Sensor fusion for analysis of gait under cognitive load

Document Version
Submitted manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Sensor Fusion for Analysis of Gait under Cognitive Load: Deep Learning Approach

Abdullah S Alharthi  
The University of Manchester, United Kingdom  
abdullah.alharthi@manchester.ac.uk

Syed U Yunas  
The University of Manchester, United Kingdom  
syed.yunas@manchester.ac.uk

Krikor B Ozanyan  
The University of Manchester, United Kingdom  
k.ozanyan@manchester.ac.uk

Abstract—Human mobility requires substantial cognitive resources, thus elevated complexity in the navigated environment instigates gait deterioration due to naturally limited cognitive load capacity. This work uses deep learning methods to study the effects of cognitive load on human gait of healthy subjects. We demonstrate classifications, achieving 86% precision with Convolutional Neural Networks (CNN), of normal gait as well as 15 subjects' gait under two types of cognitive demanding tasks. Floor sensors capturing multiples of up to 4 uninterrupted steps were utilized to harvest the raw gait spatiotemporal signals, based on the ground reaction force (GRF). A Layer-Wise Relevance Propagation (LRP) technique is proposed to interpret the CNN prediction in terms of relevance to standard events in the gait cycle. LRP projects the model predictions back to the input gait spatiotemporal signal, to generate a "heat map" over the original training set, or an unknown sample classified by the model. This allows valuable insight into which parts of the gait spatiotemporal signal have the heaviest influence on the gait classification and consequently, which gait events are mostly affected by cognitive load.

Keywords—Deep Convolutional Neural Networks (CNN), Cognitive Load, Ground Reaction Force (GRF), Layer-Wise Relevance Propagation (LRP).

I. INTRODUCTION

Gait patterns have been intensively studied in recent years to understand the relationship between cognitive load and gait disturbance, with understandable interest on changes with age. For many years, gait of healthy individuals was considered to be separate from the cognitive system, but now it is accepted that such systems are inter-related at the level of the cerebrum. Furthermore, it has been asserted that gait is not an automatic motor sequence independent from cognition, but can be affected by cognitive demanding tasks [1].

Yoge et al. [2], based on the impact of executive function and attention in gait, suggested that the latter is no longer regarded as an automated motor activity that receives minimal cognitive input. Indeed, the limb movements that produce gait with the ability to navigate within often complex surrounding to successfully reach a desired destination, influences gait. The multifaceted neuropsychological influence on humans to appropriately control the mobility and related behaviors, is intensively investigated in recent years for healthcare applications, leading to the hypothesis that the changes in individual gait under dual task (performing a demanding cognitive task while walking), are determined by the capacity decline of cognitive executive function in the normal ageing process, when the overall efficiency of the brain is reduced [1], [2], [3].

Despite the step in research on the influence of cognitive load on gait in the past few decades, the systems providing sensor data have been limited to the following: instruments attached to the body non-invasively [4], [5], force plate where the user must watch where to step [5], or monitoring of gait while walking on a treadmill [6], [7]. These methods may influence the fidelity of recording the natural gait pathology under cognitive demanding tasks, as imposed conditions are quite different from the natural environment in which an individual would walk.

The interaction of the human body with the walking surface is the point of contact with the environment, which cannot be avoided or modified at will. This interaction is typically described in terms of the ground reaction force (GRF). Thus, large area, unobtrusive floor sensor systems are called for, to simulate a real-world gait under cognitive load. This allows investigating the spatiotemporal dynamics of the forces placed on the ground during the gait cycle, by recording the signal of the cyclic behavior in varying spatial regions and time periods. Multi-sensors data analysis and fusion methodology would be required for adequate data analysis aiming to compare natural gait with that influenced by cognitive load. An obvious motivation is to study the consistency of gait patterns which can be associated with fall risk [8], [9] and mental decline, e.g. Parkinson's disease [10], [11].

Cognitive load influence studies have targeted to quantify gait deterioration using established criteria by visual monitoring, or using shallow learning methods for classification. It is difficult to claim that any of these methods, or their combination, would be able to access the maximum variability of gait deterioration and thus provide the best achievable classifications.

Therefore, the goal of this work is twofold: firstly, to categorize gait parameters in healthy adults while performing cognitive demanding tasks, using automatic extraction of optimal gait features by deep CNN; secondly, to interpret the
model performance on unseen spatiotemporal signals by applying the technique of Layer-Wise Relevance Propagation (LRP) [12] to attempt linking key known events in the gait cycle to cognitive deterioration. As a main data acquisition methodology, the subject’s unique gait signatures based on GRF signals were recorded using a set of multiple floor sensors based on plastic optical fiber (POF) technology, in a system specially designed for optimal spatiotemporal sampling [13],[23]. The main sensor data processing methodology is deep CNNs with LRP applied on the classifications to allow interpretation of the behavioral changes due to a demanding task while walking.

