Poster Abstract: Using Deep Learning to Classify The Acceleration Measurement Devices

Yuezhong Wu
University of New South Wales
Australia
yuezhong.wu@student.unsw.edu.au

Carlos Ruiz
Carnegie Mellon University
USA
carlosrd@cmu.edu

Shijia Pan
University of California, Merced
USA
span24@ucmerced.edu

Hae Young Noh
Stanford University
USA
noh@stanford.edu

Mahbub Hassan
University of New South Wales
Australia
mahbub.hassan@unsw.edu.au

Pei Zhang
Carnegie Mellon University
USA
peizhang@cmu.edu

Wen Hu
University of New South Wales
Australia
wen.hu@unsw.edu.au

ABSTRACT

Recent work has shown that two wearable devices worn on the same user can exploit gait as a secret source to generate a common key for secure pairing. The main threat of using gait comes from side-channel attackers who can use cameras to record the walking user and extract accelerations from the video to pair with legitimate devices. We propose a novel pre-step that uses a CNN-LSTM deep learning model to classify the acceleration measurement devices, i.e., between IMU vs. Camera. We prototype the pre-step and evaluate it using real subjects. Our results show that the proposed pre-step can achieve high classification success rates. The experiments with different cut-off frequencies show that the higher acceleration frequencies appear to contain more distinguishable features to classify camera from IMU.

CCS CONCEPTS

• **Security and privacy** → Mobile and wireless security;

KEYWORDS

Side-channel Attack, Deep Learning, CNN-LSTM

1 INTRODUCTION

Intuitively, two independent inertial motion sensors (IMUs) located at different body locations, e.g., chest and upper arm, would simultaneously capture the common gait of the user. Recent works in [3] exploited this observation to dynamically pair wearable devices, where devices use the common gait signal to independently produce the same cryptographic key to pair with each other. However, using gait as a secret source raises concerns about malicious attacks such as passive attack, mimicking attack and video-based side-channel attack. Passive and mimicking attacks are not capable of producing sufficiently accurate accelerations to pair with the legitimate IMUs in [3]. However, it is reported that the side-channel attacker can extract accurate accelerations from the video to pair with the legitimate devices in [1].

To prevent the video-based side-channel attack, we propose a novel pre-step that employs a CNN-LSTM deep learning model to classify the device (IMUs or videos) where the accelerations are measured. In the proposed pre-step, the accelerations from videos and IMUs will be inputted to a trained CNN-LSTM model. The model will give the prediction of the class that accelerations belong to. The accelerations measured from videos will be abandoned from further applications. In [1], 12 Hz is used to low pass filter the accelerations as it is the upper bound of human body motion frequency. In our experiments, we test a range of cut-off frequencies and show that the CNN-LSTM model achieves higher classification success rates with higher cut-offs($\geq$ 30) when classifying the accelerations from IMU and video.

2 CNN-LSTM MODEL

We employ a specifically designed CNN-LSTM model to do the classification. The CNN-LSTM is a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) recurrent network [2]. CNN-LSTM adds CNN layers before the LSTM layers for a feature transformation of input data. Then those transformed features are fed to the LSTM layers. A recent study shows that CNN-LSTM outperforms its LSTM counterparts in activity recognition [2]. So we choose CNN-LSTM as the deep learning model used in this paper.

The structure of the CNN-LSTM model in this paper is given in Figure 1. Firstly, we employ two 1d-convolutional (Conv1D) layers to extract features. A maxpooling layer is to select the most important features. The Flatten layer reshapes the output from maxpooling layer into 1-dimension time series data. Then time series data is fed to the LSTM layer. After the LSTM layer, we add two fully connected layers that are also called as dense layers. The activation function of the last dense layer is Softmax while all other layers use ReLu. We insert two dropout layers after the second Conv1D layer and LSTM layer to prevent over-fitting. The dropout rate is empirically set as 0.5, which means that 50% of inputs units of the dropout layers will be set to zeros.

