PERSON IDENTIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS ON SHORT-TERM SIGNALS FROM WEARABLE SENSORS

G. Retsinas, P. P. Filntisis, N. Efthymiou, E. Theodosis, A. Zlatintsi and P. Maragos

School of ECE, National Technical University of Athens, 15773 Athens, Greece

{gretsinas,filby,nefthymiou}@central.ntua.gr, {nzlat, maragos}@cs.ntua.gr

ABSTRACT

In this work, we explore the discriminating ability of short-term signal patterns (e.g. few minutes long) with respect to the person identification task. We focus on signals recorded by simple wearable devices, such as smart watches, which can measure movements (accelerometer and gyroscope sensors) and biosignals (heart rate monitor). To address the person identification problem, we develop a deep neural network, based on one-dimensional convolutions, which receives raw signals from three different smartwatch sensors and predicts the person wearing the smartwatch. Experimental results indicate that even with signals from wearable sensors collected at intervals of only 10 minutes, different users can be identified with notably high accuracy, revealing the existence of distinct short-term patterns of movement and heart rate between different persons.

Index Terms— person identification, wearable sensors, deep convolutional neural networks, behavioral biometrics

1. INTRODUCTION

Person identification from behavioral and physiological signals can be employed in a variety of applications, ranging from authentication and cognitive biometrics [1, 2], to leveraging the identity of the person for offering personalized services [3]. In addition, person identification from wearable signals can also provide useful insights on patients with psychiatric disorders, by identifying the distinctive behavior of these individuals [4], i.e. creating a reliable behavioral profile, and consequently detect significant changes in behavior due to their condition.

Related works in the literature typically focus only on biosignals such as ECG (electrocardiogram), EEG (electroencephalogram), or EDR (electro dermal response) [2, 5, 6]. In [7], mobile accelerometer data were used for person identification, employing a simple set of features on the extracted waveforms, while in [3] the authors leveraged accelerometer data to identify the user of specific objects in a smart home. Buriro et al. [8] used smartwatches in order to extract an arm-movement fingerprint, when the person attempts to make specific gestures. So far the literature either focused on intense, distinct activities (arm or gait movement), hard-to-collect signals (ECG, EEG) or daily activities (at a macroscopic scale). In this work, we explore person identification from data collected by low-cost off-the-shelf sensors, such as a wearable smartwatch, at at fine-grained time scales (such as 5 or 10 minutes). Hence, we have created a dataset by continuously monitoring twenty volunteers, trying to discover short-term patterns, not only for intense movements, but over every activity/state (e.g. resting or sleeping).

To address the person identification task, we propose a Deep Neural Network (DNN) architecture, which takes as input three different wearable sensor measurements (namely linear acceleration, angular velocity, and heart rate) and identifies the person who generated the signals. The proposed neural network is based on consecutive one-dimensional convolutions leading to a fast and well-performing person identification approach. Furthermore, we discuss several problems and limitations that arise from using low-cost wearable devices, such as the existence of unique noise per sensor, which can easily lead to a misinterpretation of the results (identifying the sensor rather than the person).

The rest of the paper is organized as follows: in Section 2 we present the smartwatch sensors, the collected data, and the final dataset which was created in order to evaluate our approach on person identification. The proposed deep learning approach is presented in Section 3, describing in detail the architecture and the reasoning behind each component. Section 4 contains an extensive experimental exploration of the task at hand, while Section 5 focuses on the sensor noise interference problem. Finally, conclusive remarks are drawn at Section 6.

2. COLLECTED DATA

To validate the method presented in this work, wearable data from 20 volunteers, 20 to 35 years old, were recorded over a period of twenty or more days. The continuously recorded data included 1) angular velocity (gyroscope sensor), 2) linear acceleration (accelerometer sensor with gravity compensation), and 3) heart rate measurements acquired via photoplethysmogram. An example of the waveform measured by the accelerometer (for the x axis) is depicted in Fig. 1. The raw data were recorded with a sampling frequency of 20Hz and 5Hz for the kinetic sensors and the heart rate sensor, respectively. The wearable used was a Samsung Gear S3 Frontier smart watch. In addition, by taking advantage of the provided API, we automatically recorded the sleeping schedule for each subject, as well as their steps, aggregated over periods of 1 minute length. The volunteers were instructed to continuously wear the smart watch, except for a 2-hour period during each day, which was needed to charge it.

