electronics

MDPI

Article

Sensing-HH: A Deep Hybrid Attention Model for Footwear Recognition

Yumin Yao [1,2,3], Ya Wen [4,5,*] and Jianxin Wang [1]

1 School of Computer Science and Engineering, Central South University, Changsha 410083, China; yaoyumin@csu.edu.cn (Y.Y.); jxwang@mail.csu.edu.cn (J.W.)
2 School of Information, Hunan Radio and Television University, Changsha 410004, China
3 Changsha Yumin Information Technology Co. Ltd., Changsha 410221, China
4 Massachusetts General Hospital, Harvard Medical School, Boston, MA 02129, USA
5 Ietheory Institute, Boston, MA 02129, USA
* Correspondence: yawen@ietheory.org

Received: 17 August 2020; Accepted: 10 September 2020; Published: 22 September 2020

Abstract: The human gait pattern is an emerging biometric trait for user identification of smart devices. However, one of the challenges in this biometric domain is the gait pattern change caused by footwear, especially if the users are wearing high heels (HH). Wearing HH puts extra stress and pressure on various parts of the human body and it alters the wearer's common gait pattern, which may cause difficulties in gait recognition. In this paper, we propose the Sensing-HH, a deep hybrid attention model for recognizing the subject's shoes, flat or different types of HH, using smartphone's motion sensors. In this model, two streams of convolutional and bidirectional long short-term memory (LSTM) networks are designed as the backbone, which extract the hierarchical spatial and temporal representations of accelerometer and gyroscope individually. We also introduce a spatio attention mechanism into the stacked convolutional layers to scan the crucial structure of the data. This mechanism enables the hybrid neural networks to capture extra information from the signal and thus it is able to significantly improve the discriminative power of the classifier for the footwear recognition task. To evaluate Sensing-HH, we built a dataset with 35 young females, each of whom walked for 4 min wearing shoes with varied heights of the heels. We conducted extensive experiments and the results demonstrated that the Sensing-HH outperformed the baseline models on leave-one-subject-out cross-validation (LOSO-CV). The Sensing-HH achieved the best F_m score, which was 0.827 when the smartphone was attached to the waist. This outperformed all the baseline methods at least by more than 14%. Meanwhile, the F_1 Score of the Ultra HH was as high as 0.91. The results suggest the proposed model has made the footwear recognition more efficient and automated. We hope the findings from this study paves the way for a more sophisticated application using data from motion sensors, as well as lead to a path to a more robust biometric system based on gait pattern.

Keywords: deep hybrid attention model; footwear recognition; ubiquitous computing

1. Introduction

Recently, with the wearable technology advancing at a fast pace, billions of smart devices have been equipped with built-in motion sensors such as accelerometers and gyroscopes. They can be exploited to log the body motion of users, which can be a very useful tool for the research communities studying motion sensing. More and more researchers have used the motion characteristics of the human body for various tasks, which ranged from activity recognition [1–6], gesture categorization [7], clinical condition monitoring [8], BMI predication [9], to user gait recognition [10–13]. In particular,

identity recognition using the dynamics of the walking pattern seems a promising technique in preventing the use of smart devices and other systems linked with them without the owner's permission. However, the quality of gait-based biometric systems is greatly influenced by the footwear which the subject is wearing.

The previous works [14,15] studied the gait changes related to the different shoes worn by the subjects. Their experiments were carried out with four kinds of shoes with different weights. They found that heavy footwear reduces the discrimination and the sideways motion of the foot has the most discriminating power compared to the up-down or forward-backward directions of the motion. Meanwhile, based on some previous papers on exercise physiology, the height of the heels is also a important parameter related to the human gait. A recent survey [16] summarized a list of the five main open problems for gait recognition including different kinds of shoes. Walking requires ongoing, finely tuned interactions between muscular and tendinous tissues. Wearing HH puts extra stress and pressure on various parts of the human body that would affect the subject's natural gait [17]. In common sense, an increase in the height of HH will cause a decrease in subject's walking speed and the length of stride.

Though footwear alters the gait, there is only a very limited number of studies in footwear recognition. The existing methods normally use the RGB camera [18], the specific motion capture system [19], the ground reaction force sensors [20], or Microsoft Kinect sensor [21], all of which are lab limited. In fact, there is no research on the footwear recognition in the daily life scenario, and none for the HH which about 37% to 69% of American women frequently wear [22]. Additionally, even if only considering the HH, they are categorized into many categories by the height of the heels, as shown in Table 1 and Figure 1.

Table 1. Categorizations of shoes.

Height of Heels	Categories of Shoes
0–2.54 cm (0–1 inch)	Flat
2.54–7.62 cm (1–3 inch)	Mid HH
>7.62 cm (> 3 inch)	Ultra HH

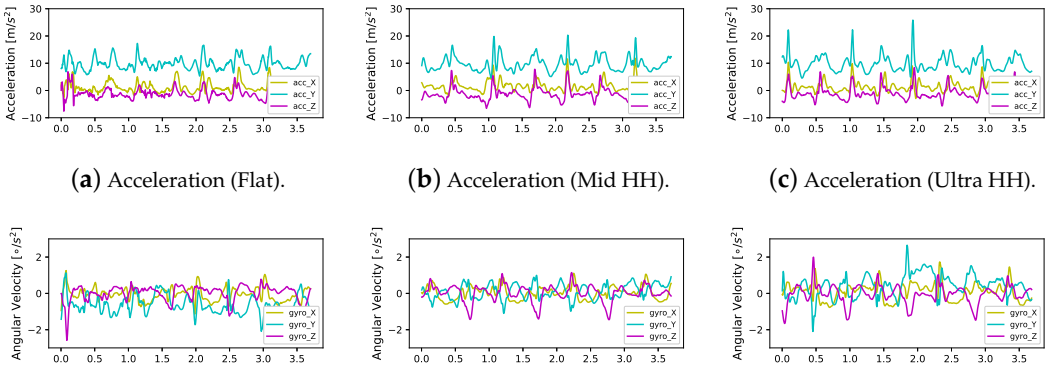
Figure 1. The location of the different smartphones.

Therefore, motion sensor-based footwear recognition using the gait characteristic in daily life is still an open challenge. One of the major challenges is that the daily life walking environment is highly dynamic and it includes a variety of environmental factors that could directly or indirectly introduce variations into the gait patterns. For example, the clothes the individual is wearing, the different walking surfaces, slopes, and obstacles on the road, can all contribute to gait changes besides footwear.