II. METHODOLOGY

A. iMAGiMAT System

The gait cycle recordings reflect the recurrent contact of the feet with the surface. The experimental work utilizes a photonic guided-path tomography sensor head, as part of an original IMAGiMAT footstep imaging system [13], for its ability to record unobstrusively temporal samples from each sensor under natural walking environment (a general retail floor carpet). The sensors constitute a carefully designed set to allow collaborative sensor fusion and deliver spatiotemporal sampling adequate for gait events.

The 1m x 2m area system (see figure 1) contains 116 POF sensors, arranged in three plies sandwiched between the carpet top pile and the carpet underlay - a lengthwise ply with 22 POF at 0° and two 47 POF plies arranged diagonally at 60° and 120°. The system electronics is concealed in a closed hard shell located at the rectangular boundaries at carpet surface level. The operational principle of the system is based on capturing the deformation caused by the GRF variations as the POF sensor’s transmitted light intensity decreases because of surface bending. This captures the specifics of foot contact and generates robust data without constraints of speed or positioning anywhere on the active surface.

B. Data Acquisition

Ethical approval was obtained through the Manchester University Research Ethics Committee (MUREC); each participant’s written consent was obtained prior to experiments and research was performed in accordance with the ethics board general guidelines. Healthy participants were invited to walk along the 2 m length direction of the iMAGiMAT sensor head normally, or while performing cognitive demanding tasks. Since the walking routine involved steps before and after contacting the active sensor area, the recorded gait spatiotemporal signals were able to capture around 4 uninterrupted footsteps at each pass.

15 healthy subjects aged 20 to 38 years, 13 male and 2 females, participated in this experiment. Each subject was asked to walk several gait cycles before and after stepping on the iMAGiMAT system for each experiment, so the captured gait data is unaffected by start and stop. With a capture rate of 20 frames/s (each frame comprising the readings of all 116 sensors), experiments yielded 5 s long time sequences each containing 100 frames.

Three manners of walking were defined as normal gait, as well as two different dual tasks, and experiments were recorded for each subject, with 10 gait trials for each manner of walking in a single assessment session, total number of samples is 450. Each manner of walking was identified as class, as follows:

- Class 0, Normal Gait: walking at normal self-selected speed.
- Class 1, Gait with serial 7 subtractions: normal walking speed attempted, while simultaneously performing serial 7 subtractions (count backward from a given random number by sevens).
- Class 2, Gait while texting: normal walking speed attempted, while simultaneously typing text on a mobile device keyboard.

C. Convolutional Neural Network (CNN)

State-of-the-art CNNs have established themselves in a variety of classification tasks, providing new insights into complex spatiotemporal signals. The models learn a high level of abstraction and pattern by applying convolution operations to extract features and perform classification. Commonly, the model consists of convolution layers, pooling layers and normalization layers, with a set of shared filters and weights. With an input \( x_i \), a kernel \( L_i \) and \((*)\) denote the element wise multiplication, the convolution operation \( C \) for layer \( s \) can be expressed as:

\[
C_s = x_i * l_i = \sum_{d=0}^{N-1} x(d) * l(i - d) \tag{1}
\]

The architecture of the CNN model engineered for this study is shown in figure 2. The model consists of 12 stacked layers to categorize all 15 subjects’ gait, normal and dual task. The model is trained and validated using a batch size of 100
samples for each iteration, 200 echoes being optimal for training. The training, validation and testing sizes are set to be 70%, 10% and 20% respectively. To improve the model performance a regularization method is utilized as follows:

i. A batch normalization followed by a dropout of size 0.5, after the last MaxPooling layer was flattened.

ii. A dropout of size 0.2 before the output layer.

Decisively, softmax classifier is placed at the final layer to classify gait spatiotemporal signals. In figure 2, each convolution layer has (kernel × feature maps × filter) labels, and they are known with the number of layers as hyperparameters. These hyperparameters are optimized based on extensive experimenting. For this study, experimentation and algorithm implementation are done using Keras libraries [14]. This tool box provides variety of commands tested until the best model for is reached. This is based on testing the model on unseen data.

D. Layer-Wise Relevance Propagation (LRP)

LRP [12], [17] is an emerging machine learning method for explaining the predictions of nonlinear deep neural networks, and it has shown notable success in image classification [15], [16] as well as subject identification based on their gait [4].

