetics whereas video footage represents 2D kinematics. Thus, the features encoded in force platform data are higher in dimensionality compared to those encoded in video footage;
- All of the features encoded in force platform data are related to the *re-ID* task whereas only a small portion of

K. Duncanson, S. Thwaites, and D. Thewlis are with the Adelaide Medical School, The University of Adelaide, SA, 5000 AUS (e-mails: kayne.duncanson@adelaide.edu.au; dominic.thewlis@adelaide.edu.au; simon.thwaites@adelaide.edu.au).

W.S.P. Robertson is with the School of Mechanical Engineering, The University of Adelaide, SA, 5000 AUS (email: william.robertson@adelaide.edu.au).

E. Abbasnejad is with The Australian Institute for Machine Learning, The University of Adelaide, SA, 5000 AUS (email: ehsan.abbasnejad@adelaide.edu.au).

Fig. 1. Components of the ground reaction force (GRF), moment, and center of pressure (COP) across the walking gait cycle measured with one force platform per foot (low-pass filtered to assist with interpretation). Functional phases of the gait cycle are separated by vertical dashed lines. Each component has unique characteristics and there are differences between left and right stance. N = Newton; m = meter.

the features encoded in video footage are related to the *re-ID* task [6, 18];

- Force platform data is stored in 1D time series format whereas video footage is stored as a set of 2D pixel arrays. This means that force platform data is more

compact and less (computationally) expensive to process compared to video footage.

Certain features of the *GRF* and *COP* have been associated with structural and physiological attributes such as anthropometrics, skeletal alignment, age, sex, pregnancy status, injury status, and disease status [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]. However, certain features of these parameters have also been associated with task and environmental constraints such as walking speed, footwear, load carriage, and surface terrain [24, 30, 31, 32, 33, 34]. Hence, for a force platform based *re-ID* system to be effective, it should highlight features that depend on consistently unique attributes (such as those pertaining to body structure) and discard features that depend on transiently unique attributes (such as those pertaining to clothing and footwear).

Deep neural networks *DNNs* are particularly well suited to the task of person *re-ID* because they automatically learn the most discriminative representations of complex data [35]. These architectures have achieved state-of-the-art performance in gait based person *re-ID* using video footage, though their utility for gait based person *re-ID* using force platform data remains largely unexplored [36, 37]. Deep 1D convolutional neural networks (*CNNs*) have shown great promise for extracting features from *GRF* and *COP* signals acquired during a single session [38, 39]. Yet, it is unknown whether these features are useful over much longer time periods. We recently conducted an experiment in our own laboratory where force platform data was acquired over two sessions separated by 3–14 days (Section II). A convolutional-recurrent neural network (*CRNN*) was trained and validated using data from 76 participants to investigate the utility of time series data augmentation [40]. The results showed that individual components of the *GRF* were not equally discriminative, and the fusion of all components led to higher accuracy than using any one of them in isolation. Nevertheless, network performance was inferred based on validation accuracy (a test set was not included), so the results could be biased toward the validation set. Based on these studies, no firm conclusions can be drawn regarding the relative utility of *GRFs*, moments, and *COP* coordinates for gait based person *re-ID*.

There were three aims to the present study. First, we wished to rate each component of force platform data in terms of its distinctiveness. Based on this, we wished to determine if representing multiple components together is more effective than representing the most distinctive component on its own. Finally, we wished to confirm the overall utility of *KineticNet* by assessing its robustness to architectural ablations. The completion of these aims will facilitate a more streamlined approach to the development of force platform based person *re-ID* systems by allowing researchers to concentrate their efforts on the most pertinent information. Below are the main contributions of this work:

- We introduce a two-stream *CRNN* named *KineticNet* that takes time normalized sequences from each stance phase of the gait cycle as an input and generates a single feature vector as the output (referred herein as the gait embedding). Following this approach, features pertaining

to each stance phase of the gait cycle can be extracted in a sequence independent manner;

- We comprehensively evaluate different components of the *GRF*, moment, and *COP* to determine which are the most useful for the *re-ID* task. We found that F_z was the most discriminant component and fusing additional components with F_z did not significantly improve *re-ID* accuracy;
- We present a freely available force platform dataset, named *ForceID*, that is tailor made for research on gait based person *re-ID*. The dataset was collected in a repeated measures design consisting of two sessions separated by a minimum of 3 and a maximum of 14 days. It contains force platform data and demographic measures from 111 healthy individuals who walked the length of our laboratory at preferred walking speed in personal clothes and footwear.

The remainder of this paper is organised as follows: Section II describes the dataset; Section III details the signal pre-processing method and architecture of *KineticNet*; Section IV outlines the experimental protocol; Section V presents the results; Section VI includes a discussion of key findings; and finally, Section VII concludes the paper.

II. FORCE PLATFORM DATASET

The dataset used in this study is from a repeated measures experiment that was approved by the Human Research Ethics Committee (approval no. H-2018-009). The experiment was conducted at The University of Adelaide gait analysis laboratory. Participants were recruited from the general population if they were in good health and had no known neurological disorder that could affect their gait (e.g. stroke [41], cerebral palsy [29], Parkinson’s disease [42], etc.) In total, 118 individuals volunteered to participate in the experiment (57% M, age 27 ± 7 years, mass 71.41 ± 16.52 kg, height 1.72 ± 0.10 m). Each participant attended two sessions, with a minimum of 3 and a maximum of 14 days between each session. They wore their own personal clothing and footwear, though footwear had to be consistent between sessions. Age, sex, mass, height, and footwear type were recorded at the start of each session (footwear was photographed for future reference). Participants completed five trials of walking along the length of the laboratory at preferred walking speed (10 m). Embedded into the floor surface at the center of the laboratory, two OPT400600-HP force platforms (Advanced Mechanical Technology Inc., USA) measured *GRFs* and moments during left and right stance. These measures, along with the calculated *COP* coordinates, were acquired through Vicon Nexus at 2000 Hz. Two Vicon Vue video cameras (Vicon, Oxford, UK) that were time synchronized with the force platforms captured motion from front and side views at 50 Hz with 1920×1080 pixel resolution so that foot contact regions could be inspected. Trials were included for further analysis if each foot contacted within the sensing area of each force platform, as identified from the video footage.

In total, 1156 trials were included in the dataset comprising 8–10 trials per participant. From each trial there were eight 1D

Fig. 2. Comparison between raw and pre-processed versions of F_z and C_y for two random participants. In this case, the inter-participant variation is easier to interpret in the pre-processed versions.

time series signals per stance side: F_x , F_y , F_z , M_x , M_y , M_z , C_x , and C_y . *GRFs* were expressed in units of N, moments were expressed in units of N mm, and *COP* coordinates were expressed in units of mm. The *ForceID* dataset contains data from 111 participants because seven participants did not consent to their data being made publicly available. It can be found at https://adelaide.figshare.com/articles/dataset/ForceID_Dataset/14482980 (DOI:10.25909/14482980).