In this section, we evaluate the diffculty of the task by visualizing the raw signals, as shown in Figure 2. A participant of medium build (average weight) was asked to walk back and forth three times on the same surface, wearing different shoes (flat, mid HH and ultra HH), each time with a

smartphone placed on her waist. From the visualized data, we find that the gait of ultra HH (9.8 cm) is significantly different from the previous two scenes. Its acceleration component has a sharper peak, and especially the angular velocity has a lateral rotation lasting for one second. We believe that this is due to the reduced stability and the changes of the center of gravity caused by the ultra HH.

Inspired by the deep neural networks, some very recent works employ them to motion sensor-based recognition, such as Convolution Neural Network (CNN) [3,5,23], which are competent in capturing the local characteristics of multi-channel signals; Recurrent Neural Network (RNN) [24], and its variant, LSTM units [1,25], which are designed to extract the temporal dependencies and incrementally learn information over time.



(**a**) Acceleration (Flat). (**b**) Acceleration (Mid HH). (**c**) Acceleration (Ultra HH).

(**d**) Angular Velocity (Flat). (**e**) Angular Velocity (Mid HH). (**f**) Angular Velocity (Ultra HH).

Figure 2. Some sub-sequences of a participant walking three times wearing different shoes (Flat, Mid high heels (HH) and Ultra HH).

Recently, the combination of CNNs and LSTM in a unified stack framework has already offered state-of-the-art results in sensor-based recognition [7]. In our previous study, we developed a hybrid deep neural network [9] for gait analysis using data captured from built-in motion sensors in smartphones. The hybrid deep neural network overcomes the challenge of environmental factors. In this study, we extended our prior work by incorporating some extensions of attention mechanism to the previous model, and tested its performance by investigating gait changes related to footwear. The extensions we introduced in this study and the major contributions of this paper are summarized in three points:

(1) To the best of our knowledge, we are the first to recognize the subject's footwear by the dynamics of gait changes acquired from smartphone sensors in daily life. We categorize the shoes into 3 classes by the height of the heels (flat, mid HH and ultra HH). We propose Sensing-HH, a novel deep attention model, which can automatically learn a hierarchical feature representation and the infinite temporal contexts from raw signals through the hybrid net structures. It also has the ability to implicitly learn to suppress irrelevant parts in the raw signals and to highlight salient features useful for this specific task by adding the attention mechanism.

(2) We established a dataset with 35 young females wearing 3 kinds of shoes. All of them were asked to walk for 4 min on a flat surface, with 3 smartphones as recording devices, which at the same time were held by their hands, attached to their waists, and placed in their handbags, respectively.

(3) We conducted comprehensive experiments on this dataset to evaluate the proposed Sensing-HH model. The results showed that our model achieved competitive performance with a mean F1-score (F_m) of 0.827 when the smartphone was attached to the waist, from different classes, through cross verification. Meanwhile, the F_1 Score of the Ultra HH was as high as 0.91.

The remaining part of this paper is structured as follows:

In Section 2, we give a brief overview of the state of some related literature. In Section 3, we present how the dataset was established. In Section 4, we illustrate the Sensing-HH, a deep attention model. Experimental results with the baseline methods are presented in Section 5. Section 6 gives the conclusions.

2. Related Work

In this section we summarize research that are most relevant to our proposed approach, grouping them in three domains: footwear related gait analysis using motion sensors, previous deep learning approaches for motion sensor-based recognition and attention mechanism for sensor data processing.

2.1. General Footwear Related Gait Analysis Using Motion Sensors

Wearable motion sensors makes gait analysis [16,26] much easier. The research about general gait analysis using motion sensors are focused on model-based methodology [27,28], which needs to first model gait based on a comprehensive understanding of the gait mechanism, and then convert the sensor signal into some gait-related physiological parameters [29–31], such as gait rhythm, step length, symmetry, inner foot distance, ankle shape, detection of gait phases [32], or kinematic parameters (joint angle measurement) [33]. The recent review [34] has proved the wearable sensors to be very useful in monitoring and analyzing the stability of subjects.

2.2. Previous Deep Learning Approaches for Motion Sensors-Based Recognition

Over the past few years, deep neural networks emerged as a family of learning models for automating feature design, and have achieved tremendous successes in many application domains [35–40]. Particularly, Yosinski et al. [41] demonstrated that features learned were not specific to a particular task and could be useful for multiple related tasks. Some studies employed deep neural networks for motion sensor-based recognition tasks. It is common to use Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), and recently some researchers have paid attention to the hybrid network which consists of CNN and RNN. Gadaleta et al. [12] presented IDNet, a user authentication framework from smartphone-acquired motion signals. The stacked convolutional layers were used as a series of feature extractors, and then One-Class SVM (OSVM) was used as a classifier for gait recognition. The experiments exploited an in-house dataset with data collected from 50 subjects during six months. Data are acquired using different smartphone models positioned in the right front pocket of trousers. Subjects were asked to walk at their normal pace in different walking sessions for about 5 min. The accelerometer, gyroscope were both used in the recognition process for recording. Zou et al. [13] proposed a CNN-RNN structure for robust gait feature representation, with which features of the space and time domains were successively abstracted by the hybrid network. Two datasets were collected for identification and verification. In a previous work, we also proposed a hybrid deep neural network [9] to predict the BMI of smartphone users, which was also based on the characteristics of body movement captured by the smartphone's built-in motion sensors.

2.3. Attention Mechanism for Sensor Data Processing

The attention mechanism is popular in deep learning areas [42]. It has been successfully applied to image recognition [43,44], natural language processing [45,46] and speech recognition [47], which is originally a concept in biology and psychology that illustrates how we restrict our attention to something crucial for better cognitive results. Recently, some researchers have explored the potential of using attention models for processing sensor data, such as Electroencephalography (EEG) and wearable sensor data. Zhang et al. [48] presented a Convolutional Attention Model (CAM) for EEG-based human movement intention recognition in the subject-independent scenario. In the study, the integrated attention mechanism was utilized to focus on the most discriminative information of EEG signals during the period of movement imagination while omitting other less relative parts.

Zhang et al. [49] introduced a selective attention mechanism into the reinforcement learning scheme to focus on the crucial dimensions of the multimodal wearable sensor data. This mechanism helped to capture extra information from the signal and thus it was able to significantly improve the discriminative power of the classifier. Zeng et al. [50] proposed two attention models for human activity recognition: temporal attention and sensor attention. These two mechanisms adaptively focused on important signals and sensor modalities. Wang et al. [51] presented an attention-based convolutional neural network for human recognition from weakly labeled data.

Our proposed attention model is focused on a long sequence of sensor data, and it not only improves the performance of the model but also has better interpretability.

3. Dataset

To our best knowledge, there is no existing dataset that specifically studied the motion sensor-based gait recognition of HH wearing in a daily environment. In this section, we describe our strategy for motion sensor data collection to build the dataset.