To organize this large collection of everyday raw sensor data, we split the recordings into intervals of predefined duration (e.g. 10 minutes). The selected duration should be few minutes long for two main reasons: 1) we want to explore the information in a micro-scale, which is usually discarded by mainstream analysis of biosignals (simple statistics) and 2) we want to create a dataset of significant size in order to efficiently apply machine learning techniques.

This research has been financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project code:T1EDK-02890, acronym: e-Prevention).

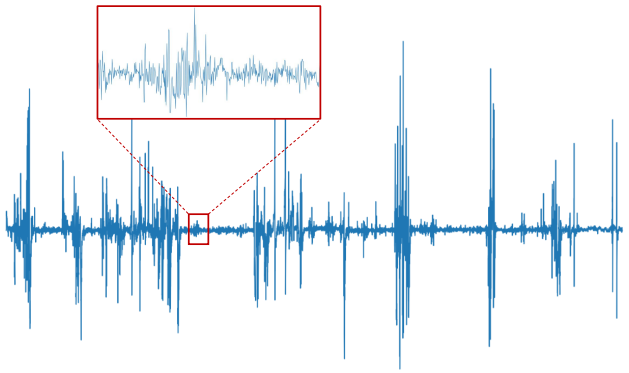

Fig. 1. Raw data (only x-axis) drawn from the accelerometer sensor.

For the sake of brevity, we annotate the raw signals of accelerometer, gyroscope and heart rate sensor as *acc*, *gyr* and *hrm* respectively. As we already have mentioned, we also store an indication of the sleeping and walking/running states. Such information would be very helpful in the upcoming analysis and therefore we distinguish three categories: 1) *sleep*: the person is asleep, 2) *walk*: the person is walking, 3) *other*: the person is neither sleeping nor walking (resting, standing, minor hand movements etc.).

After splitting the recordings, and discarding intervals where data were not recorded, either due to sensor/application malfunction, or because the subject was not wearing the sensor, we have a total of 60199 intervals, of which 17953 intervals correspond to *sleep*, 23235 intervals to *walk*, and 19011 intervals to *other*. The presented dataset is publicly available at `http://cvsp.cs.ntua.gr/research/person-id-dataset/`.

3. A DEEP NEURAL NETWORK APPROACH

To effectively address the problem of person identification we employ a Deep Neural Network (DNN), which takes as input the three different signals from the sensors of the smart watch and returns a class id, which corresponds to the person wearing the watch.

Due to the sequential nature of the input data, an immediate assumption would be that recurrent neural networks (RNNs) should be a perfect fit for the task. Nonetheless, RNNs have been proven to be cumbersome to train (convergence related problems) [9], while they cannot be parallelized, leading to high inference times. On the other hand, 1D convolutional layers are fully parallelized and can capture local context over sequences. When we stack several such layers we can effectively capture large context, resulting to a fixed sized representation of the entire sequence.

Based on the aforementioned observation, we build our architecture for the task at hand by stacking multiple 1D convolutional layers. Each convolutional layer is followed by a batch normalization layer (which speeds up convergence [10]) and a ReLU non-linearity. Similar architectures, based on 1D convolutions, have been previously used on ECG signals for compressing [11], detecting arrhythmia [12] or even predicting the sleep state [13].

An important aspect of this problem is the existence of three different streams of data, not necessarily synchronized. Data recorded by the accelerometer and the gyroscope may display small shifts in timestamps amongst them, while the heart rate is recorded at a different sampling frequency (5Hz versus 20Hz of *acc* and *gyr* data). In order to avoid merging the data streams into one common stream from the start (e.g. concatenate *acc* and *gyr* along with an interpolated version of *hrm*), we merge the data streams after they have been processed by a block of 1-d convolutional layers, specifically designed for the respective stream. In this case, this means that the blocks corresponding to *acc* and *gyr* data have more convolutional layers compared to the *hrm* block, to compensate for the signal length difference. For the same reason, they also have two pooling layers of stride 2. By merging a more abstract representation at a higher level, the network does not have to learn the synchronization variations of the data streams, bypassing this way adding an unnecessary complexity to the problem.