III. METHODS

A. Problem Formulation

We formulated the *re-ID* task as a $1 : K$ matching problem where K represents the number of identities in the dataset. Suppose we have a query sample x_q with an associated query label y_q and N_{ref} samples in the reference set $\{(x_i, y_i = i) | i = 1, \dots, N_{ref}\}$. The task was to predict y_q by calculating the similarity between x_q and each of the samples in the reference set.

B. Signal pre-processing

The pre-processing method transformed the signals to a common reference frame to highlight their unique characteristics (Fig 2). Specifically, the following steps were completed:

- 1) *GRF* and moment signals were trimmed to include only the ground contact period as defined by the first and last frame at which F_z was above 50 N. *COP* signals were trimmed a further 5% on each end to avoid potential inaccuracies at low force values (e.g. C_x from force platform two in Fig. 1).
- 2) All signals were filtered with a 4th order bi-directional low-pass Butterworth filter (cut-off= 30 Hz) to minimize high frequency noise.
- 3) The free moment (M_f) was calculated as:

$$M_f = M_z - ((C_x \times F_y) - (C_y \times F_x))$$

M_f is the force couple exerted by the foot on the ground about a vertical axis originating at the *COP* [43]. This parameter provides an indication of torsion between the pelvis and the lower limb during stance [44]. It may be useful for gait based person *re-ID* but has not yet been implemented. M_f was calculated for the middle 90 % of stance (based on how much the *COP* signals had been trimmed), so the remaining 10 % was zero padded;

- 4) All signals (including M_f) were interpolated to a fixed length of 400 frames, given that the shortest ground contact period was 440 frames.
- 5) *COP* coordinates were re-scaled to start at (0,0).

C. Network Architecture

One major contribution of this work is the two-stream *CRNN* architecture *KineticNet*, which is the main component of the *re-ID* system presented. In this section, the architecture of *KineticNet* is described, including the two-stream structure, the convolutional layers used to extract local kinetic features from input sequences, the long-short term memory (*LSTM*) layers used to detect temporal dependencies, and the fully-connected (*FC*) layers used to synthesise key features.

1) *Two-stream Network*: *KineticNet* extracts features from both left and right stance phases to generate a representation of the entire gait cycle (Fig. 3). This means that *KineticNet* detects features that do not depend on footstep order, which is important because footstep order often changed between trials in our experiment. The same phenomenon can be expected in many practical applications where only one gait cycle worth of kinetic data is available. We wished to implement the *CRNN* architecture as it has recently been applied with success in related tasks such as video based person *re-ID* and gait phase recognition [45, 46]. The input to each *CRNN* is a length normalized sequence of shape $s_b \times C_{in} \times 400$, where s_b is batch size. As can be seen in Fig. 3, features of kinetic data from each stance phase are extracted separately before the feature vectors are summed element-wise to produce the gait embedding. *LSTM* units are able to retain memory across numerous time steps, though short sequences are less likely to cause vanishing gradients during back-propagation [47, 48]. As such, convolutional layers were implemented before the *LSTM* layer to extract local kinetic features and reduce the sequence length [49] (Fig. 4).

2) *Convolutional layers*: The first operation in each convolutional layer is a 1D convolution where kinetic features are detected within overlapping time windows [50]. The convolution filter parameters (kernel k , stride s , and padding p) are set at $k = 3$, $s = 1$, and $p = 1$ such that the input and output lengths are equal and minimal padding is required. This ensures that the convolution operation is efficient and all of the information encoded in the input is retained. The number of features doubles with each convolution except the first, where the number of output features is 32 and thus the ratio depends on C_{in} .

After each convolution is batch normalization, Exponential Linear Unit (*ELU*) activation, and local average pooling. Batch normalization reduces internal covariate shift, particularly in

Fig. 3. Illustration of the global architecture of *KineticNet* containing two Convolutional Recurrent Neural Networks (*CRNNs*). An input sequence from left stance (x_L) and the input sequence from right stance (x_R), each comprising 400 time steps (t), are processed separately to produce two vectors with 128 features (f) each. Subsequently, these (feature) vectors are summed element-wise to produce the gait embedding (e).

cases where the input contains multiple scales [51]. As can be seen in Fig. 1, each channel of kinetic data has a different scale, so batch normalization is expected to help when multiple components are used as the input. Local average pooling reduces the sequence length in a way that minimizes information loss [50]. The pooling parameters (kernel k , stride s , and padding p) for the first three layers are $k = 2$, $s = 2$, and $p = 0$, whereas the pooling parameters for the final convolutional layer are $k = 3$, $s = 3$, and $p = 0$ as determined experimentally ($k, s = \{2, 3, 4\}$). After the convolutional layers the sequence length has reduced from 400 to 16 and the number of features has increased from C_{in} to 256.

3) *LSTM and FC layers*: The *LSTM* layer identifies temporal relationships between kinetic features. The input to this layer is shaped $s_b \times 16 \times 256$, meaning that there are 16 time steps and each time step is represented a 256 element feature vector. A magnified view of an *LSTM* cell is shown in Layer 5 of Fig. 4. The i^{th} input for a given time step t (x_i^t) is transformed to a hidden state (h_i^t) that depends on the previous hidden state (h_{i-1}^{t-1}) as well as the current cell state (c_i^t). The hidden state can only incorporate short term contextual information, whereas the cell state can incorporate long term contextual information if necessary. The hidden size (s_h), representing the number of *LSTM* cells at each time step, is set to 256 as determined experimentally ($s_h = \{64, 128, 256, 512\}$). After the *LSTM* layer is the *FC* layer that takes the hidden state vector from each time step (shape $bs \times 16 \times 256$) and returns a 128 element feature vector for each time step (shape $bs \times 16 \times 128$). Only the feature vector from the final time step

Fig. 4. Illustration of the local Convolutional Recurrent Neural Network architecture comprising four convolutional layers followed by one long-short term memory (LSTM) layer and one fully-connected (FC) layer. Between convolution (Conv) and average pooling (Pool) operations is batch normalization and Exponential Linear Unit activation (respectively). Grey shading indicates zero padding. C_{in} = number of input channels, x = LSTM input, c = cell state, h = hidden state, i = index, t = time.

is returned based on the assumption that relevant information is transferred across all 16 time steps (Fig. 4 - Layer 6).