3.1. Participants Selection

We recruited female participants who wear HH for at least 5 days a week, for an average of 12 h a day (including walking, sitting and standing). In order to avoid other factors such as age, height, and weight to impact the results, we selected 35 subjects with the age range from 19 to 27, and with similar builds. Participant details are shown below: age: 23 ± 4 years; height: 164.3 ± 12.4 cm; mass: 51.8 ± 7.6 kg. Each of the participants was informed before the experiment of its aim and the measuring method. All of them signed a consent to participate in the study. Prior to the gait measurement, we conducted a short survey asking questions about the preferred types of footwear and how frequently they wear HH. Two-thirds of the participants answered that they preferred flat shoes in their day to day life. One-third of them preferred high heeled shoes, even with the heels more than 8 cm in height. All of them wore 3 kinds of shoes (flat, mid HH and ultra HH) for this study.

3.2. Data Collection

All of the motion sensor data were recorded by a log application from 3 different android smartphones (Samsung Galaxy S10, Samsung Galaxy Note8, and Smartisan Pro2). Table 2 summarizes sensor specifications for the devices.

Table 2. Sensors specifications with the max. sampling rate.

Smartphone	Accelerometer	Gyroscope	Magnetometer
Samsung S10	STMicro LSM6DSO (416 Hz)	STMicro LSM6DSO (416 Hz)	AK09918 (50 Hz)
Samsung Note8	STMicro LSM6DSL (400 Hz)	STMicro LSM6DSL (400 Hz)	AK09916C (50 Hz)
Smartisan Pro2	Bosch BMI160 (200 Hz)	Bosch BMI160 (200 Hz)	AK09918 (50 Hz)

The tri-accelerometer and the tri-gyroscope are motion sensors equipped by the smartphones we used. The tri-accelerometer is based on the basic principle of acceleration and it is used to measure the acceleration (including gravity) in the X, Y and Z directions of the smartphones. The tri-gyroscope captures the angular velocity of a smartphone during its rotation in space. Both of them reflect the gait characteristic of smartphone users.

All of the participants were asked to walk for 4 min on a flat ground, as shown in Figure 3, the recording devices, the 3 smartphones that was mentioned before were held on their hands, attached to their waists, and placed in their handbags, respectively, as shown in Figure 4.

Figure 3. Photos of participants wearing Ultra HH.

Figure 4. The locations of the different smartphones.

4. Methodology

In this section, we give an overview of the development of Sensing-HH. First, we define the notations used in this study. Second, we introduce the proposed Sensing-HH model in details.

4.1. Notation and Definitions

To avoid ambiguity, we are clarifying the following terms used in this paper: Sequence, Sub-Sequence and Instance:

4.1.1. Sequence

The sequence S is all recordings of one subject, it is an ordered list of multi-dimensional time series that are typically recorded in temporal order at fixed intervals. Given the dataset with total N subjects, the m-th subject, $m \in [1, N]$, the sequence is S_m, and T_m is the total number of intervals.

$$S_m = \{d_m^1, \ldots, d_m^i, \ldots, d_m^{T_m}\} \tag{1}$$

d_m^i denotes the m-th subject's sensor recording (tri-axis accelerometer and tri-axis gyroscope) at the i-th sampling point and $i \in [1, T_m]$, as follows:

$$d_m^i = \begin{bmatrix} a_{m,x}^i \\ a_{m,y}^i \\ a_{m,z}^i \\ g_{m,x}^i \\ g_{m,y}^i \\ g_{m,z}^i \end{bmatrix} \tag{2}$$

In this paper, the sequence S_m will be segmented into a series of sub-sequences by a sliding windows strategy.

4.1.2. Sub-Sequence

The de-facto standard workflow for processing sensor data in ubiquitous computing treats individual sub-sequences x_m^k as statistically independent.

x_m^k, $k \in [1, L]$, is the k-th sub-sequence of the sequence S_m:

$$S_m = \{x_m^1, \ldots, x_m^k, \ldots, x_m^L\} \tag{3}$$

$L = \lfloor \frac{T_m - w}{\theta} \rfloor$, w is the length of each sub-sequence, and θ is the step between the start intervals of two consecutive sub-sequences. Concretely, x_m^k has the sampling points from $d_m^{(k-1)\times\theta}$ to $d_m^{(k-1)\times\theta + w}$.

$$x_m^k = \{d_m^{(k-1)\times\theta + 1}, \ldots, d_m^{(k-1)\times\theta + w + 1}\} \tag{4}$$

4.1.3. Instance

In practice, the instance i_m^k refers to the data fed into the recognition model, which is the suitable transforming format of sub-sequence x_m^k by data preprocessing function $\mathcal{H}(*)$.

$$i_m^k = \mathcal{H}(x_m^k) \tag{5}$$

In this paper, the task is to learn a function $f : \mathcal{I} \to \mathcal{Y}$ from a given data set. Where $\mathcal{I}$ denotes the instance space, $\mathcal{I} = \{i_m^1, \ldots, i_m^k, \ldots, i_m^L\}$, $k \in [1, L]$, and $\mathcal{Y}$ is the set of class labels, $\mathcal{Y} = \{'Flat', 'Mid - HH', 'UltraHH'\}$.

Given the unified representation f, we simultaneously optimize the network by minimizing a loss function L, which makes it possible to shorten the distance between the predicted label and ground truth.

4.2. Sensing-HH: A Deep Attention Model

This subsection introduces our proposed deep attention network, which consists of two streams, and takes acceleration and angular velocity as inputs respectively. Each stream is composed of four different Modules: a signal preprocessing module, a deep hybrid connection network module, an attention network module, and a fusion module. As illustrated in Figure 5, and the details are presented as follows:

