A Deep Learning Approach on Gender and Age Recognition using a Single Inertial Sensor

1st Yingnan Sun
Dept. of Computing
Imperial College London
London, UK
y.sun16@imperial.ac.uk

2nd Frank P.-W. Lo
The Hamlyn Center
Imperial College London
London, UK
po.lo15@imperial.ac.uk

3rd Benny Lo
The Hamlyn Center
Imperial College London
London, UK
benny.lo@imperial.ac.uk

***Abstract*—Extracting human attributes, such as gender and age, from biometrics have received much attention in recent years. Gender and age recognition can provide crucial information for applications such as security, healthcare, and gaming. In this paper, a novel deep learning approach on gender and age recognition using a single inertial sensors is proposed. The proposed approach is tested using the largest available inertial sensor-based gait database with data collected from more than 700 subjects. To demonstrate the robustness and effectiveness of the proposed approach, 10 trials of inter-subject Monte-Carlo cross validation were conducted, and the results show that the proposed approach can achieve an averaged accuracy of 86.6%±2.4% for distinguishing two age groups: teen and adult, and recognizing gender with averaged accuracies of 88.6%±2.5% and 73.9%±2.8% for adults and teens respectively.**

***Index Terms*—Age recognition, gender recognition, soft biometrics, gait biometrics, inertial sensors**

I. Introduction

In recent years, biometrics has been widely adopted in security applications such as mobile phone authentication. These applications are often focused on the uniqueness of hard biometrics - typical physiological traits, such as face and fingerprint, and behavioral traits, such as gait and voice [1]. Although hard biometrics are the core metrics for biometric systems, much research has showed that soft biometrics, such as age, skin color, and gender, can also improve the performance of biometric systems [2]–[4]. Soft biometrics, especially gender and age, can also provide personal specific information which could benefit in business, healthcare, robotic, and gaming applications. The state-of-the-art human gender and age recognition methods are often based on the static facial features [5] or whole body images [6], and dynamic features from voice [7] and gait [8]. Gait, the walking pattern of a person, can be captured by a camera from a distance, or captured by inertial sensors attached to the person [9]. Similar to face and iris, a gait pattern of a person is unique because bones, joints and muscles used for walking are very different from person to person.

Gender and age recognition using gait sequences captured by a camera has gained more popularity in the past few years. For example, Li et al [10] proposed a vision-based gender recognition method using different components of human walking silhouettes, and Makihara et al [11] proposed a vision-based age estimation method also using human walking silhouettes. Vision-based gender and age recognition is robust and effective in a controlled environment, such as a situation where a person walking in front of a camera at a fixed location. It is difficult to extract gait silhouettes if the target subject is occluded by another person or objects. These problems can be solved by using inertial sensors to capture gait biometrics. Although inertial sensors have to be attached to or carried by the person, they can be used in uncontrolled environments, such as a group of people walking closely together in a pedestrian area, a situation where vision-based approaches could not be applied.

This research is supported by EPSRC project - SenTH+ - PETRAS - IoT EP/N023242/1 Cyber Security of the Internet of Things.

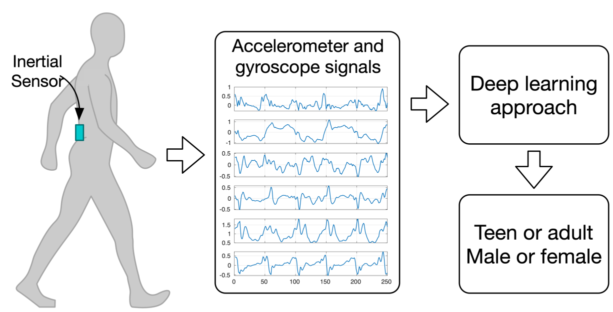


Fig. 1: Illustration of the proposed inertial sensor-based gender and age recognition approach

Although inertial sensor-based gait biometrics is widely used for authentication, it has not been fully exploited for gender and age recognition. Riaz et al [12] studied the estimation of gender, age and height using a trained random forest classifiers with hand-crafted features of single-step inertial signal recordings. The dataset collected by the authors consisted of only 26 subjects with a balanced gender ratio and an averaged age of 48.1±12.7 years. The authors have demonstrated the feasibility of gender and age recognition using inertial sensors on a small population using 10-fold cross validation. But the hand-crafted feature extraction technique used by the authors suffered significant performance drop (from over 85% to around 65%) when using inter-subject cross validation for age estimation, failing to show the robustness of the proposed approach.