D. Loss Function and Optimizer

KineticNet's loss was calculated using the triplet margin metric with an online hard negative sampling strategy [52]. Triplet loss has been found superior to classification and verification losses for the *re-ID* task in terms of performance and scalability [53]. In general, the loss L was formulated as follows:

$$L = \sum_i^N [\|f((x_L^a, x_R^a)) - f((x_L^p, x_R^p))\|_2^2 - \|f((x_L^a, x_R^a)) - f((x_L^n, x_R^n))\|_2^2 + \alpha]_+ \quad (1)$$

N represents the number of triplets selected on each forward pass, with each triplet containing an anchor (a) sample, a positive (p) sample, and a negative (n) sample. Based on equation 1, the goal is to minimise the distance (in feature space) between pairs of embeddings that belong to the same identity

and maximise the distance (in feature space) between pairs of embeddings that belong to different identities. Details on how the loss was implemented (including triplet selection) can be found in Section IV-C. Weights were optimized using the AMSGrad optimizer to guarantee convergence and preserve the learning rate [54].

IV. EXPERIMENTS

A. Overview

Three phases of experimentation were conducted to evaluate different components of force platform data and confirm the utility of *KineticNet* for gait based person *re-ID*. In phase 1, each individual component (represented by a single channel) was used as the input to *KineticNet* to determine its effect on *re-ID* accuracy. In phase 2, the top N_c components were concatenated along the channel dimension and used as the input to *KineticNet*, where $N_c = \{2, 3, \dots, 9\}$. This allowed us to determine if using multiple components leads to higher accuracy than using the most discriminant component in isolation. Finally, in phase 3, we conducted ablation studies

on *KineticNet* using the most effective input from phases 1 and 2 to quantify the significance of certain modules within its architecture.

B. Sample distribution and feature standardization

All of the experiments across phases 1, 2, and 3 were conducted according to the protocol outlined in Fig. 5. This protocol was designed to facilitate repeated runs of network training, validation, and testing using either a single component or multiple components as the input to *KineticNet*. Specifically, ten runs were conducted per input to account for stochastic effects due to random sample allocation. Each participant had n samples per component comprising n_{S1} samples from session one $\{(x_{iL}, x_{iR}, y_i = i) | i = 1, \dots, n_{S1}\}$ and n_{S2} samples from session two $\{(x_{iL}, x_{iR}, y_i = i) | i = \{1, \dots, n_{S2}\}$. Depending on the participant, n was 8–10, n_{S1} was 3–5, and n_{S2} was 4–5. The n samples were distributed into training, validation, and test subsets using the following method:

- 1) The set of n_{S1} samples from session one were randomly shuffled and the set of n_{S2} samples from session two were randomly shuffled;
- 2) One sample from each session was allocated to the test subset and one sample from each session was allocated to the validation subset. The remaining samples were allocated to the training subset. This meant that all three subsets contained samples from both sessions.

For each component, there were 684 samples for training ($\sim 60\%$), 236 samples for validation ($\sim 20\%$), and 236 samples for testing ($\sim 20\%$). The features within each subset were standardized by removing the mean and scaling to unit variance. Validation and test subsets were standardized using the mean and standard deviation of training subsets.

C. Implementation details

Samples from each subset were loaded and processed in mini-batches to improve the generalization performance of *KineticNet* by introducing stochasticity [55]. A batch size of 64 was chosen as it led to the highest accuracy ($s_b = \{32, 64, 128\}$). Training samples were loaded using a weighted random sampler as the training set was imbalanced. During training, triplets were sampled based on the Euclidean distance matrix of all embeddings within each batch. Namely, the *hardest positive* input (x_L^p, x_R^p) and the *hardest negative* (x_L^n, x_R^n) input were sampled for each anchor input (x_L^a, x_R^a). The *hardest positive* input was an input from the same identity as the anchor that formed the maximum Euclidean distance in feature space. Conversely, the *hardest negative* input was the input from a different identity to the anchor that formed the minimum Euclidean distance in feature space. The hardest triplets in each batch can be thought of as moderate triplets across the entire training subset. The purpose of sampling moderate triplets was to allow *KineticNet* to learn from challenging examples (i.e. similar samples from different identities) without oversampling outliers [53]. Finally, the margin value α was set to 0.3 as determined experimentally ($m = \{0.1, 0.2, 0.3, 0.4\}$). In calculating the loss on the

Fig. 5. Diagram showing each step in the experimental protocol. C_{in} = number of input channels.

validation and test subsets, we wished to replicate a scenario where an individual has been enrolled to the reference set and then must be re-identified at a later date (3–14 days later in this case). Accordingly, only inputs from session two trials were used as anchors and only inputs from session one trials were used as positives or negatives. The *easiest positive* input and the *hardest negative* input were sampled for each anchor input. The *easiest positive* input was an input from the same identity as the anchor, other than the anchor itself, that formed the minimum Euclidean distance in feature space. The *hardest negative* input was defined in the same way as for training (see above).

The primary performance metric in these experiments was rank-1 accuracy; that is, the number of correct predictions over the total number of predictions. Regardless of the subset in question, a prediction was considered correct (for a given triplet) if the anchor-positive distance was smaller than the anchor-negative distance:

$$\|f((x_L^a, x_R^a)) - f((x_L^p, x_R^p))\|_2^2 < \|f((x_L^a, x_R^a)) - f((x_L^n, x_R^n))\|_2^2 \quad (2)$$

During training, the accuracy was calculated as an average value based on the number of batches processed per epoch. The weights that resulted in the highest validation accuracy were used to process the test subset. Time taken for training and validation, referred to as run time, was the secondary performance metric. This indicated the computational efficiency of *KineticNet* for a given input. Training and validation were conducted over 150 epochs at a learning rate of 0.0001 using an Intel(R) Core(TM) i9-9900K CPU (with 32GB RAM) and an NVIDIA GeForce RTX 2080 SUPER GPU. The AMSGrad optimizer was implemented with default hyper-parameters ($\beta = (0.9, 0.999)$, $\epsilon = 1 \times 10^{-8}$, weight decay = 0).

D. Statistical Analyses

Within each experimental phase, differences in test accuracy were compared using the independent samples t-test. Normality was assumed on the basis of D’Agostino and Pearson’s Skewness test because the Kurtosis test is invalid with fewer than 20 samples (in this case there were 10) [56, 57]. Equal variance was assumed if both groups were not skewed. Significance was determined using an alpha value of

0.01 and the family-wise error rate was controlled using the Holm-Bonferroni correction method [58].

V. RESULTS

A. Phase one – Individual components

Table I provides a summary of *KineticNet*'s performance using individual components as the input. The differences in test accuracy between F_z and F_y as well as F_y and C_y were significant ($p = 6.29 \times 10^{-4}$ and $p = 3.75 \times 10^{-6}$, respectively). C_y was 5.51% more accurate than F_x and 6.95% more accurate than M_x , though neither of these differences were statistically significant ($p = 0.09$ and $p = 0.06$, respectively). The difference between M_x and M_f was significant ($p = 1.44 \times 10^{-5}$), whereas the difference between M_z and M_y was non-significant ($p = 0.50$). All other differences in test accuracy were found to be significant (Supplementary Material - Table 1).