The merge is performed by concatenating the three different signals after transforming them into having the same length and dimension. After merging the generated feature sequences, we apply the final convolution block, consisting of multiple 1D convolutional layers along with batch-normalization and ReLU. All convolutional layers have a kernel of length 5. This block contains also dropout layers after each ReLU non-linearity in order to avoid over-fitting and converge to a generalizing solution. The sequence length is reduced per few convolutional layers using an average pooling operation. At the top of the network, we apply average pooling over the entire remaining sequence, followed by a fully connected layer. The basic architectural structure of the proposed network is visualized in Figure 2.

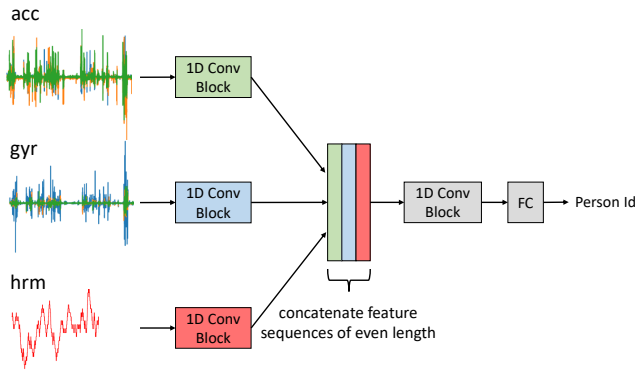

Fig. 2. Architecture of the proposed network. Each data stream will be individually processed before merging. The reported blocks consist of consecutive 1D convolutional layers along with batch-normalization and ReLU.

Note that in this work we do not perform an architecture exploration, since we propose a well-performing architecture, meticulously crafted for the specific problem, aiming to explore the discriminating capability of short-term signals. It is highly probable to attain slightly better accuracy results by fine-tuning the proposed architecture.

4. EXPERIMENTAL EXPLORATION

In this section, we explore the ability to accurately identify a person from micro-behavioral and physiological data with respect to the person's states (sleep, walk, other), the sensors' contribution, the duration of the recorded interval and the sampling frequency.

The dataset is split into train/test sets as follows: the intervals corresponding to the first 75% of the overall user's recorded days are added to the train set while the rest 25% are added to the test set. Approximately, for each user, we have 15 days for training and 5 days for testing. Note that the number of intervals per day varies between users. According to this setup, the resulting 10-minute intervals are ~ 45000 and ~ 15000 for training and testing, respectively.

For the upcoming experiments the architecture is the one described in Sec. 3 unless stated otherwise. The training procedure is the same for all experiments.[1]

4.1. State Dependency and Sensor Contribution

Given the sleep and walk attributes per interval, we can further explore the significance of these states with respect to the identification task. For an in-depth analysis, we report the sleep/walk importance for all possible sensor combinations. Specifically, in order to understand the contribution of each sensor's data on the person identification task, we also evaluate all seven possible combinations of the three sensors (drop the remaining conv block of the omitted signals). The results are accumulated in Table 1, lending to insightful observations: 1) Walk state provides the best performing results, while sleep the worst. This was to be expected, since intense movements provide valuable information. 2) Kinetic data (acc and gyr) help considerably the identification task compared to hrm data. However, if combined, extra performance boost is granted.

	sleep	walk	other	overall
acc	57.67%	86.36%	80.06%	75.34%
gyr	67.97%	85.23%	80.12%	78.13%
hrm	71.09%	44.29%	54.97%	55.84%
acc+gyr	85.85%	94.69%	90.34%	90.49%
acc+hrm	78.84%	89.11%	85.84%	84.78%
gyr+hrm	81.90%	88.38%	86.14%	85.72%
all	86.17%	94.17%	92.65%	91.50%

Table 1. Exploring the impact of person's states and sensors' combinations on the person identification accuracy.