Furthermore, Jain and Kanhangad [13] studied gender classification using a built-in inertial sensors of smartphones when users are walking at different speeds. The authors also used hand-crafted features in the proposed approach, and tested it on two datasets containing 46 and 63 subjects separately. The subjects in these datasets are mostly adults and age from 19 to 36 years, and people younger than 19 years or older than 36 years were not considered in the experiments. Another related work carried out by Bales et al [14] also based on inertial sensors, however, the sensors were installed beneath the floor of a building instead of attaching to the body. The authors proposed a machine learning based gender classification approach using data collected from only fifteen subjects.

In this paper, we propose a deep learning approach for gender and age recognition using a single inertial sensor attached to the lower back of the subjects. Deep learning approaches are widely exploited in vision-based gender recognition, but to our best knowledge, it has not been used for inertial sensor-based gender and age recognition. The proposed approach was evaluated on the largest inertial sensor-based gait database available [15], which has inertial data collected from 744 subjects. 640 out of 744 subjects (whose gender information is provided) with a gender ratio of 1:1 and age range from 2 to 79 years, were used in the experiments. 10 trials of inter-subject Monte-Carlo cross validation were carried out for all the experiments to demonstrate the robustness and effectiveness of the proposed approach.

II. Methodology

As shown in Fig. 1, the proposed deep learning-based age and gender recognition approach requires only a single inertial sensor attached to the lower back of the subject. The deep learning approach consists of three blocks as shown in Fig. 2: a signal pre-processing block, a convolutional feature extraction block, and a fully connected classifier. In the signal pre-processing block, a sliding window is applied to the accelerometer and gyroscope signals collected from the inertial sensor. Then, the partitioned signal data is fed into the convolutional feature extraction block to extract features inside each sliding window. At last, a 2-class fully connected classifier will then classify either teen or adult, or male or female.

A. Convolutional Feature Extraction

The partitioned signal data fed into the first layer is in a 3D matrix form of $(B \times W \times N)$, in which batch size $B = 10$, sliding window size $W = 100$ (which is 1 second), and the number of channels $N = 6$ (i.e. 3-axis accelerometer and 3-axis gyroscope data). In the first 1D convolutional layer, there are 200 filters/kernels with kernel size set to 5 and stride set to 2. The output feature map of the first convolutional layer has a shape of $(25 \times 50 \times 200)$, and it is fed into a max pooling layer whose pool size is set to 2 and stride is set to 3. The dimension of feature map is reduced based on the maximum value of

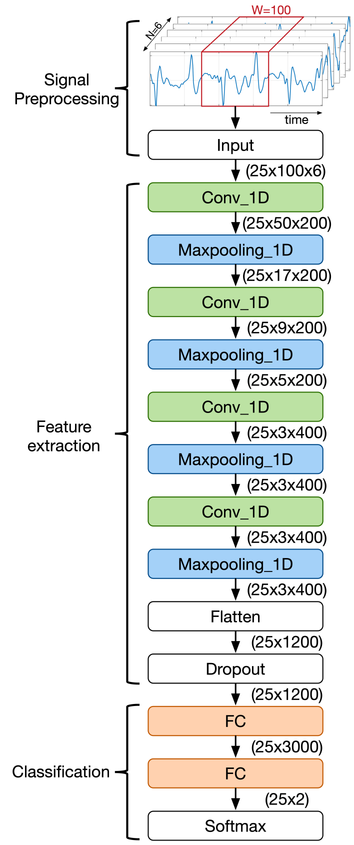


Fig. 2: Network architecture of the proposed deep learning approach

each pool, and its shape is reduced to $(25 \times 17 \times 200)$. The same feature extraction procedure is repeated 3 more times as indicated in Fig. 2. The mathematical expression of the output feature maps of l^{th} $(l = [1, 2, 3, 4])$ 1D convolutional layer is

$$\theta^l = \left[\gamma_1^l(j), \gamma_2^l(j), \gamma_3^l(j), \cdots, \gamma_m^l(j)\right] \quad (1)$$