It took between 59.28 and 60.12 sec to train and validate *KineticNet* regardless of which component was used as the input. Throughout training, the accuracy and loss exhibited similar yet opposite trends for all components (Fig. 6). The same was true during validation aside from the first two epochs where the validation accuracy and loss both increased for all but F_z and F_y . Relative differences in accuracy and loss at the first epoch were largely indicative of relative differences in accuracy and loss at the final epoch. However, the differences were much larger on the validation subset compared to the training subset. This is also true for the maximum values, as can be seen in Table I. Depending on the component, the maximum training accuracy ranged from $99.97 \pm 0.05\%$ to $82.87 \pm 3.93\%$ (F_z vs. M_y , respectively), whereas the maximum validation accuracy ranged from $95.42 \pm 1.09\%$ to $40.85 \pm 4.37\%$ (F_z vs. M_y , respectively). Convergence characteristics also varied greatly depending on the component. As indicated by the slope of the lines on the left side of Fig. 6, F_z and F_y had fully converged by the end of training, with F_x close to full convergence. The other components appeared to require more training, however, their accuracy and loss on the validation subset had almost fully converged by epoch 120. All of the components except for F_z over-fitted the training subset by more than 11%, with M_z over-fitting the most at 43.84%.

B. Phase two – Multiple components

Table II provides a summary of *KineticNet*'s performance using the top N_c components from phase 1 (in terms of test accuracy) as the input, where $N_c = \{2, 3, \dots, 9\}$. Fusing F_z , F_y , and C_y led to the highest test accuracy at $96.02 \pm 1.26\%$. However, the 3.31% improvement in accuracy using F_z , F_y , and C_y compared to using F_z alone was not statistically significant ($p = 1.36 \times 10^{-2}$). Accuracy and loss curves are not shown as *KineticNet* had similar convergence patterns and achieved perfect training accuracy regardless of the input (Table II).

Fig. 6. The mean accuracy and loss of *KineticNet* during training and validation using individual components as the input. The large variation in accuracy and loss indicates that component selection is important in defining convergence properties.

Fig. 7. The mean accuracy and loss of different versions of *KineticNet* during training and validation using $\{F_z, F_y, C_y\}$ as the input. In the first 30 epochs, the variant without batch normalization (*BN*) and the variant without long-short term memory (*LSTM*) had considerably different convergence properties compared to the original.

C. Phase three – Ablations

F_z , F_y , and C_y were used as the input to *KineticNet* for ablation studies because their fusion led to the highest test accuracy across phases 1 and 2. Four ablations were conducted to determine how much *KineticNet* relies on key modules within its architecture:

- 1) Average pooling was swapped for max pooling in the convolutional layers. Max pooling is a common alternative to average pooling that can improve shift invariance in *CNNs* [59];
- 2) The *ELU* activation function was swapped for the rectified linear unit (*ReLU*) activation function in the convolutional layers. *ReLU* is the default activation function

TABLE I

PERFORMANCE METRICS FOR *KineticNet* USING INDIVIDUAL COMPONENTS AS THE INPUT (REPORTED AS MEAN $\pm$ STANDARD DEVIATION). ROWS ARE SORTED BY TEST (T_e) ACCURACY (A) IN DESCENDING ORDER. THE VERSION THAT RESULTED IN THE HIGHEST TEST ACCURACY IS BOLD. Tr = TRAIN, Va = VALIDATION.

Input	$max(A_{Tr})$ (%)	$max(A_{Va})$ (%)	A_{Te} (%)	Run time (sec)
F_z	99.97 ± 0.05	95.42 ± 1.09	92.71 ± 1.90	59.28 ± 0.34
F_y	99.90 ± 0.07	88.22 ± 2.69	85.00 ± 3.52	60.12 ± 0.09
C_y	98.14 ± 0.61	73.14 ± 3.92	70.85 ± 3.68	60.00 ± 0.09
F_x	99.61 ± 0.27	65.51 ± 2.86	65.34 ± 3.48	59.85 ± 0.14
M_x	94.41 ± 0.66	67.88 ± 4.48	63.90 ± 4.77	60.12 ± 0.13
M_f	89.91 ± 1.69	53.31 ± 3.80	50.42 ± 3.20	59.33 ± 0.19
M_z	86.81 ± 3.11	42.97 ± 5.99	39.75 ± 6.16	59.89 ± 0.17
C_x	82.32 ± 2.63	41.36 ± 3.05	36.61 ± 3.43	59.78 ± 0.13
M_y	82.87 ± 3.93	40.85 ± 4.37	35.85 ± 7.61	59.82 ± 0.13

TABLE II

PERFORMANCE METRICS FOR *KineticNet* WITH THE TOP N_c COMPONENTS FROM PHASE 1 (IN TERMS OF TEST ACCURACY) AS THE INPUT, WHERE $N_c = \{2, 3, \dots, 9\}$ (REPORTED AS MEAN $\pm$ STANDARD DEVIATION). FOR REFERENCE, THE FIRST ROW CONTAINS THE METRICS OBTAINED USING F_z ALONE. THE VERSION THAT RESULTED IN THE HIGHEST TEST (T_e) ACCURACY (A) IS BOLD. Tr = TRAIN, Va = VALIDATION.

Input	$max(A_{Tr})$ (%)	$max(A_{Va})$ (%)	A_{Te} (%)	Run time (sec)
F_z	99.97 ± 0.05	95.42 ± 1.09	92.71 ± 1.90	59.28 ± 0.34
$\{F_z, F_y\}$	100 ± 0.00	95.51 ± 1.52	95.59 ± 1.55	59.45 ± 0.31
$\{F_z, F_y, C_y\}$		97.80 ± 1.09	96.02 ± 1.26	59.84 ± 0.13
$\{F_z, F_y, C_y, F_x\}$		96.53 ± 1.71	95.59 ± 1.30	59.49 ± 0.15
$\{F_z, F_y, C_y, F_x, M_x\}$		95.85 ± 1.34	95.34 ± 0.57	59.50 ± 0.10
$\{F_z, F_y, C_y, F_x, M_x, M_f\}$		96.36 ± 1.01	94.41 ± 2.15	59.38 ± 0.16
$\{F_z, F_y, C_y, F_x, M_x, M_f, M_z\}$		95.85 ± 2.13	95.17 ± 1.65	59.08 ± 0.14
$\{F_z, F_y, C_y, F_x, M_x, M_f, M_z, C_x\}$		96.19 ± 1.71	93.64 ± 2.25	59.49 ± 0.31
$\{F_z, F_y, C_y, F_x, M_x, M_f, M_z, C_x, M_y\}$		95.85 ± 1.58	93.47 ± 1.20	59.10 ± 0.23

TABLE III

PERFORMANCE METRICS FOR *KineticNet* UNDER DIFFERENT ABLATION CONDITIONS USING $\{F_z, F_y, C_y\}$ AS THE INPUT (REPORTED AS MEAN $\pm$ STANDARD DEVIATION). ROWS ARE SORTED BY TEST (T_e) ACCURACY (A) IN DESCENDING ORDER. FOR REFERENCE, THE FIRST ROW CONTAINS THE METRICS FOR THE ORIGINAL VERSION. THE VERSION THAT RESULTED IN THE HIGHEST TEST ACCURACY IS BOLD.