4.2. Minimum Distinctive Interval Duration

The initial choice of 10 minutes intervals has a very clear intuition: select the interval small enough to constrain the network to learn micro-behaviors and not daily habits, and large enough in order to capture basic activities, such as walking, running etc. Even if this interval can be considered notably short to provide such accurate results, one would wonder *what is the minimum interval without significant drop in accuracy?* To this end, we consider a set of different interval lengths from 10 second up to 10 minutes. The dependency of the system's performance over the duration of the interval is depicted in Fig. 3. As one would suspect, performance is significantly dropping along with the interval duration/ Therefore, we conclude that our initial suggestion of using ten-minute intervals is indeed a helpful one.

4.3. Minimum Distinctive Sampling Frequency

After finding the minimum interval capable of extracting distinctive patterns, we consider the case of aggregating consecutive measurements. Since we have a rather high sampling frequency on our initial setting, it is important to explore the impact of sampling frequency to the task at hand. In this setting, we will explore different aggregation lengths over the initial raw data. Aggregation will be performed by averaging, i.e. we will apply a mean filter (also called moving average filter) which corresponds to a low-pass filter. As we can observe in Fig. 4, the identification accuracy is not affected significantly, even though it slightly drops, up to a window length of 4,

[1]Overall 80 epochs / SGD optimizer with 0.8 momentum and initial learning rate of 0.01 / learning rate decreased to 0.001 at 40 epochs.

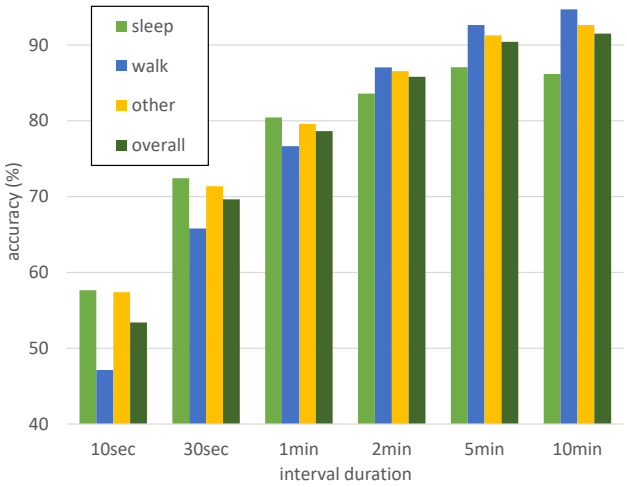

Fig. 3. Accuracy (%) dependency on interval duration.

which corresponds to an effective sampling frequency of 5Hz (concerning the accelerometer and the gyroscope sensors). Using larger window lengths, or smaller sampling frequency, leads to a notable drop in performance.

Note that for window size of 256, sleep data are better performing than walk data, a counter-intuitive result, which can be also observed in Fig. 3 for the case of the minimum interval of 10 seconds. This phenomenon will be revisited at the next section.

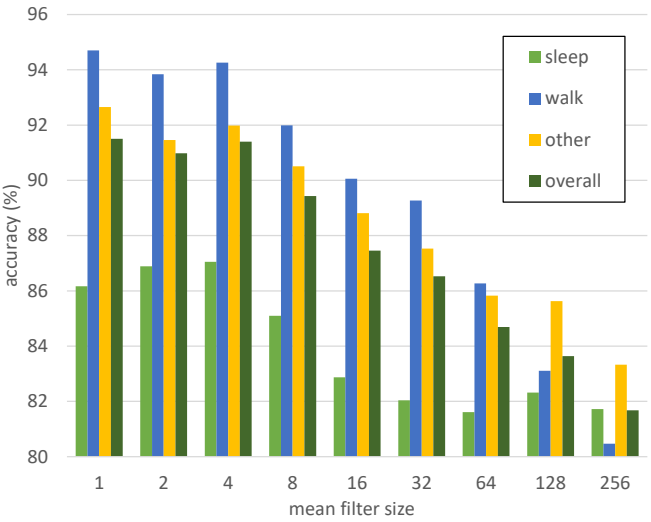

Fig. 4. Accuracy (%) dependency on aggregation length.