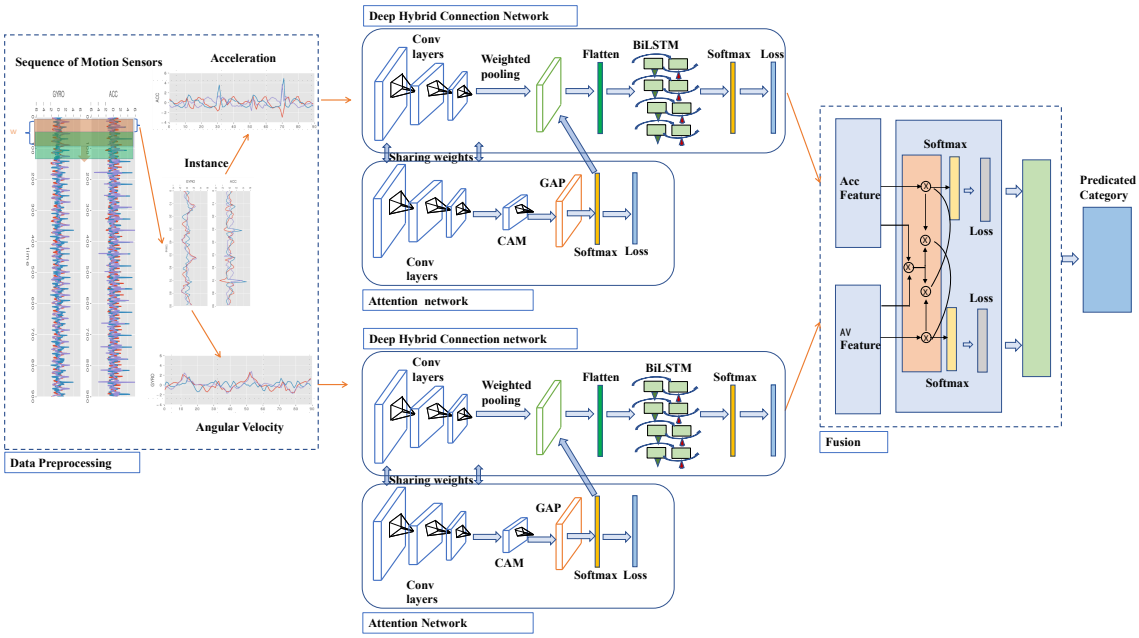


Figure 5. Overall architecture of the Sensing-HH model.

4.2.1. Data Preprocessing Module

In practice, the data preprocessing module includes three main steps:

Step 1: Resampling and Interpolating

Unlike some specific sensors which are used under constrained experimental conditions, the sampling frequencies in most smartphones' built-in sensors are time-varying because their processing unit and operating system were designed for multitasking [11]. Additionally, different sensors have different sampling rates to guarantee the data from all types of the motion sensors can be processed simultaneously, resampling and interpolating steps are required to transform the sequence of raw signals into equally spaced time series. In this paper, the motion sensor data time series are interpolated using cubic spline method [52] and resampled at f = 200 Hz.

Step 2: Gravity Filtering

Raw accelerometer data include gravity components, which makes it difficult to use motion sensors to reflect the change of celerity and position of a smartphone at the time. In this paper, we applied a novel gravity filtering method based on the combination of EMD (Empirical Mode Decomposition) and the wavelet threshold, which is proposed by Lu et al. [53].

Step 3: Normalization

After filling up the missing values by resampling and interpolating, we normalize the training data by setting data mean to 0 and standard deviation to 1, and as usual we use the training data mean and standard deviation to normalize the test data.

4.2.2. Deep Hybrid Connection Network Module

As a result, learning the inter-modality correlations along with the intra-modality information is one of the major challenges in HH recognition from multi-modalities of signals. The current researches of sensor-based recognition are usually accomplished with multiple different sensors such as accelerometer and gyroscope. Generally, using the diverse sensing modalities can obtain better results than using only one particular sensor. Our proposed deep hybrid connection neural networks

consist of two-stream CNN-BiLSTM (Bidirectional LSTM) networks with the stacked convolution layers and bi-directional Long Short-Term Memory that encode features from multiple perspectives.

There are two main components in the CNN-BiLSTM, the first one is the stacked 2-Dimensional CNNs, which is applied to extract spatial features from processed sensory data such as acceleration and angular velocity. The second one is the BiLSTM which is responsible for learning the bidirectional long-term dependencies of salient features extracted by CNNs.

In practice, the stacked CNNs are competent in capturing the local connections of sensory data in spatial scale. In order to learn a rich representation of the input, the convolutional layers produce a set of multiple feature maps. Although the cells in adjacent convolutional layers are locally connected, various significant patterns of input signals at different levels can be obtained by stacking several convolutional layers to form a hierarchical structure of gradually more abstract features. The 2-Dimensional convolution layer l with its operation of calculating a feature map $c_{i,j}^{l,M}$ as:

$$c_{i,j}^{l,m} = \varphi\left(\sum_{m'=0}^{M'-1}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} w_{x,y,m'}^{l-1,k} z_{i+x-1,j+y-1}^{l-1,M'} + b^{l-1,m}\right) \tag{6}$$

where X and Y are the size of convolution kernel running over space and time, respectively, M' is the number of feature maps in the convolutional layer $(l-1)$, $w^{l-1,m'} \in \mathbb{R}^X \times Y \times M'$ is a local filter weight tensor, and $b^{l-1,k} \in \mathbb{R}$ is a bias, and $\varphi(*)$ is the Rectified Linear Units (ReLU) nonlinear function.

One shortcoming of conventional LSTM is that they are only able to make use of the previous context. Following Bi-LSTM, the same input data are fed into a forward LSTM and a backward LSTM. Then two hidden states are concatenated to compute the final output of Bi-LSTM y_t as:

$$\overrightarrow{h}_t = LSTM\left(x_t, \overrightarrow{h}_{t-1}\right) \tag{7}$$

$$h_t = LSTM\left(x_t, h_{t-1}\right) \tag{8}$$

$$y_t = W_{\vec{h}} h_t + W_h h_t + b \tag{9}$$

where $\overrightarrow{h}_t$ is the forward LSTM hidden state and h_t is the backward LSTM hidden state simultaneously at each time step t, $LSTM(*)$ denotes the LSTM operation, $W_{\vec{h}}$ and W_h represent the weights of the forward LSTM and the backward LSTM, respectively, and b is the bias at the output layer.

4.2.3. Attention Network Module

As shown in Figure 5, the attention network is constructed based on the deep hybrid connection network we mentioned before. We generate the class activation maps [54] using global average pooling (GAP) in the CNNs parts, where GAP outputs the spatial average of the feature map of each unit at the last convolutional layer. A weighted sum of these values is used to generate the final output. Similarly, we compute a weighted sum of the feature maps of the last convolutional layer to obtain our class activation maps. We describe this in details below for the case of classification using softmax.

The weights of the softmax layer are propagated back to the convolution layers for decomposing the multi-dimensional time series into salient and non-salient regions. The so-called salient regions are considered to contain information on discriminative gait patterns of wearing high heels, which provide indications and important information associated to pre-defined shoe categories, and the non-salient regions that are less relevant to the footwear categories.