Version	$max(A_{Tr})$ (%)	$max(A_{Va})$ (%)	A_{Te} (%)	Run time (sec)
Original	100 ± 0.00	97.80 ± 1.09	96.02 ± 1.26	59.84 ± 0.13
Max pooling	100 ± 0.00	96.78 ± 1.55	95.68 ± 1.63	60.49 ± 0.10
No batch normalization		95.17 ± 1.26	93.81 ± 2.43	51.07 ± 0.16
No <i>LSTM</i>		96.19 ± 2.01	93.39 ± 1.89	57.27 ± 0.33
<i>ReLU</i> activation		93.56 ± 2.15	93.31 ± 2.32	60.36 ± 0.14

for *DNNs* [60];

- 3) Batch normalization was not completed in convolutional layers. Previous research suggests that combining batch normalization with *ELU* may confer no additional benefit compared to using *ELU* alone [61]. This is because both batch normalization and *ELU* function to reduce internal covariate shift;
- 4) The *LSTM* and *FC* layers were replaced with two *FC* layers to determine whether the order of the kinetic features (throughout stance phase) was important. Batch normalization and an *ELU* activation were included after the first *FC* layer. Given that the output from layer 4 was shaped $s_b \times 16 \times 256$, the last two dimensions were flattened into a 4096 element feature vector. The number of output features in the first *FC* layer was set to 2048 and the number of output features in the final

FC layer was set to 1024 (to reduce the number of features at each layer by a factor of two). Even though the number of features in the gait embedding differed between this version and the original version (1024 vs. 128), we thought that it was more important to maintain the overall depth of the network.

Table III shows the performance metrics for each ablated version of *KineticNet* compared to the original. With respect to the test subsets, the original version was the most accurate whereas the version with *ReLU* activation instead of *ELU* activation was the least accurate ($96.02 \pm 1.26\%$ vs. $93.31 \pm 2.32\%$, respectively). However, none of the differences between ablated versions of *KineticNet* were statistically significant (Supplementary Material - Table 3).

Fig. 7 shows the mean accuracy and loss throughout training and validation for each ablated version compared to the

original version. Within the first 30 epochs, the version without the *LSTM* layer went from being 2.17 times as accurate on the training set to only 1.12 times as accurate. However, these differences in accuracy did not reflect on the validation subset. This version also started out with a loss 2.27 times higher, yet had a loss 0.18 times as high by epoch 30. On the validation subset, though, the loss was notably higher and fluctuated considerably throughout all 150 epochs. Finally, the version without batch normalization appeared to converge at a loss value of 0.28 on both training and validation subsets during the first 10 epochs. Aside from these exceptions, convergence characteristics were quite similar for all versions of *KineticNet*.

VI. DISCUSSION

The aims of this study were to evaluate different components of force platform data and confirm the utility of *KineticNet* for gait based person *re-ID*. The maximum test accuracy achieved across all experiments was $96.02 \pm 1.26\%$ using the combination of F_z , F_y , and C_y . However, using F_z alone led to a test accuracy of $92.71 \pm 1.90\%$. In previous studies it has been indicated that F_z is the most discriminant component of the *GRF*, followed by F_y and F_x (respectively) [40, 62]. This study is the first to show that F_z is the most discriminant component of all force platform data. It was also found that the inclusion of additional components along with F_z (at the input level) confers only a marginal increase in *re-ID* accuracy. Based on these findings, developers can focus their efforts towards the most discriminant components of force platform data and thus accelerate the development of gait based person *re-ID* systems.

The maximum accuracy obtained in this study is competitive with the previous benchmark for single session *re-ID* and forms the new benchmark for inter-session *re-ID* (using force platforms). In all but one previous investigation, data was acquired from each participant during a single session [10, 40, 63]. This may have reduced the variability of the data and in turn simplified the *re-ID* task (Table IV) [64]. The benchmark for single session *re-ID* was 97.38 % accuracy on 200 people, an additional 1.36 % compared to our best result on 118 people [65]. Still, this comparison is vague because the authors did not provide a detailed description of the experimental protocol or dataset characteristics. Finally, the initial benchmark for inter-session *re-ID* was set at 86.1 % using a subset of the current dataset (76 participants) and *KineticNet* superseded this benchmark by 9.92 % [40]. In summary, the present results are promising for the prospect of gait based person *re-ID* using force platforms and they should be used as a reference point moving forward.

The superiority of F_z over other components suggests that its pattern differs between individuals yet remains consistent within individuals over at least 3–14 days (i.e. the range included in this study). Previous studies have shown that key features of F_z have a smaller within-participant coefficient of variation (CV) compared to key features of F_y and F_x (respectively) (Table IV). With that being said, F_z appears to have a similar within-participant CV to C_y (in a single session). In our study, the use of F_z as the input to *KineticNet*

TABLE IV
PREVIOUSLY REPORTED COEFFICIENTS OF VARIATION (CVs) FOR INDIVIDUAL COMPONENTS OF KINETIC DATA. *DATA ACQUIRED USING PRESSURE SENSING INSOLES.

Study	Sample population	Number of sessions	Component	CV (%)
[71]	10 males	1	F_x	~ 15
			F_y	5–10
			F_z	< 5
[72]	10 males	1	F_x	~ 8 –10
			F_y	~ 5 –7
			F_z	< 5
[29]	8 boys 7 girls	3	F_x	23.24–51.95
			F_y	11.44–20.38
			F_z	7.05–15.50
[73]*	4 males 11 females	1	C_x	1.4–46.5
			C_y	0.3–3.3

resulted in 21.86 % more accuracy compared to the use of C_y . Perhaps this is because the CVs were calculated based on 2–3 manually pre-defined features whereas *KineticNet* relied on 128 automatically learnt features. Alternatively, the additional accuracy obtained using F_z may be explained by its uniqueness rather than its consistency.