and feed forwarding process for each neuron is

$$\gamma_m^l(j) = \beta^l(j) + \sum_{i=1}^{i \leq K^l} (\theta_{i,j}^{l-1} w_1(l,m,i) + \theta_{i,j+1}^{l-1} w_2(l,m,i)) \quad (2)$$

where $\gamma_m^l(j)$ is the j^{th} neuron in the m^{th} kernel of the l^{th} 1D convolutional layer, and $w_k(l,m,i)$ refers to the weights of the l,m,i^{th} filter used in the neural network. $\beta^l(j)$ is the bias of the j^{th} neuron in the l^{th} 1D convolutional layer. K^l is the number of kernel being used in the l^{th} 1D convolutional layer, and $\theta_{i,j}^l$ is the i,j^{th} element of the output feature maps from the max pooling layer following the $(l-1)^{th}$ 1D convolutional layer. The output of the final max pooling layer is flatten to a shape of (25×1200), and it passes through a dropout layer with a keep probability of 0.95 to prevent overfitting. Then, the final feature map is fed into a classifier with two fully connected layers to produce the final output, which is the probabilities of the two classes (teen/adult or male/female). The softmax layer at the end calculates the loss, which is used for optimizing the

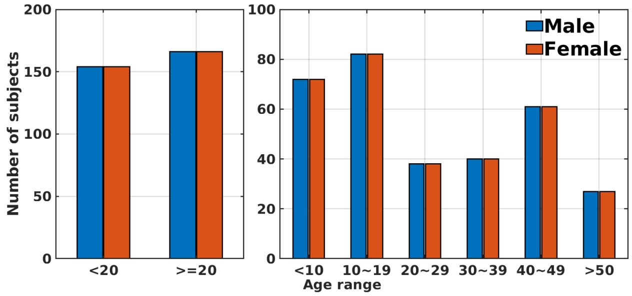


Fig. 3: Distributions of gender and age of the selected subjects from the OU-ISIR gait database

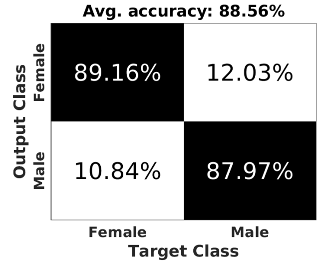
neurons of the network in the training phase.

B. Database

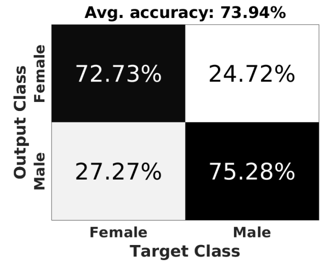
To evaluate the proposed gender and age recognition approach on the whole generation, the largest available inertial sensor-based OU-ISIR gait database [15] was used in the experiments. To ensure a balanced gender ratio in all age groups, we followed protocol 5.6 [15], in which 640 subjects are selected. The distributions of gender and age of the selected subjects as shown in Fig. 3. The subplot on the left side shows the number of subjects of teens and adults respectively, and the subplot on the right side shows the number of subjects for 6 age groups. Each subject has two sequences of level walking inertial sensor recordings, which contains about 7 to 12 steps.

C. Training and Testing

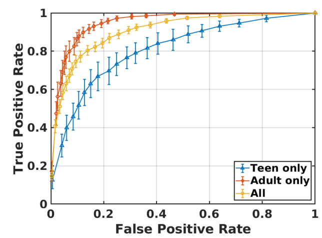
To demonstrate the robustness and effectiveness of the proposed approach, inter-subject Monte-Carlo cross validation was carried out 10 times, and the means and standard deviations of the results across 10 trials are presented in the experimental results section for both gender and age recognition. For age recognition, there were two classes: teen (age$<$20) and adult (age$\geq$20). In each trial, 70% of the subjects in each class were randomly selected for training, and the rest was reserved for testing. For gender recognition, two experiments were carried out: 1) in the first experiment, all the subjects, either teens or adults, were trained in the same network; 2) in the second experiment, gender recognition for teens and adults was carried out using two separate networks (i.e. a subject who is below 20 will not be considered at all in the network for recognizing gender for adults). In addition, the data of the selected testing subjects were not used for training the network, which eliminates the possibilities of over-fitting the network for better testing results. It can be also proven that the proposed approach is capable of recognizing gender and age of subjects from other dataset with high accuracy. As shown in Fig. 2, a sliding window is applied to partition the sequences of inertial sensor recordings into slices of data with shape (100$\times$6). Then, a batch of 25 slices is fed into the network together in each iteration. The initial learning rate of the network is set to 0.001, and it decays by 4% after every 10 thousand iterations. The weights and biases in the convolutional and fully connected layers are randomly