978-1-7281-5497-8/20/$31.00 ©2020 IEEE
DOI 10.1109/IPSN48710.2020.00-11

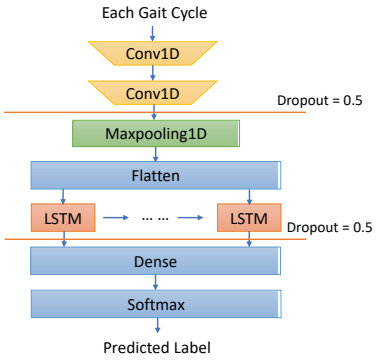


Figure 1: The CNN-LSTM model structure

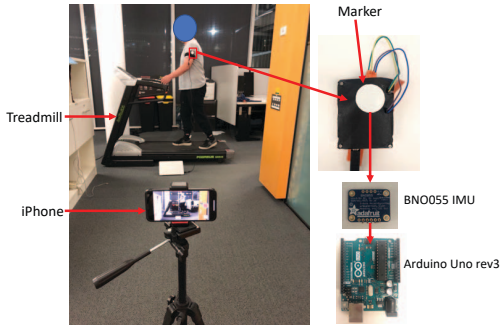


Figure 2: The setup of data collection

3 EVALUATION

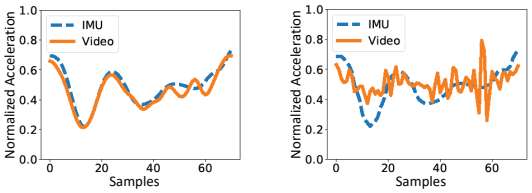
3.1 Data Collection

In the experiments, we collect data from 2 subjects [1]. As shown in figure 2, each subject is wearing a BNO055 IMU [2] on the upper arm with a marker attached upon the IMU. The marker is a paper-made white circle. The IMU is connected to an Arduino Uno REV3 board [3]. Each subject is walking on a treadmill for about one minute. We collect 3-axis acceleration and quaternion from the IMU at 100Hz. The camera with two lenses of an iPhone X is used to record the walking subject in slow-motion mode. The frame rate is 120Hz and the resolution is 720p. The iPhone is mounted on a tripod at a height of 1 meter. The tripod is placed 3 meters away from the treadmill. We use a software called Tracker [4] to extract the accelerations of the white circle marker.

3.2 Data Preprocessing

For IMUs, we transfer the accelerations from the IMU's coordinate system to the world coordinate system. We only use the accelerations along the gravity direction. For Videos, we directly use the acceleration along the gravity direction extracted by Tracker. After applying a low pass filter, we normalize the data to the range of 0 and 1. Then we do gait cycle detection. Each gait cycle is one window in the dataset. Finally, we label gait cycles from IMU as class 0 and gait cycles from the video as class 1.

[1] Ethical approval for carrying out this experiment has been granted by the corresponding organization (Approval Number HC17008)
[2] https://www.bosch-sensortec.com/products/smart-sensors/bno055.html
[3] https://store.arduino.cc/usa/arduino-uno-rev3
[4] https://physlets.org/tracker/



(a) Cut-off freqeuncy = 12Hz (b) Cut-off freqeuncy = 40Hz

Figure 3: Accelerations of One Gait cycle after low pass filters using different Cut-off frequencies

Table 1: Classification success rate of classification

Cut-off frequency of low pass filter	12Hz	20Hz	30Hz	40Hz	raw data
Classification success rate	60.34%	75.81%	90.76%	93.33%	94.59%

3.3 Results

The comparison of accelerations of one gait cycle from IMU and video is depicted in Figure 3. When the cut-off of the low pass filter is set as 12 Hz. we can see that the accelerations of IMU and video are quite similar to each other. Then the cut-off is increased to 40 Hz, the difference between IMU and video becomes much more obvious.

Based on the above observation, We apply different cut-offs on the data to train and test the CNN-LSTM model. The results are under 3-fold cross-validation. We use the classification success rate to measure the performance of the model. In table 1, we can see that the success rate is improved with the increasing cut-off. When the cut-off is 12 Hz, it is only 60.34%. It can reach higher than 90% with the cut-off $\geq$ 30 and finally achieve the highest of 94.59% when using the raw data without any low pass filter. The results show that higher acceleration frequencies appear to contain more distinguishable features to classify camera from IMU. To find the frequency that maximize the classification success rate and minimize the information overhead in frequency $\geq$ 12 Hz will be the future work.

REFERENCES

[1] A. Bruesch, L. Nguyen, D. Schürmann, S. Sigg, and L. C. Wolf. 2019. Security Properties of Gait for Mobile Device Pairing. *IEEE Transactions on Mobile Computing* (2019), 1–1. https://doi.org/10.1109/TMC.2019.2897933
[2] Juan C. Núñez, Raúl Cabido, Juan J. Pantrigo, Antonio S. Montemayor, and José F. Vélez. 2018. Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition. *Pattern Recognition* 76 (2018), 80 – 94. https://doi.org/10.1016/j.patcog.2017.10.033
[3] W. Xu, G. Revadigar, C. Luo, N. Bergmann, and W. Hu. 2016. Walkie-Talkie: Motion-Assisted Automatic Key Generation for Secure On-Body Device Communication. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 1–12. https://doi.org/10.1109/IPSN.2016.7460726