The amount of force applied to ground during stance, and thus the resultant *GRF*, depends on descending neural commands as well as soft tissue properties [66]. F_z acts in the opposite direction to body weight and is regulated such that the vertical trajectory of the body center of mass (*COM*) forms a sinusoidal pattern [67]. In line with this, a significant amount of the variation in F_z between individuals can be explained by variations in body mass [68]. Previous studies on gait based person *re-ID* have found reductions of more than 25 % in *re-ID* accuracy using weight normalized *GRFs* [62, 69]. In one study, mass values extracted from *GRF* measures were 39 % accurate in identifying 62 children [70]. Using our dataset, a comparison of the mass measurements (obtained using standard scales) between sessions revealed that approximately 20 % *re-ID* accuracy could be obtained using mass alone. Overall, these findings suggest that most of the additional accuracy conferred by F_z may be attributable to variations in body mass.

One unexpected finding from this study is that the original version of *KineticNet* was only 2.63 % more accurate than the version without the *LSTM* layer. Perhaps this is because only the hidden state vector from the final time step was used to generate the gait embedding in the original version whereas the features from all time steps were used to generate the gait embedding in the version without the *LSTM* layer. Some useful information may have been lost by discarding features from 15 of the 16 time steps. Nevertheless, the version without the *LSTM* layer appeared to over fit the training data to a larger extent than the original version as can be seen from the convergence characteristics in Fig. 7. In order for firm conclusions to be drawn, a version of *KineticNet* with an *LSTM* layer that uses all *LSTM* hidden state vectors (to generate the gait embedding) should be compared to a version of *KineticNet*

without an *LSTM* layer that has a regularization component such as dropout between *FC* layers.

Findings from this study should be interpreted within the context of its limitations. Firstly, all participants were in good health and had no known neurological disorder that could affect their gait. Also, each participant's footwear was consistent between sessions and the dataset only includes preferred walking speed trials where each foot landed completely within the sensing area of each force platform. Lastly, participants were not allowed to walk while carrying objects in either session. These constraints cannot be applied in many real-world scenarios, particularly those pertaining to surveillance. Further experiments are required to determine the impact of varying footwear, walking speed, and carriage conditions on inter-session *re-ID* accuracy. On top of this, the utility of signal segmentation should be explored to facilitate the use of partial footstep data. In this study, samples were distributed into training, validation, and test subsets using a within-participant distribution method. This meant that *KineticNet* was trained and evaluated on different samples from the same participants. Reserving a separate set of participants for evaluation would allow us to determine the network's ability to generalise to new participants.

VII. CONCLUSION

In this paper we presented a novel *SCRNN* architecture (*KineticNet*) that re-identifies individuals based on their walking gait kinetics. We first evaluated the network using individual components of force platform data and found that F_z encodes the most characteristic features of gait, followed by F_y , and C_y (respectively). Next, we tested the utility of combining multiple components at the input level and found the fusion of F_z , F_y , and C_y to be most effective. Specifically, the version of *KineticNet* that relied on these three components was able to re-identify 118 participants with $96.02 \pm 1.26\%$ accuracy. Overall, these findings warrant further investigation into the utility of force platforms as an accessory or alternative to video cameras for gait based person *re-ID* in authentication and surveillance scenarios.

ACKNOWLEDGMENT

This work was supported by an Australian Government Research Training Program (RTP) Scholarship. Defence Science and Technology Group and the National Health and Medical Research Council (ID: 1126229) also funded elements of the work.

REFERENCES

- [1] J. Perry and J. R. Davids, "Gait analysis: normal and pathological function," *Journal of Pediatric Orthopaedics*, vol. 12, no. 6, pp. 815–815, 11 1992.
- [2] M. W. Whittle, *Gait Analysis: An Introduction*. Butterworth-Heinemann, 5 2014.
- [3] S. V. Stevenage, M. S. Nixon, and K. Vince, "Visual Analysis of Gait as a Cue to Identity," *Applied Cognitive Psychology*, vol. 13, no. 6, pp. 513–526, 12 1999.
- [4] M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang, and S. V. Stevenage, "Automatic Gait Recognition," in *Biometrics*, A. K. Jain, R. Bolle, and S. Pankanti, Eds. Boston, MA: Springer, 1999, pp. 231–249.
- [5] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. Springer Science & Business Media, 10 2007.
- [6] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 12, pp. 1505–1518, 12 2003.
- [7] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 1 2004.
- [8] K. M. Newell, R. E. A. Van Emmerik, and P. V. McDonald, "Biomechanical constraints and action theory," *Human Movement Science*, vol. 8, no. 4, pp. 403–409, 8 1989.
- [9] C. Wan, L. Wang, V. V. Phoha, and editors, "A survey on gait recognition," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 8 2018.
- [10] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Computer Vision and Image Understanding*, vol. 167, pp. 1–27, 2 2018.
- [11] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4. IEEE, 8 2006, pp. 441–444.
- [12] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1511–1521, 6 2012.
- [13] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition," *IPSI Transactions on Computer Vision and Applications*, vol. 10, no. 1, pp. 1–14, 12 2018.
- [14] J. Neves, F. Narducci, S. Barra, and H. Proença, "Biometric recognition in surveillance scenarios: a survey," *Artificial Intelligence Review*, vol. 46, no. 4, pp. 515–541, 12 2016.
- [15] A. Michelini, A. Eshraghi, and J. Andrysek, "Two-dimensional video gait analysis: A systematic review of reliability, validity, and best practice considerations," *Prosthetics and Orthotics International*, vol. 44, no. 4, pp. 245–262, 8 2020.
- [16] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 209–226, 3 2017.
- [17] D. A. Winter, A. E. Patla, and J. S. Frank, "Assessment of balance control in humans," *Med prog technol*, vol. 16, no. 1-2, pp. 31–51, 5 1990.