The aim is to quantify the contribution of a single component of an input \(x_i\) (in our case, an iMAgiMAT sensor signal at a specific time frame) to the prediction of \(f_c(x)\) (as gait class \(c\)) made by the CNN classifier \(f\). Gait class prediction is redistributed to each intermediate node via backpropagation until the input layer. The LRP outputs a heat map over the original signal to highlight the signal sections of highest contributions to the model prediction. To understand this methodology, we first note that a CNN network consists of multiple learning parameters as follows [4]:

\[
f_c(x) = \omega^T x_i + b_c = \sum_i \omega_{ic} x_i + b_c
\]  

(2)

Here, \(\omega_{ic}\) is the weight of a neuron \(i\) in a specific layer \(s\) for class \(c\) and \(b_c\) is the bias term for class \(c\); these parameters are learned in the CNN during supervisory training by fusing the POF sensor signals. The output \(f_c(x)\) is evaluated in a forward pass (see figure 3 (1)) and the parameters \((\omega_{ic}, b_c)\) are updated by back-propagating using model error or the loss. For the latter, we base the computations on categorical cross entropy.

The LRP decomposes the CNN output for a given prediction function of gait class \(c\) as \(f_c\) for an input \(x\) and generates a “relevance score” \(R_i\) for all neurons \(j\) that receive relevance from neuron \(k\) for the prediction of \(f_c(x)\), such that

\[
\text{LRP} = \sum_i R_i = j = \sum_j R_{ij} = f_c(x).
\]

The LRP starts at the CNN output layer after removing the Softmax layer. In this process, a gait class \(c\) (e.g. normal gait, subtraction 7 gait, or walking while texting) is selected as input to LRP and the other classes are eliminated.

Assuming we have a single output neuron \(j\) in one of the model layers, this neuron receives a relevance score \(R_j\) from an upper layer neuron \(k\) (see figure 3(2)) or the output of the model (class \(c\). The scores are redistributed between the connected neurons throughout the network layers, based on the contribution of the inputs \(i\) in signal \(x\) using the activation function (computed in the forward pass and updated by back-propagating during training) of neuron \(j\).

The neuron \(k\) will hold a certain relevance score based on the activation function of that neuron, and it passes its influence to consecutive neurons (\(j\) in reverse direction). Finally, the method outputs relevance scores for each sensor

![Fig. 2. CNN network architecture for gait classification, and LRP](image-url)
signal at a specific time frame. These scores represent a heat map, where the strong values at specific time frames highlight the areas that contributed most to the model classifications. The entire set of model relevance scores \( R_i \) can be represented as follows [17]:

\[
R_i^{(s-1)} = \sum_j \frac{x_j^{(s-1)}}{\sum_j x_j^{(s-1)}} R_k^{(s)}
\]

Here \( i \) indexes a neuron for layer \( s \), and \( j \) is computed over all neurons joined to neuron \( i \). Equation 3 it satisfies the relevance conservation property \( \sum_i R_i = R_i \). Figure 3 shows the forward-pass of the CNN model. Figure 4 shows the spatial average \( SA \) of the spatiotemporal gait signal based on the GRF:

\[
SA[t] = \frac{1}{116} \sum_{i=1}^{116} (x_i[t])
\]

to capture gait events due to the stepping behavior. Here \( x_i \) are the readings from individual sensors and \( t \) enumerates the frames in each sample.

A. Gait spatiotemporal Classifications

The proposed model is trained and validated on 360 spatiotemporal signals samples of size 100x116, which is 80% of the 15 subject’s gait trials. The remaining 90 samples of the same size 100x116 are used for testing the model prediction and computing the LRP relevance map.

Three types of gait including cognitive demanding task patterns are learned. This is based on fusing 116 POF sensors in the deep layers of the CNN model, to extract gait patterns automatically. The model learns gait events for different subjects while they perform cognitive demanding tasks. The classifications result showing in figure 5, indicates that the classification accuracy is 85% for normal gait, 90% for walking with 7s subtraction and 79% for walking while texting. The highest confusion was between the two dual tasks. These accuracies were obtained after extensive Optimization of the CNN model shown in figure 2, with the goal to achieve best prerequisite for performing LRP analysis.

B. Gait Spatiotemporal Classification Decomposition

The CNN model in figure 2 was frozen and transferred using transfer learning methods. Deep Taylor decomposition [12] based on the LRP (see II.D) was utilized for this work using the iNNvestigate library [19]. The LRP method decomposes the CNN prediction, to time-resolved input relevance scores \( R_i \), for each spatiotemporal input \( x_i \). This highlights which gait cycle events are relevant towards informing the model’s final classification outcome. Figure 6 displays an example overlay of a 1-D “heat map” for dual task samples classified with 100% true positives.