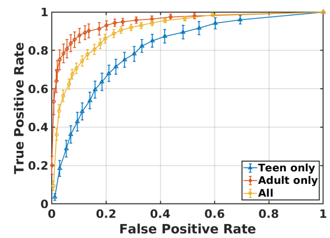
(a) Confusion matrix for adults



(b) Confusion matrix for teens



(c) ROC curves for female



(d) ROC curves for male

Fig. 4: Experimental results for gender recognition: (a) and (b) show confusion matrices (sum up for 10 trials) for adult-only and teen-only respectively. (c) and (d) show ROC curves for female class and male class separately, with point-wise confidence bounds calculated for 10 trials

TABLE I: Gender recognition results (averaged for 10 trials)

		Accuracy	Sensitivity	Specificity	F1-score
per recording	Teen	0.739±0.028	0.766±0.092	0.713±0.066	0.744±0.043
	Adult	**0.886±0.025**	**0.878±0.053**	**0.893±0.058**	**0.885±0.026**
	All	0.828±0.028	0.822±0.040	0.833±0.042	0.827±0.029
per sliding window	Teen	0.702±0.017	0.728±0.074	0.676±0.058	0.706±0.033
	Adult	0.836±0.016	0.821±0.055	0.853±0.052	0.837±0.019
	All	0.787±0.015	0.788±0.036	0.787±0.028	0.790±0.019

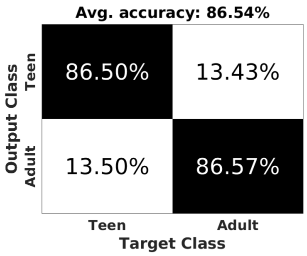
initialized and optimized using ADAM optimization algorithm [16], and the training process is stopped when reaching 5 epochs.

III. Experimental Results

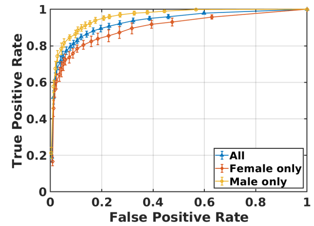
In this section, the performance of the proposed approach is presented, using evaluation matrices for typical biometric systems, including confusion matrices and Receiver Operating Characteristic (ROC) curves. The recognition performance per sliding window and per recording was reported using accuracy, sensitivity, specificity, and F1-score, which are averaged across 10 trials.

A. Gender Recognition

As aforementioned, there are two experiments conducted for gender recognition using the proposed approach: the first one using the entire dataset and the second one splits the dataset into teen group and adult group for separate training. This is to investigate how the age of the subject affects the performance of the proposed approach on gender recognition. Fig. 4(a) and (b) show confusion matrices for gender recognition using



(a) Confusion matrix



(b) ROC curves

Fig. 5: (a) shows the confusion matrix of 10 trials for age recognition, and (b) shows the ROC curves for all, female-only and male-only respectively for age recognition with point-wise confidence bounds calculated across 10 trials

adult-only dataset and teen-only dataset respectively. The proposed approach can distinguish gender for adults (age≥20) with an averaged accuracy of 88.56% across 10 trials, whereas it performs poorly, with only an averaged accuracy of 73.94% for teens. This is expected because the muscle and bones of teens are still growing, which makes their gait less predictive. In addition, female is more recognizable then male for adults, and male is more recognizable than female for teens. Fig. 4(c) and (d) show the ROC curves for the proposed approach for gender recognition, and they also indicate that gender is more distinctive for adults than teens. More details for the gender recognition performance are listed in Table I, where accuracy, sensitivity, specificity, and F1-score for teen-only, adult-only, and all age group across 10 trials are presented. The results for each inertial data recording, which contains about 5 to 10 steps, are aggregated from the recognition results from each sliding window. Therefore, accuracy per recording is better than that of per sliding window.