For a given instance of signals, we denoted $f_k(c,t)$ represent the activation of unit k in the last convolutional layer at spatial location (c,t) , where c means the channel of signals and t means the timestamps of signals. Then for a certain category m, we denote the corresponding weight of unit k

and the corresponding input of softmax layer as w_k^m, and the the result of performing global average pooling as F^k can be obtained

$$F^k = \sum_{c,t} f_k(c,t) \tag{10}$$

Thus, for a given class m, the input to the softmax, S_m, can indicate the overall importance of convolutional activations for category m, we obtain that

$$S_m = \sum_k w_k^m F_k = \sum_{c,t} \sum_k w_k^m f_k(c,t) \tag{11}$$

Also we can define Att_m the class activation map as class m, and it can directly indicate the importance of the activation at spatial location (c, t) for category m

$$Att_m(c,t) = \sum_k w_k^m f_k(c,t) \tag{12}$$

$$S_m = \sum_{c,t} Att_m(c,t) \tag{13}$$

Finally, after all these processes, we have a set of compatibility score for the output of class m by a softmax function:

$$P_m = \frac{\exp(S_m)}{\sum_m \exp(S_m)} \tag{14}$$

This way, we transfer the spatial attention into deep hybrid connection network to emphasize the salient regions with discriminative information. This attention model is also able to revisit the previous information and focus on more important parts to learn a better representation.

4.2.4. Fusion Module

In the previous work [9], we used the fully connected (FC) layer on top of two-stream CNN-LSTM to produce probability scores on target labels. However, in this paper, to overcome the "one-stream-dominating-the-network" problem, the designed fusion module is combined with the attention weighted learning strategy. On one hand, directly concatenating the convolutional features and feeding it into FC layers may result in over-parameterization, which makes training difficult, especially for a high heels gait dataset on a limited scale. On the other hand, the low accuracy of the previous model is not only due to the over-fitting problem but also because only one type of sensor dominates the network while the other source only has a small impact on the final prediction. In this paper, we modify the attention mechanism to take two sources as input and have the compute attention weight from each source to produce a prediction for the current input by the softmax layer. This assumption is also confirmed by the following stream selection approach. We took two-stream CNN-BiLSTM as input and compute weights for each stream, as follows:

$$s = W_1' x_1 + W_2' x_2 \tag{15}$$

$$e_i = v^T \tanh\left(s + W_i' x_i\right) \tag{16}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{k=1}^{2} \exp(e_k)} \tag{17}$$

$$o_i = \sum_{k=1}^{2} \alpha_i x_i \tag{18}$$

where W_1', W_2' are the weighted parameters of the different streams, and x_1, x_2 are the learned features from accelerometer and gyroscope, respectively. The attention weights are normalized by softmax to create the attention map α_i for each type of sensor.

5. Experiments

In this section, to evaluate the performance of the proposed Sensing-HH model for real-world application scenarios, we carefully conducted an experimental evaluation on a real-world dataset collected by ourselves and compared the results with several baselines methods. Additionally, we tested if there were significant signal differences between using footwear as measuring standard verses not using it.

5.1. Experimental Settings

5.1.1. Baselines

To illustrate the difficulty of the task we also compared the approach proposed in this work with standard classification methods typically used for automated assessment systems in other sensor-based recognition [9,55–58].

RF [55]. The random forest (RF) is an ensemble classifier which, besides classifying data, can be used for measuring attribute importance. RF builds many classification trees, where each tree votes for a class and the forest chooses the classification having the most votes from the trees.

SVM [56]. The recognition process starts with the acquisition of the sensor signals, which were subsequently pre-processed by applying noise filters and then sampled in fixed-width sliding windows. From each window, a vector of 17 features is obtained by calculating variables from the signals in the time and frequency domain. Finally, these patterns are used as input of the trained SVM Classifier for the recognition.

CNN [57]. The stacked 2-Dimensional CNNs were designed to introduce a degree of locality in the patterns matched in the input data and to enable translational invariance with respect to the precise location (i.e., time of occurrence) of each pattern within a frame of movement data.

BiLSTM [58]. The model was based on a bidirectional Long Short-Term Memory Recurrent Neural Network (BLSTM-RNN), which is designed to take contextual information into account. The network can process data gathered from different positions, which results in a system that is invariant to transformations and distortions of the input patterns.

CNN-LSTM [9]. The CNN was designed to capture the spatial relationship, and the LSTM can make use of the temporal relationship. Combining CNN and LSTM enhances the ability to recognize the varied time span and signal distributions.

5.1.2. Setup

The handcrafted feature-based methods use WEKA toolkit [59] and the settings from previous papers [55,56]. Sensing-HH and other deep learning benchmark models [9,57,58] are performed on Keras 2.3.0 and Tensorflow 2.0. For such deep learning models, tuning hyper-parameters is a time-consuming and challenging task due to the fact that numerous parameters need to be configured. In this paper, we applied the functional ANOVA framework proposed by Hoos et al. [60] to estimate the impact of each hyperparameter on the performance observed across all experiments. Six common hyper-parameters, namely the optimizer, learning rate, number of epochs, batch size, dropout rate, and regularizer, are optimized, see Table 3.

Table 3. Setting of hyper-parameters.

Parameter	DL Models Search Space	Optimal Values			
		Sensing-HH	**CNN** [57]	**BiLSTM** [58]	**CNN-LSTM** [9]
Optimizer	**(Rmsprop, Adam, Sgd)**	**Adam**	**Sgd**	**Rmsprop**	**Adam**
Learning Rate	(0.001, 0.01)	0.001	0.001	0.001	0.001
Epochs	(20, 50, 100)	50	50	20	100
Batch Size	(20, 30, 40, 50)	10	10	30	10
Dropout Rate	(0, 0.3, 0.5)	0.5	0.3	0.5	0.5
Regularizer	(L1, L2)	L2	L2	L2	L2

5.1.3. Cross-Validation Strategies

In order to obtain an unbiased evaluation of the classification performance, a leave one subject out cross-validation (LOSO-CV) is adopted. Suppose a dataset with N subjects. For each experiment, we used $N-1$ subjects' sensor data for training and the rest of the subjects' sensor data for testing. At first, in LOSO-CV, the subjects $\{S_n\}_{n=1}^{N}$ are partitioned into N groups. The samples are then partitioned by the groups into N sub-samples $\{D_n\}_{n=1}^{N}$ of the N sub-samples. A single sub-sample D_{test} is retained for testing the model, and the remaining $N-1$ sub-samples D_{train} are used as the training data. Then the cross-validation process is repeated N times, with each of the N sub-samples used exactly once as the validation data. Compared with k-fold cross-validation, the LOSO-CV not only ensures that the testing procedure covers all the participants but also makes it closer to the real-world application.

5.1.4. Evaluation Criteria

Since high precision and high recall are both desired in this application, and the datasets utilized in this work are possibly biased as it is limited by the selection of the subjects. We used the mean $F1$ score (F_m) to estimate the overall performance of different models, which corresponds to the harmonic mean of precision and recall:

$$F_m = \frac{1}{C}\sum_{i=1}^{C}\frac{2\cdot precision_i \cdot recall_i}{precision_i + recall_i} \tag{19}$$

Here, $i = 1, \ldots, C$ is the set of classes considered.

$$precision_i = \frac{TP_i}{TP_i + FP_i}, \quad recall_i = \frac{TP_i}{TP_i + FN_i} \tag{20}$$

TP_i, FP_i represents the number of true and false positive, respectively and FN_i is the number of false negatives.