For more in-depth analysis, we compare in figure 7 the \( SA \) of the raw IMAGiMAT POF signals from (4) with a relevance

![Fig. 4. Gait spatial average of gait spatiotemporal signals (see equation 4), y axis: transmitted light intensity [%]; x axis: frames. Gait events recorded by the sensors in a typical gait cycle[18]: 1- Heel strike, 2- Foot-flattening, 3- Single support, 4- Opposite Heel strike, 5- Opposite Foot-flattening, 6- Double support, 7-Toe-off, 8- Foot swing, 9- Heel strike, 10- Double support, 11- Toe-off, 12- Foot swing, 13- Opposite Heel strike, 14- Single support, 15-Toe-off.](Image)

![Fig. 5. CNN model confusion matrix classification of 15 subjects.](Image)

![Fig. 6. Raw IMAGiMAT POF signals top, corresponding LRP scores signals bottom.](Image)
temporal quasi-signal generated from LRP calculations (1-D heat maps), as follows (further explanation in IV): the LRP generates relevance scores (see figure 6) over the input space of the original spatiotemporal signal; these scores are then used to outline the input variable most contributing to the prediction of a certain class. Figure 7 shows the temporal sequences with identified gait events for three subjects (raw sensor signal and computed LRP quasi-signal row pairs for each subject).

IV. DISCUSSION AND CONCLUSION

In this paper, GRF-derived spatiotemporal signals from healthy control subjects are recorded with the floor sensor iMAGiMAT system, under the conditions of normal walking, as well as walking while performing cognitive demanding dual tasks. Deep learning methods are applied to extract gait features automatically end-to-end from raw data. Gait under cognitive load is classified with 86% weighted precision. The LRP relevance scores point out which parts of the gait spatiotemporal signal were most relevant for classification, as shown in figure 7 for three subjects. To understand this approach, we note that on multiple repetitive occasions each subject will initiate a gait cycle (explained in [18]) by heel strike, strictly followed by other gait events described in figure 4.

Figure 4 details a subject’s recorded gait temporal signal, where it starts by heel strike and ends by toe off when the user steps out of the iMAGiMAT floor sensor. Consistent with this, each gait spatiotemporal sample (classified by the CNN network) is assigned a corresponding spatiotemporal LRP computed quasi-signal as shown in figure 6. The inherent noise in the latter suggests a focus on the gait events on the surface of the carpet, rather than particular computed numbers. Therefore, we use (4) on the LRP output to generate a heat map, which is possible to overlap with the temporal activity on the carpet. The overlap with events numbered from 1 to 9 on figure 7 implies that the effect of cognitive load on gait can be summarized as follows.

i. Normal gait: the signals most influential for the CNN to identify this class are after the heel strike and before the foot-flattening (figure 7, event numbers 1, 4, 7).

ii. Walking while performing 7s subtraction: the CNN classification of this gait is based strongly on the
transition between foot-flattening and opposite toe-off (figure 7, event numbers 2, 5, 8).

iii. Walking while writing a text in smartphone: LRP scores for this event, are based on the transition between foot-flattening and double or single support (figure 7, event numbers 3, 6, 9).

Overall, the LRP analysis indicates that subjects’ normal gait is characterised by an emphatic heel strike (high GRF due to the heel landing); gait while performing 7s subtraction highlights flat-foot (shift of emphasis away from hill strike due to less GRF resulting from weakened heel landing), gait while texting is characterised by weak support (compromised balance while walking).

The results are promising, but apart from assessing the particular classification accuracy numbers, the main weakness in the conclusions comes from the limited number of subjects yet included in the currently reported starting phase of this work. A larger number of subjects, grouped in clusters of 10–15, is expected to yield better results in studying the effect of cognitive load on gait using LRP. Immediate future work will involve a larger number of subjects, better coverage of age variations while maintaining gender balance.

It is worth mentioning that the results reported at this point, and the LRP decomposition of gait classifications in particular, indicate a possible bridge between the output of artificial intelligence systems for processing of high quality gait data and decisions based on visual observations by humans, or quantitative parameters derived from such observations, (c.f. figure 1 in [18]) which are tested and routinely implemented in current practice. In healthcare alone, these approaches may address some of the significant interest in the detection of the onset of Parkinson disease [20] and Alzheimer's [21], freezing of gait [22], and fall risks [8]. Other potential spin-offs are in the areas of biometrics and security.

REFERENCES