B. Age Recognition

The proposed approach is capable of distinguishing two age groups: teen and adult. The confusion matrix in Fig. 5(a) show that the average accuracies for teens and adults are 85.50% and 86.57% respectively. It indicates that proposed approach has no bias towards either age group. Fig. 5(c) shows the ROC curves of the age recognition for all the subjects, female subjects, and male subjects. The proposed approach performs better for age recognition on male subjects than female subjects. This is also shown in Table II, where the age recognition accuracy of male subjects per recording is 88.7%, 4.4% higher than that of female subjects. Moreover, age recognition using the proposed approach for male subjects has higher sensitivity but less specificity than female subjects.

IV. Conclusion

In this paper, a deep learning approach on gender and age recognition using a single inertial sensor is proposed, and the proposed approach was tested on the largest available inertial sensor-based gait database. The results from 10 trials of inter-subject Monte-Carlo cross validation show that the proposed approach is robust and effective. The proposed approach is capable of recognizing either teen or adult with an averaged accuracy of 86.6%±2.4%, and recognizing gender with averaged accuracies of 88.6%±2.5% and 73.9%±2.8% for adults and teens separately.

TABLE II: Age recognition results (averaged for 10 trials)

		Accuracy	Sensitivity	Specificity	F1-score
per recording	Female	0.843±0.028	0.785±0.033	0.896±0.054	0.827±0.028
	Male	**0.887±0.026**	**0.913±0.029**	**0.864±0.037**	**0.885±0.026**
	All	0.866±0.024	0.849±0.025	0.880±0.042	0.857±0.023
per sliding window	Female	0.775±0.025	0.728±0.050	0.814±0.052	0.743±0.027
	Male	0.830±0.019	0.867±0.026	0.798±0.033	0.827±0.019
	All	0.802±0.019	0.798±0.031	0.806±0.040	0.787±0.018

References

[1] A. Mahfouz, T. M. Mahmoud, and A. S. Eldin, "A survey on behavioral biometric authentication on smartphones," *Journal of Information Security and Applications*, vol. 37, pp. 28–37, 2017.

[2] R. Zewail, A. Elsafi, M. Saeb, and N. Hamdy, "Soft and hard biometrics fusion for improved identity verification," in *The 2004 47th Midwest Symposium on Circuits and Systems, 2004. MWSCAS '04.*, vol. 1, Jul. 2004, pp. I–225.

[3] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739–777, Jan. 2011.

[4] M. Abreu and M. Fairhurst, "Enhancing identity prediction using a novel approach to combining hard- and soft-biometric information," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 599–607, Sep. 2011.

[5] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, Dec. 2014.

[6] H. A. Perlin and H. S. Lopes, "Extracting human attributes using a convolutional neural network approach," *Pattern Recognition Letters*, vol. 68, pp. 250–259, 2015.

[7] M. H. Sedaaghi, "A comparative study of gender and age classification in speech signals," *Iranian Journal of Electrical and Electronic Engineering*, vol. 5, no. 1, pp. 1–12, 2009.

[8] H. Hediyeh, T. Sayed, and M. H. Zaki, "Use of spatiotemporal parameters of gait for automated classification of pedestrian gender and age," *Transportation Research Record*, vol. 2393, no. 1, pp. 31–40, 2013.

[9] Y. Sun, C. Wong, G.-Z. Yang, and B. Lo, "Secure key generation using gait features for body sensor networks,"

[10] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 145–155, Mar. 2008.

[11] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in *Proceedings of 2011 International Joint Conference on Biometrics*, IEEE, 2011, pp. 1–6.

[12] Q. Riaz, A. Vogele, B. Kruger, and A. Weber, "One small step for a man: Estimation of gender, age and height from recordings of one step by a single inertial sensor," *Sensors*, vol. 15, no. 12, pp. 31 999–32 019, 2015.

[13] A. Jain and V. Kanhangad, "Gender classification in smartphones using gait information," *Expert Systems with Applications*, vol. 93, pp. 257–266, 2018.

[14] D. Bales, P. A. Tarazaga, M. Kasarda, D. Batra, A. Woolard, J. D. Poston, and V. S. Malladi, "Gender classification of walkers via underfloor accelerometer measurements," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1259–1266, 2016.

[15] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication," *Pattern Recognition*, vol. 47, no. 1, pp. 228–237, 2014.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.