5.2. Experimental Results and Analysis

Extensive experiments were conducted on footwear recognition tasks on the real-world dataset collected by ourselves, as mentioned in Section 3. We first compare our method with different state-of-the-art works under different locations of devices, held in their hands, attached to their waists, and placed in their handbags, respectively. Then, to demonstrate how well the Sensing-HH works in real-world applications. An additional experiment was performed on a new fusion scene.

5.2.1. Comparsion with Baselines

In this subsection, we extensively compare our model with a set of baseline methods under different scenes for footwear recognition.

Table 4 presents the comparison between the proposed Sensing-HH and the state-of-the-art methods as well as baselines, in three groups of sliding windows parameters settings for example to quantitatively show the different performance of the models, and the best performance is emphasized in bold. In general, the deep neutral network-based models [9,57,58] indeed improve considerably due to the captured complex features from raw signals. On the other hand, the handcrafted feature-based method only has satisfactory results in the waist scene. Sensing-HH achieved the best F_m score which was 0.896, when the smartphone was attached to the waist. Meanwhile, we found that the suitable size of sliding windows for this recognition task was 2 s. Clearly, in this scene, the performance improvements of Sensing-HH over the RF [55], SVM [56], CNN [57], BiLSTM [58], CNN-LSTM [9] models are 23.1%, 17.8%, 14.2%, 16.5% and 15.4%, respectively.

Table 4. Performance comparison measured by F_m, where W_s and 'O' are short for 'Size of Sub-sequence' and 'Overlap of the adjacent sub-sequence or cycle', where 'H', 'W' and 'B' are short for 'Hand', 'Waist' and 'Bag'. Sensing-HH achieves the best results on most parameters, significantly outperforming the state-of-the-art.

Model	W_s:1s, O : 50%			W_s:2s, O : 50%			W_s:5s, O : 50%		
	H	**W**	**B**	**H**	**W**	**B**	**H**	**W**	**B**
RF [55]	0.601	0.649	0.521	0.613	0.671	0.557	0.582	0.648	0.428
SVM [56]	0.637	0.688	0.582	0.672	0.701	0.564	0.601	0.652	0.536
CNN [57]	0.682	0.715	0.604	0.707	0.723	0.641	0.653	0.694	0.572
BiLSTM [58]	0.676	0.693	0.609	0.662	0.709	0.583	0.699	0.714	0.565
CNN-LSTM [9]	0.683	0.711	0.614	0.703	0.716	0.659	0.704	0.723	0.534
Sensing-HH	0.716	0.759	0.627	0.743	0.826	0.636	0.721	0.786	0.617

Overall, the Sensing-HH has robust performance in most of the scenes, regardless of the device locations. The reason could be that it has attention-based two-stream deep hybrid networks. We will discuss this further in the following subsection.

5.2.2. Ablation Study

In this subsection, to demonstrate the efficiency of our framework design, we performed a careful ablation study to examine the contributions of the proposed components to the model's classification performance. Specifically, we removed each component one at a time in our Sensing-HH framework. First, we named the different versions of Sensing-HH with different components removed as follows:

(1) HHw/oATT: The Sensing-HH model without the attention component.
(2) HHw/oLSTM: The Sensing-HH model without the BiLSTM component.

For different variants, we tuned the hidden dimension of models, so that they had similar numbers of model parameters to the completed Sensing-HH, to remove the performance gain induced by model complexity.

The experiment measures use W_s: 2 s, O: 50% settings. The results are shown in Figure 6, with comparison to other deep learning models. Some observations from these results were worth noting:

(1) The best recognition performance was obtained with the smartphone attached to the waist. The Sensing-HH significantly outperformed other deep models in this scene. But the differences amongst all the deep models in other scenes, i.e., held by the hand or in the bag, were not significant.
(2) Removing the attention component (in HHw/oATT) from the Sensing-HH caused the most significant performance drop in the waist scene, which dropped nearly 14.6%. This suggests the importance of the attention component in this mode.

(3) Removing the BiLSTM component (in HHw/oLSTM) from the Sensing-HH caused a performance drop of nearly 4–6% in most of the scenes.

The conclusion is that all of the components together lead to the robust performance of Sensing-HH in all of the scenes.

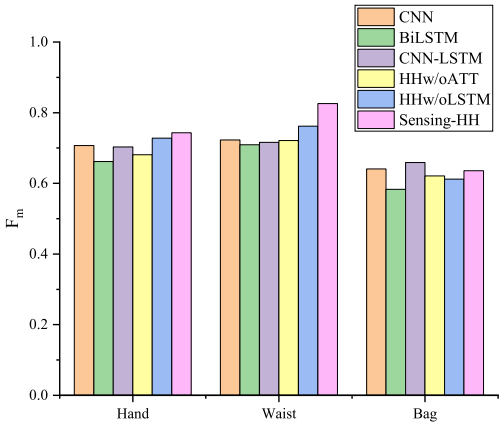


Figure 6. F_m of six models on different class in waist scene.

5.2.3. Failure Cases

To analyze failure cases of our proposed Sensing-HH, we visualized the confusion matrix of the result of misclassification. The details of the instances as shown in Table 5.

Table 5. The instances of different shoes categories (sliding windows of 200 samplings, 2 s).

Categories by Shoes	Number of Instances		
	Hand	Waist	Handbag
Flat	9937	9451	9703
Mid HH	4391	4479	4532
Ultra HH	6591	6902	6811

From the confusion matrix in Figure 7, we found that the recognition accuracy of the Flat and Mid HH classes in the cross-view benchmark were relatively lower than the Ultra HH class, which had Precision 0.92 and Recall 0.89.

Furthermore, we paid attention to the specific failure cases of the Mid HH and Ultra HH classes, as shown in Table 6. We found that gait pattern changes related to the different shoes seemed to be impacted by the subject's body height and weight.

Table 6. Some subjects of failure cases.