- [18] G. V. Veres, L. Gordon, J. N. Carter, and M. S. Nixon, "What image information is important in silhouette-based gait recognition?" in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 6 2004, pp. 2–2.
- [19] C. D. Stickley, S. N. Andrews, E. A. Parke, and R. K. Hetzler, "The effectiveness of scaling procedures for comparing ground reaction forces," *Journal of Biomechanics*, vol. 77, pp. 55–61, 8 2018.
- [20] L. Wong, A. Hunt, J. Burns, and J. Crosbie, "Effect of Foot Morphology on Center-of-Pressure Excursion During Barefoot Walking," *Journal of the American Podiatric Medical Association*, vol. 98, no. 2, pp. 112–117, 3 2008. [Online]. Available: <https://doi.org/10.7547/0980112>
- [21] N. Chockalingam, P. H. Dangerfield, A. Rahmatalla, E. N. Ahmed, and T. Cochrane, "Assessment of ground reaction force during scoliotic gait," *European Spine Journal*, vol. 13, no. 8, pp. 750–754, 12 2004.
- [22] C. C. Gao, J. S. Chern, C. J. Chang, P. L. Lai, and C. W. Lung, "Center of pressure progression patterns during level walking in adolescents with idiopathic scoliosis," *PLoS ONE*, vol. 14, no. 4, pp. –, 4 2019.
- [23] M. C. Chiu and M. J. Wang, "The effect of gait speed and gender on perceived exertion, muscle activity, joint motion of lower extremity, ground reaction force and heart rate during normal walking," *Gait and Posture*, vol. 25, no. 3, pp. 385–392, 3 2007.
- [24] M. C. Chiu, H. C. Wu, and L. Y. Chang, "Gait speed and gender effects on center of pressure progression during normal walking," *Gait and Posture*, vol. 37, no. 1, pp. 43–48, 1 2013.
- [25] J. R. Franz and R. Kram, "Advanced age affects the individual leg mechanics of level, uphill, and downhill walking," *Journal of Biomechanics*, vol. 46, no. 3, pp. 535–540, 2 2013.
- [26] M. C. Chiu, H. C. Wu, L. Y. Chang, and M. H. Wu, "Center of pressure progression characteristics under the plantar region for elderly adults," *Gait and Posture*, vol. 37, no. 3, pp. 408–412, 3 2013.
- [27] J. K. Lymbery and W. Gilleard, "The stance phase of walking during late pregnancy: Temporospatial and ground reaction force variables," *Journal of the American Podiatric Medical Association*, vol. 95, no. 3, pp. 247–253, 5 2005.
- [28] B. Pietrosimone, M. K. Seeley, C. Johnston, S. J. Pfeiffer, J. T. Spang, and J. T. Blackburn, "Walking Ground Reaction Force Post-ACL Reconstruction: Analysis of Time and Symptoms," *Medicine and Science in Sports and Exercise*, vol. 51, no. 2, pp. 246–246, 2 2019.
- [29] R. White, I. Agouris, R. D. Selbie, and M. Kirkpatrick, "The variability of force platform data in normal and cerebral palsy gait," *Clinical Biomechanics*, vol. 14, no. 3, pp. 185–192, 3 1999.
- [30] J. Nilsson and A. Thorstensson, "Ground reaction forces at different speeds of human walking and running," *Acta Physiologica Scandinavica*, vol. 136, no. 2, pp. 217–227, 6 1989.
- [31] D. J. Stefanyshyn, B. M. Nigg, V. Fisher, B. O'flynn, and W. Liu, "The Influence of High Heeled Shoes on Kinematics, Kinetics, and Muscle EMG of Normal Female Gait," *Journal of Applied Biomechanics*, vol. 16, no. 3, pp. 309–319, 8 2000.
- [32] Z. B. Barrons and G. D. Heise, "The impact of shod vs unshod walking on center of pressure variability," *Gait and Posture*, vol. 81, pp. 116–119, 9 2020.
- [33] X. A. Zhang, M. Ye, and C. T. Wang, "Effect of unilateral load carriage on postures and gait symmetry in ground reaction force during walking," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 13, no. 3, pp. 339–344, 6 2010.
- [34] S. A. Birrell, R. H. Hooper, and R. A. Haslam, "The effect of military load carriage on ground reaction forces," *Gait and Posture*, vol. 26, no. 4, pp. 611–614, 10 2007.
- [35] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 5 2015.
- [36] K. Sundararajan and D. L. Woodard, "Deep learning for biometrics: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–34, 5 2018.
- [37] A. S. Alharthi, S. U. Yunas, and K. B. Ozanyan, "Deep learning for monitoring of human gait: A review," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9575–9591, 7 2019.
- [38] H. Fabian, S. Lapuschkin, W. Samek, K. R. Muller, and W. Schöllhorn, "Explaining the unique nature of individual gait patterns with deep learning," *Scientific Reports*, vol. 9, no. 1, pp. 1–3, 2 2019.
- [39] P. Terrier, "Gait Recognition via Deep Learning of the Center-of-Pressure Trajectory," *Applied Sciences*, vol. 10, no. 3, pp. 774–774, 1 2020.
- [40] J. Yeomans, S. Thwaites, W. S. P. Robertson, D. Booth, B. Ng, and D. Thewlis, "Simulating Time-Series Data for Improved Deep Neural Network Performance," *IEEE Access*, vol. 7, pp. 131 248–131 255, 9 2019.
- [41] G. Chen, C. Patten, D. H. Kothari, and F. E. Zajac, "Gait differences between individuals with post-stroke hemiparesis and non-disabled controls at matched speeds," *Gait and Posture*, vol. 22, no. 1, pp. 51–56, 8 2005.
- [42] J. M. Hausdorff, M. E. Cudkowicz, R. Firtion, J. Y. Wei, and A. L. Goldberger, "Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease," *Movement Disorders*, vol. 13, no. 3, pp. 428–437, 5 1998.
- [43] J. P. Holden and P. R. Cavanagh, "The free moment of ground reaction in distance running and its changes with pronation," *Journal of Biomechanics*, vol. 24, no. 10, pp. 887–897, 1 1991.
- [44] T. Ohkawa, T. Atomi, K. Hasegawa, and Y. Atomi, "The free moment is associated with torsion between the pelvis and the foot during gait," *Gait and Posture*, vol. 58, pp. 415–420, 10 2017.
- [45] L. Wu, C. Shen, and A. V. D. Hengel, "Convolutional lstm networks for video-based person re-identification," *arXiv preprint*, vol. 1, no. 11, 6 2016.
- [46] D. Kreuzer and M. Munz, "Deep convolutional and LSTM networks on multi-channel time series data for gait phase recognition," *Sensors*, vol. 21, no. 3, pp. 789–