5. PERSON VS SENSOR

Experimental results indicate that we can distinguish every person with notably high accuracy, even if the person is asleep (86.2%). Such accuracy is alarming, since sleeping only contains minor changes in movement and heart rate at sparse intervals. Consequently, we questioned the validity of our results and the generalization capability of our network. One major concern was the fact that each person wears a unique watch, which generates a unique noise. If the network cannot find discriminative enough features at

a higher level, it would focus on classifying the sensors' noise. In other words, the actual measurements are coupled with a "fingerprint" noise which may distract the trained system. The case of sleep data is a good example with minor actual measurements, which can mislead the training procedure. In this section, we will evaluate the person vs sensor problem and propose solutions in order to train sensor-independent neural networks.

5.1. Validation Dataset for Sensor Coupling

In order to measure the entanglement of sensor and person information which was propagated to the output of the neural network and thus the identification results, we created a small validation dataset. Two of the initial volunteers were given watches belonging to different volunteers for two additional days and we collected these data as a distraction dataset. As an entanglement metric of the DNN classifier we will report the probability of detecting the person vs the probability of detecting the sensor, using a softmax activation of the DNN output. Evaluation on the validation dataset is performed by using an already trained model (on the original dataset) as a probability estimator. Table 2 contains the aforementioned probabilities of detecting the person or the sensor for the different person's states. Along with the overall system we evaluate the cases of using only kinetic data (*acc* and *gyr*) or only heart rate measurements.

The results strongly suggest that the initial system learned to classify the sensor along with the person. Intervals with intense movement were correctly classified, while intervals with no movement (e.g. sleep) are very frequently missclassified to the previous owner of the watch (i.e. sensor), confirming our initial hypothesis of entanglement. The network which uses only hrm measurements, contrary to the one based on kinetic measurements, is very robust towards correctly classifying the person, even when asleep, hinting that the sensor noise resides at the kinetic data.

sensors	source	sleep	walk	other	overall
all	person	0.07	0.54	0.14	0.20
	sensor	0.58	0.07	0.19	0.37
acc-gyr	person	0.04	0.34	0.12	0.16
	sensor	0.65	0.07	0.25	0.41
hrm	person	0.52	0.20	0.29	0.41
	sensor	0.14	0.08	0.07	0.11

Table 2. Person Vs Sensor probabilities for different states (sleep, walk, other) and sensor combinations.

5.2. Decoupling vs Contaminating

One straightforward solution to the aforementioned problem is to decouple the two signals, i.e. separate the actual measurements from the sensor noise. Nonetheless, this is a non trivial task, since we cannot accurately define the noise, nor its frequency. Further, even if we make the assumption that noise resides in the high frequency plane, we cannot simply apply a low-pass filter since useful information may exist in high frequency components. This claim is supported by Fig. 4, where mean filters of trivial size (e.g. 2 or 4) lead to a slight accuracy decrease for the case of the action "walk". In other words, high frequency components contain critical information for intense movements and cannot be discarded.

Our first approach is to define a method for accurately detecting the sensor noise and subsequently remove it. We assume that the noise has a much smaller magnitude compared to the real signal and is periodic. Under these assumptions, we can detect such low-magnitude periodic signals from the Short Time Fourier Transform (STFT) domain by simply detecting frequency values of similar low magnitude across the majority of the STFT windows. Next, we subtract these frequency components, corresponding to the periodic noise, from the initial magnitude of the STFT and finally we get the noise-independent signal with the inverse STFT. Performance-wise this approach did not solve the sensor independence problem and was abandoned.