No.	Heels (cm)	Height (cm)	Weight (kg)	True Label	Predicted Label
2	7.0	152	46.3	Mid HH	Ultra HH
22	7.1	154	43.8	Mid HH	Ultra HH
12	7.7	170	58.4	Ultra HH	Flat
31	8.1	176	51.6	Ultra HH	Mid HH

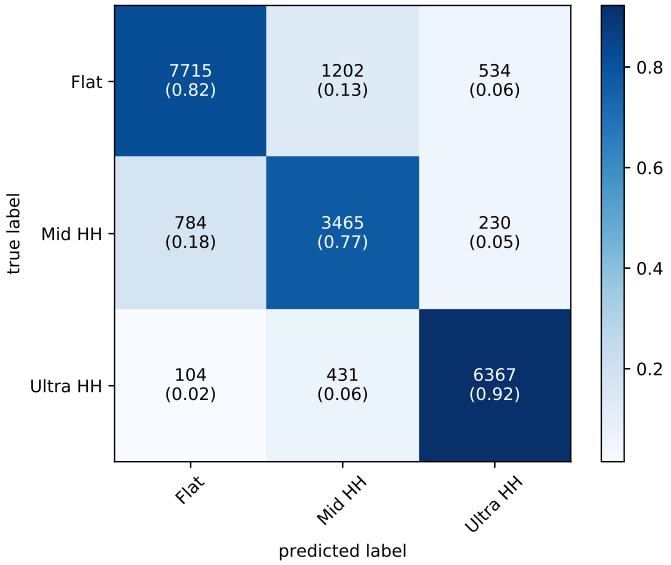


Figure 7. Confusion matrix for the proposed Sensing-HH in waist scene.

6. Conclusions

In summary, we developed Sensing-HH for footwear recognition based on daily life gait data captured by built-in motion sensors from smartphones. To our best knowledge, we are the first to recognize the subject's footwear by the dynamics of gait changes acquired from smartphone sensors in daily life. We categorize the shoes into 3 classes by the height of the heels (flat, mid HH and ultra HH). Sensing-HH is a novel deep attention model which can automatically learn a hierarchical feature representation and the infinite temporal contexts from raw signals through the hybrid net structures. It also has the ability to implicitly learn to suppress irrelevant parts in the raw signals and to highlight useful salient features for this specific task by adding the attention mechanism. We used a daily life gait dataset to evaluate the performance of Sensing-HH and other baseline models. Comparing to three existing deep neural networks and two shallow models, Sensing-HH performed significantly better in most scenarios.

The results show that the proposed model is able to make footwear recognition more efficient and automated. It also can be applied to a large population as it only requires data from smartphones and it can accurately recognize footwear using daily life gait data with no restriction to the location of the measuring devices. Sencing-HH has the potential to extend use of the motion sensor data. For example, to help build a robust biometric system that includes gait pattern analysis. Future studies will focus on how to accurately recognize footwear in a dataset having a wider range of varied heights and weights of the subjects, so that the model would be able to work under an even closer-to-reality scenario.

Author Contributions: Conceptualization, Y.Y., J.W. and Y.W.; software, Y.Y.; validation Y.Y. and Y.W.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.W. and J.W.; project administration, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Guan, Y.; Plötz, T. Ensembles of deep lstm learners for activity recognition using wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–28. [CrossRef]
2. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [CrossRef]

3. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional neural networks for human activity recognition using mobile sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, Austin, TX, USA, 6–7 November 2014; pp. 197–205.
4. Jiang, W.; Yin, Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 13 October 2015; ACM: New York, NY, USA, 2015; pp. 1307–1310.
5. Ha, S.; Choi, S. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 381–388.
6. Yang, J.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
7. Zhu, G.; Zhang, L.; Shen, P.; Song, J. Multimodal gesture recognition using 3-D convolution and convolutional LSTM. *IEEE Access* **2017**, *5*, 4517–4524. [CrossRef]
8. Um, T.T.; Pfister, F.M.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 216–220.
9. Yao, Y.; Song, L.; Ye, J. Motion-To-BMI: Using Motion Sensors to Predict the Body Mass Index of Smartphone Users. *Sensors* **2020**, *20*, 1134. [CrossRef] [PubMed]
10. Nickel, C.; Brandt, H.; Busch, C. Classification of acceleration data for biometric gait recognition on mobile devices. In Proceedings of the BIOSIG 2011 Biometrics Special Interest Group, Darmstadt, Germany, 8–9 September 2011.
11. Zhao, Y.; Zhou, S. Wearable device-based gait recognition using angle embedded gait dynamic images and a convolutional neural network. *Sensors* **2017**, *17*, 478. [CrossRef] [PubMed]
12. Gadaleta, M.; Rossi, M. Idnet: Smartphone-based gait recognition with convolutional neural networks. *Patt. Recognit.* **2018**, *74*, 25–37. [CrossRef]
13. Zou, Q.; Wang, Y.; Wang, Q.; Zhao, Y.; Li, Q. Deep Learning-Based Gait Recognition Using Smartphones in the Wild. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3197–3212. [CrossRef]
14. Gafurov, D.; Snekkenes, E. Towards understanding the uniqueness of gait biometric. In Proceedings of the 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–8.
15. Gafurov, D.; Snekkenes, E.; Bours, P. Improved gait recognition performance using cycle matching. In Proceedings of the 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, Perth, Australia, 20–13 April 2010; pp. 836–841.
16. Marsico, M.D.; Mecca, A. A survey on gait recognition via wearable sensors. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–39. [CrossRef]
17. Cronin, N.J.; Barrett, R.S.; Carty, C.P. Long-term use of high-heeled shoes alters the neuromechanics of human walking. *J. Appl. Physiol.* **2012**, *112*, 1054–1058. [CrossRef]
18. Sarkar, S.; Phillips, P.J.; Liu, Z.; Vega, I.R.; Grother, P.; Bowyer, K.W. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 162–177. [CrossRef]
19. Kim, M.; Kim, M.; Park, S.; Kwon, J.; Park, J. Feasibility study of gait recognition using points in three-dimensional space. *Int. J. Fuzzy Log. Intell. Syst.* **2013**, *13*, 124–132. [CrossRef]
20. Marcin, D. Human gait recognition based on ground reaction forces in case of sport shoes and high heels. In Proceedings of the 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, 3–5 July 2017; pp. 247–252.
21. Derlatka, M.; Bogdan, M. Recognition of a Person Wearing Sport Shoes or High Heels through Gait Using Two Types of Sensors. *Sensors* **2018**, *18*, 1639. [CrossRef] [PubMed]
22. Frey, C.; Thompson, F.; Smith, J.; Sanders, M.; Horstman, H. American Orthopaedic Foot and Ankle Society women's shoe survey. *Foot Ankle* **1993**, *14*, 78–81. [CrossRef]