- 789, 1 2021.
- [47] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proceedings*. Presses universitaires de Louvain, 4 2015, pp. 89–94.
 - [48] M. Daniluk, T. Rocktäschel, J. Welbl, and S. Riedel, "Frustratingly Short Attention Spans in Neural Language Modeling," *ArXiv preprint*, 2017. [Online]. Available: <http://arxiv.org/abs/1702.04521>
 - [49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
 - [50] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, 3 2017.
 - [51] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning*, vol. 1. PMLR, 6 2015, pp. 448–456.
 - [52] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, 2015, pp. 815–823.
 - [53] A. Hermans, L. Beyer, and B. Leibe, "In Defense of the Triplet Loss for Person Re-Identification," *arXiv preprint*, 3 2017. [Online]. Available: <http://arxiv.org/abs/1703.07737>
 - [54] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," *arXiv preprint*, 4 2019.
 - [55] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," *arXiv preprint*, 4 2018.
 - [56] R. B. D'Agostino, A. Belanger, and J. R. B. D'Agostino, "A suggestion for using powerful and informative tests of normality," *American Statistician*, vol. 44, no. 4, pp. 316–321, 11 1990.
 - [57] R. B. D'Agostino, "An omnibus test of normality for moderate and large size samples," *Biometrika*, vol. 58, no. 2, pp. 341–348, 8 1971.
 - [58] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, vol. 6, no. 2, pp. 65–70, 1 1979.
 - [59] M. Ranzato, F. J. Huang, Y. L. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 6 2007, pp. 1–8.
 - [60] V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 1 2010.
 - [61] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," *arXiv preprint*, 11 2015.
 - [62] Z. M. Yao, X. Zhou, E. D. Lin, S. Xu, and Y. N. Sun, "A novel biometric recognition system based on ground reaction force measurements of continuous gait," in *3rd International Conference on Human System Interaction, HSI'2010 - Conference Proceedings*. IEEE, 5 2010, pp. 452–458.
 - [63] M. Derlatka, "Human gait recognition based on signals from two force plates," in *International Conference on Artificial Intelligence and Soft Computing*, vol. 7268 LNAI, no. PART 2. Berlin, Heidelberg: Springer, 4 2012, pp. 251–258.
 - [64] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, "The effect of time on gait recognition performance," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 543–552, 11 2011.
 - [65] M. Derlatka and M. Bogdan, "Ensemble kNN classifiers for human gait recognition based on ground reaction forces," in *2015 8th International Conference on Human System Interaction (HSI)*. IEEE, 6 2015, pp. 88–93.
 - [66] E. Ayyappa, "Normal human locomotion, Part 2: Motion, ground-reaction force and muscle activity," *Journal of Prosthetics and Orthotics*, vol. 9, no. 2, pp. 49–57, 4 1997.
 - [67] P. G. Adamczyk and A. D. Kuo, "Redirection of center-of-mass velocity during the step-to-step transition of human walking," *Journal of Experimental Biology*, vol. 212, no. 16, pp. 2668–2678, 8 2009.
 - [68] D. R. Mullineaux, C. E. Milner, I. S. Davis, and J. Hamill, "Normalization of Ground Reaction Forces," *Journal of Applied Biomechanics*, vol. 22, no. 3, pp. 230–233, 8 2006.
 - [69] M. D. Addlesee, A. H. Jones, F. Livesey, and F. Samaria, "The ORL active floor [sensor system]," *IEEE Personal Communications*, vol. 4, no. 5, pp. 35–41, 10 1997. [Online]. Available: <https://www.researchgate.net/publication/2514465>
 - [70] J. Jenkins and C. Ellis, "Using ground reaction forces from gait analysis: body mass as a weak biometric," in *International conference on pervasive computing*. Berlin, Heidelberg: Springer, 5 2007, pp. 251–267.
 - [71] G. Giakas and V. Baltzopoulos, "Time and frequency domain analysis of ground reaction forces during walking: an investigation of variability and symmetry," *Gait & Posture*, vol. 5, no. 3, pp. 189–197, 6 1997.
 - [72] K. Masani, M. Kouzaki, and T. Fukunaga, "Variability of ground reaction forces during treadmill walking," *Journal of applied physiology*, vol. 92, no. 5, pp. 1885–1890, 5 2002. [Online]. Available: <http://www.jap.org>
 - [73] Y. Wang and K. Watanabe, "Limb Dominance Related to the Variability and Symmetry of the Vertical Ground Reaction Force and Center of Pressure," *Journal of Applied Biomechanics*, vol. 28, no. 4, pp. 473–478, 8 2012. [Online]. Available: www.JAB-Journal.com

biomechanics, and soft biometrics.

Kayne Duncanson was born in Bega, NSW, Australia in 1996. He received a Bachelors degree (Hons.) in Exercise and Nutrition Sciences from The University of Queensland, QLD, Australia in 2019. In his Honours project he investigated the biomechanical effects of altered longitudinal foot stiffness using a passive elastic foot exoskeleton. He is currently pursuing the Ph.D. degree in Medicine at The University of Adelaide, SA, Australia, under the supervision of Assoc. Prof. D. Thewlis. His current research interests include deep learning, gait

processing, and machine learning. He is a member of the International Society of Biomechanics, and the Australian and New Zealand Orthopaedic Research Society.

Dominic Thewlis was born in Leeds, U.K. in 1982. He received the Ph.D. degree in biomechanics, in 2009, under the supervision of J. Richards from the University of Central Lancashire, U.K., where he was a Research Fellow from 2005 to 2009. He is currently an Associate Professor and National Health and Medical Research Fellow with Adelaide Medical School, The University of Adelaide. His research interests include biomechanics, computational modelling of the human musculoskeletal system, motion capture methods, pose estimation, biological signal

from The University of Adelaide. His research interests include sports and clinical biomechanics, with a current focus on orthopaedics and trauma.

Simon Thwaites was born in Perth, WA, Australia, in 1991. He received the B.Eng. degree (Hons.) in mechanical and sports engineering and the M.Eng. degree in mechanical engineering from The University of Adelaide, SA, Australia, in 2013 and 2014, respectively, where he is currently pursuing the Ph.D. degree in Medicine, under the supervision of Assoc. Prof. D. Thewlis. From 2015 to 2017, he was a Research Assistant with the School of Mechanical Engineering, and in 2018, he was a Research Officer with the School of Medicine, both

David Booth received the B.Sc. degree in computer science from Wolverhampton University, in 1984, and the M.Phil. and Ph.D. degrees in applied mathematics from Oxford Brookes University, in 1986 and 1991, respectively. From 1986 to 2008, he had various roles at the Defence Science and Technology Laboratory, Malvern, U.K. and its predecessors. Between 2008 and 2020 he was a Senior Research Scientist with the Defence Science and Technology Group, Edinburgh, latterly as a member of the Biometrics Group. He is currently an independent scientist with wide ranging interests including the application of novel sensors and video analytics.

scientist and founding member of DeepSightX who won the second prize in the global mineral discovery challenge in 2019.

Ehsan Abbasnejad is a Senior Research Fellow at the Australian Institute for Machine Learning (AIML) and a Principal Researcher at Centre for Augmented Reasoning (CAR) at the University of Adelaide. He was awarded his PhD degree in 2015 in Computer Science from the Australian National University (ANU). He has extensive experience in deep learning for a range of applications of machine vision and language in agriculture, mining, defense, etc. His past experience in industry includes Microsoft, Xerox and NEC. He has been a research

William Robertson was born in Adelaide, Australia, in 1981. He received the B.Eng. (Hons.) and Ph.D. degrees in mechanical engineering, in 2002 and 2013, respectively, under the supervision of B. Cazzolato, from The University of Adelaide, Australia, where he is currently a Lecturer with the School of Mechanical Engineering. His research interests include magnetic levitation, vibration control, and biomechanics. He has taught six courses in dynamics and control and sports engineering, and is a Lecturer and a Coordinator for the honours project course.