A simple alternative to defining a precise decoupling technique is to contaminate the data with additional artificial noise, capable of covering the initial sensors' noise. We consider the following contamination techniques: 1) randomly zeroing a considerable percentage of small STFT components while adding white noise on the STFT magnitude, and 2) randomly zeroing *acc* or *gyr* data entirely, since they most likely contain significant sensor noise and the system should learn to identify the user even when one of the input streams is missing. The aforementioned techniques act as an augmentation of the raw signals. Trained networks under contaminated data present increased fluctuations in performance (compared to our initial approach) and therefore we exhibit the results by averaging over five networks (an ensemble of networks may solve this problem, but it will not be explored in this paper). Table 3 contains the accuracy and the person vs sensor probability of the contamination approach. The results indicate that we indeed created sensor-independent networks, albeit slightly decreasing the overall performance. Note that we do not aim to perfectly classify 10-minutes intervals, but to discover distinctive short-term patterns. Therefore the person identification can be performed in larger intervals (hour or day) through the aggregation of persons' probabilities over several smaller intervals.

sleep	walk	other	overall
84.32%	93.67%	89.19%	88.78%

(a)

source	sleep	walk	other	overall
person	0.29	0.45	0.26	0.27
sensor	0.09	0.03	0.08	0.08

(b)

Table 3. (a) Identification accuracy (b) Person VS Sensor probabilities for the contamination approach.

6. CONCLUSIONS AND FUTURE WORK

In this paper we studied the ability to successfully identify a person given raw data recorded by noisy wearable sensors and applying deep learning techniques. Our main focus was to determine if signals of small duration (e.g. 10 minutes) withhold discriminative patterns for the person identification task, which could lead to creating an interesting behavioral profile of each person. One major hindrance towards our goal was the existence of unique sensor noise, which misled the DNN to classify the sensor instead of the person. This was effectively addressed by contaminating the initial raw signals with artificial noise, capable of covering the sensors' noise. The proposed neural network reports outstanding identification results, especially when the user is walking, verifying our initial hypothesis of unique short-term patterns per person.

Concerning future work, we should extensively evaluate the sensor vs person problem on a larger scale in order to generate sensor-independent neural networks: extend the validation dataset, consider alternative noise removal/contaminating approaches or merging techniques (based on hrm sensor independence) or even consider the STFT domain as input for a modified neural network with 2D convolutional layers.

7. REFERENCES

[1] C. Camara, P. Peris-Lopez, J. E. Tapiador, and G. Suarez-Tangil, "Non-invasive multi-modal human identification system combining ecg, gsr, and airflow biosignals," *Journal of Medical and Biological Engineering*, vol. 35, no. 6, pp. 735–748, 2015.

[2] C. Camara, P. Peris-Lopez, and J. E. Tapiador, "Human identification using compressed ecg signals," *Journal of medical systems*, vol. 39, no. 11, pp. 148, 2015.

[3] J. Ranjan and K. Whitehouse, "Object hallmarks: Identifying object users using wearable wrist sensors," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 51–61.

[4] M. H. Aung, M. Matthews, and T. Choudhury, "Sensing behavioral symptoms of mental health and delivering personalized interventions using mobile technologies," *Depression and anxiety*, vol. 34, no. 7, pp. 603–609, 2017.

[5] H. Hussain, C.-M. Ting, F. Numan, M. N. Ibrahim, N. F. Izan, M.M. Mohammad, and H. Sh-Hussain, "Analysis of ecg biosignal recognition for client identifiction," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2017, pp. 15–20.

[6] R. Palaniappan and D. P. Mandic, "Eeg based biometric framework for automatic identity verification," *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 49, no. 2, pp. 243–250, 2007.

[7] W. U. Rehman, A. Laghari, and Z. Memon, "Exploiting smart phone accelerometer as a personal identification mechanism," *Mehran University Research Journal of Engineering & Technology*, vol. 34, no. S1, 2015.

[8] A. Buriro, R. Van Acker, B. Crispo, and A. Mahboob, "Airsign: A gesture-based smartwatch user authentication," in *2018 International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2018, pp. 1–5.

[9] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*, 2013, pp. 1310–1318.

[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[11] O. Yildirim, R. San Tan, and U. R. Acharya, "An efficient compression of ecg signals using deep convolutional autoencoders," *Cognitive Systems Research*, vol. 52, pp. 198–211, 2018.

[12] Ö. Yıldırım, P. Pławiak, R. San Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ecg signals," *Computers in biology and medicine*, vol. 102, pp. 411–420, 2018.

[13] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel eeg using convolutional neural networks," *arXiv preprint arXiv:1610.01683*, 2016.