23. Bevilacqua, A.; MacDonald, K.; Rangarej, A.; Widjaya, V.; Caulfield, B.; Kechadi, T. Human Activity Recognition with Convolutional Neural Networks. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Dublin, Ireland, 10–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 541–552.
24. Sabir, A.T.; Maghdid, H.S.; Asaad, S.M.; Ahmed, M.H.; Asaad, A.T. Gait-based Gender Classification Using Smartphone Accelerometer Sensor. In Proceedings of the 2019 5th International Conference on Frontiers of Signal Processing (ICFSP), Marseille, France, 18–20 September 2019; pp. 12–20.
25. Steven Eyobu, O.; Han, D. Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network. *Sensors* **2018**, *18*, 2892. [CrossRef] [PubMed]
26. Sprager, S.; Juric, M.B. Inertial sensor-based gait recognition: A review. *Sensors* **2015**, *15*, 22089–22127. [CrossRef]
27. Yun, X.; Bachmann, E.R. Design, implementation, and experimental results of a quaternion-based Kalman filter for human body motion tracking. *IEEE Trans. Robot.* **2006**, *22*, 1216–1227. [CrossRef]
28. Liu, T.; Inoue, Y.; Shibata, K. Development of a wearable sensor system for quantitative gait analysis. *Measurement* **2009**, *42*, 978–988. [CrossRef]
29. Renaudin, V.; Susi, M.; Lachapelle, G. Step length estimation using handheld inertial sensors. *Sensors* **2012**, *12*, 8507–8525. [CrossRef]
30. Schepers, H.M.; Koopman, H.F.; Veltink, P.H. Ambulatory assessment of ankle and foot dynamics. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 895–902. [CrossRef]
31. Sabatini, A.M.; Martelloni, C.; Scapellato, S.; Cavallo, F. Assessment of walking features from foot inertial sensing. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 486–494. [CrossRef]
32. Favre, J.; Aissaoui, R.; Jolles, B.M.; de Guise, J.A.; Aminian, K. Functional calibration procedure for 3D knee joint angle description using inertial sensors. *J. Biomech.* **2009**, *42*, 2330–2335. [CrossRef] [PubMed]
33. Seel, T.; Raisch, J.; Schauer, T. IMU-based joint angle measurement for gait analysis. *Sensors* **2014**, *14*, 6891–6909. [CrossRef] [PubMed]
34. Rucco, R.; Sorriso, A.; Liparoti, M.; Ferraioli, G.; Sorrentino, P.; Ambrosanio, M.; Baselice, F. Type and location of wearable sensors for monitoring falls during static and dynamic tasks in healthy elderly: A review. *Sensors* **2018**, *18*, 1613. [CrossRef] [PubMed]
35. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1995; Volume 3361.
36. Abdel-Hamid, O.; Mohamed, A.r.; Jiang, H.; Penn, G. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 4277–4280.
37. Abdel-Hamid, O.; Mohamed, A.r.; Jiang, H.; Deng, L.; Penn, G.; Yu, D. Convolutional neural networks for speech recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 1533–1545. [CrossRef]
38. Liu, J.; Pan, Y.; Li, M.; Chen, Z.; Tang, L.; Lu, C.; Wang, J. Applications of deep learning to MRI images: A survey. *Big Data Min. Anal.* **2018**, *1*, 1–18.
39. Kong, Y.; Gao, J.; Xu, Y.; Pan, Y.; Wang, J.; Liu, J. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. *Neurocomputing* **2019**, *324*, 63–68. [CrossRef]
40. Zeng, M.; Li, M.; Fei, Z.; Yu, Y.; Pan, Y.; Wang, J. Automatic ICD-9 coding via deep transfer learning. *Neurocomputing* **2019**, *324*, 43–50. [CrossRef]
41. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? Advances in neural information processing systems. *arXiv* **2014**, arXiv:1411.1792.
42. Lee, J.B.; Rossi, R.A.; Kim, S.; Ahmed, N.K.; Koh, E. Attention models in graphs: A survey. *ACM Trans. Knowl. Discov. Data (TKDD)* **2019**, *13*, 1–25. [CrossRef]
43. Ba, J.; Mnih, V.; Kavukcuoglu, K. Multiple object recognition with visual attention. *arXiv* **2014**, arXiv:1412.7755.
44. Peng, Y.; He, X.; Zhao, J. Object-part attention model for fine-grained image classification. *IEEE Trans. Image Process.* **2017**, *27*, 1487–1500. [CrossRef] [PubMed]
45. Luong, M.T.; Pham, H.; Manning, C.D. Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 7–11 September 2015; pp. 1412–1421.

46. Yin, W.; Schütze, H.; Xiang, B.; Zhou, B. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 259–272. [CrossRef]
47. Zeyer, A.; Irie, K.; Schlüter, R.; Ney, H. Improved training of end-to-end attention models for speech recognition. *arXiv* **2018**, arXiv:1805.03294.
48. Zhang, D.; Yao, L.; Chen, K.; Wang, S. Ready for Use: Subject-Independent Movement Intention Recognition via a Convolutional Attention Model. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Turin, Italy, 22–26 October 2018; pp. 1763–1766.
49. Zhang, X.; Yao, L.; Huang, C.; Wang, S.; Tan, M.; Long, G.; Wang, C. Multi-modality sensor data classification with selective attention. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 3111–3117.
50. Zeng, M.; Gao, H.; Yu, T.; Mengshoel, O.J.; Langseth, H.; Lane, I.; Liu, X. Understanding and improving recurrent networks for human activity recognition by continuous attention. In Proceedings of the 2018 ACM International Symposium on Wearable Computers, Singapore, 8–12 October 2018; pp. 56–63.
51. Wang, K.; He, J.; Zhang, L. Attention-Based Convolutional Neural Network for Weakly Labeled Human Activities' Recognition With Wearable Sensors. *IEEE Sens. J.* **2019**, *19*, 7598–7604. [CrossRef]
52. Chiewchanwattana, S.; Lursinsap, C. FI-GEM networks for incomplete time-series prediction. In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290), Honolulu, HI, USA, 12–17 May 2002; Volume 2, pp. 1757–1762.
53. Lu, Q.; Pang, L.; Huang, H.; Shen, C.; Cao, H.; Shi, Y.; Liu, J. High-G calibration denoising method for high-G MEMS accelerometer based on EMD and wavelet threshold. *Micromachines* **2019**, *10*, 134. [CrossRef] [PubMed]
54. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
55. Casale, P.; Pujol, O.; Radeva, P. Human activity recognition from accelerometer data using a wearable device. In Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis, Las Palmas de Gran Canaria, Spain, 8–10 June 2011; pp. 289–296.
56. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International Workshop on Ambient Assisted Living*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 216–223.
57. Hammerla, N.Y.; Halloran, S.; Plötz, T. Deep, convolutional, and recurrent models for human activity recognition using wearables. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 1533–1540.
58. Edel, M.; Köppe, E. Binarized-blstm-rnn based human activity recognition. In Proceedings of the 2016 International conference on indoor positioning and indoor navigation (IPIN), Sapporo, Japan, 18–21 September 2016; pp. 1–7.
59. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]
60. Hoos, H.; Leyton-Brown, K. An efficient approach for assessing hyperparameter importance. In Proceedings of the International Conference on Machine Learning, Bejing, China, 22–24 June 2014; pp. 754–762.