

Received November 26, 2020, accepted December 8, 2020, date of publication December 14, 2020, date of current version December 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044497

Deep Learning-Based Vibration Signal Personnel Positioning System

YANG YU^{1,2}, MARIAN WALTEREIT¹, VIKTOR MATKOVIC¹,
WEIYAN HOU², AND TORBEN WEIS¹

¹Distributed Systems Group, University of Duisburg-Essen, 47057 Duisburg, Germany

²School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Yang Yu (yang.yu@uni-due.de)

We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen. We acknowledge and thank Evonik Digital for this research work's financial support.

ABSTRACT In this work, we present a person localization system based on ground vibration caused by walking persons. The system is designed for production plants and large buildings to track the movement of workers. Position and movement in these settings are especially safety-relevant in emergencies. Our approach is privacy-preserving, because it requires neither video nor sound. Instead, piezo sensors on the floor measure vibrations, which are analyzed with machine learning to derive a person's position from the vibration signals. This way, our system can determine where a person is moving, but it is not straightforward to attach names to the detected persons. Due to the anisotropic characteristic of the ground vibration wave, classical analysis methods are not applicable. We show that a deep learning-based approach is feasible. Our experiments show that we can determine the position with an average F1 score of 0.95.

INDEX TERMS Vibration signal, localization, pattern recognition, deep learning, privacy protection, robustness, piezo sensor.

I. INTRODUCTION

Localization technology has attracted attention from industry and academics [1]. It is the cornerstone for a large number of person localization services [2]–[6]. Person localization data can be used in scenarios like post-disaster rescue, situational awareness, security defense, and intrusion detection. At the same time, people also hope that the data applied while obtaining location information does not violate their privacy.

However, there are many studies on surveillance and situation awareness, but rarely the issue of privacy is considered. In some countries or regions, such as Europe, privacy is extremely important due to legal restrictions. At the same time, people's need for intelligent situational awareness is growing. The motivation of our research is to balance these two contradictory points in order to enable intelligent situational awareness while preserving privacy.

In this paper, we introduce an approach that is based on the analysis of structural vibration waves caused by footsteps, measured with cheap piezo sensors on the floor. The main research challenge is to determine the position of individual footsteps on the floor. This information can be used to deter-

mine a higher-level context, such as the walking direction or the walking path of a person.

The advantage of our approach is that we do not rely on persons wearing any senders or receivers. Meanwhile, our approach uses very cheap piezo sensors rather than expensive geophones [7], which makes a large-scale commercial deployment possible. Furthermore, our system is less privacy-invasive than cameras, and it does not detect who is walking. Especially for rescue scenarios in large production plants, it is imperative to know the location of people and their direction of movement. Besides, our system works in smoked areas where cameras do not work anymore.

Our approach is based on the observation that the steps of a walking person cause a mechanical vibration of the ground. At this time, the position of the person is the position of the vibration source. Nevertheless, the time difference and amplitude of the recorded vibration signals are related to the vibration source position. Thus, we assumed that the vibration source could be determined by an analysis of the vibration signals' characteristics.

However, due to the non-uniform nature of the ground structure materials and factors caused by the ground construction process, the ground substances' property is anisotropic. Also, ground waves show multipath effects caused by various obstacles like walls, etc. in the indoor space. Furthermore,

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

the recorded vibration signal has the characteristics of low signal-to-noise ratio (SNR) [7], which leads to an inevitable large measurement error in signal processing. Due to these characteristics, person localization based on mechanical ground waves is challenging. The traditional time differences of arrival (TDOA), time-of-arrival (TOA), angle-of-arrival (AOA), Doppler shift frequency-difference-of-arrival (FDOA) as well as received signal strength (RSS) based methods are widely used in wireless radio-based localization approaches with sensor arrays [7]–[10]. Due to mechanical wave propagation characteristics on the indoor floor, these methods cannot achieve good performance in our scenario.

Meanwhile, there are methods based on time difference distribution models [11] to recognize the ball impact localization on table tennis rackets using piezo-electric sensors. In this research, the sensors are distributed relatively close to each other, and the vibration is spreading on wood. The starting point of the wave signal can be identified relatively clearly with low measurement errors. In contrast, as mentioned above, in our scenario the quality of the signal propagation on the concrete floor is much worse than on the wood floor, and the measurement error of the waveform arrival time is immense. So we need to find a different method to overcome this problem.

Neural networks are capable of learning complex nonlinear relationships [12]. Convolutional neural networks show good performance in feature extraction from data [13]. Meanwhile, residual connections make deep networks easier to train without increasing the complexity of the network [14]. There is existing research that uses deep residual convolutional networks to analyze time serial data with good performance [15]. Therefore, in this paper, we show that the analysis of ground vibration signals is feasible with deep residual neural networks.

Our approach's starting point is a customized positioning application in a specific area of a chemical plant with the feature of privacy protection. In the future, our approach can be used on vibration-based positioning applications in general scenarios. The cost of one set of the sensor matrix is lower than 2 Euro. With custom circuits and mass production, the hardware cost involved in our approach can be greatly reduced.

Meanwhile, our sensor deployment requires one unit every $9m^2$. The function of each unit is relatively independent. When there is a larger area, more units can be deployed. When pedestrians walk across different units, the entire integrated system can give global positioning information to pedestrians.

Our approach's sensor can be embedded into floor tiles to unify the distribution characteristics of training data and data from actual application scene, thereby improving the feasibility of our approach in practical applications. Meanwhile, there is ongoing research about domain adaptation [16], [17]. The domain adaptation technique makes the deep learning-based approach work on the target domain data,

which has different distribution characteristics than the training data. Furthermore, the dimensionality reduction technology can further reduce the dimensionality of the input data, thereby reducing the computing cost [18], [19]. The introduction of domain adaptation methods and dimensionality reduction methods to our system will be further studied in future work.

The contribution of this paper is as follows. We present a novel fine-grained deep learning [20] based localization technique using cheap vibration sensors, which is a passive detection that does not need the pedestrian to carry a device. The proposed technique can well tolerate noise in the indoor environment. We evaluated and validated the accuracy and robustness of the approach with experiments in a real scenario.

II. RELATED WORK

Localization systems for outdoor scenarios are often realized with GPS. However, GPS does not work indoor [21]. Pedestrian Dead Reckoning-based methods use inertial sensors, and they cannot avoid the error accumulation problem [22]–[24].

The radiofrequency or WiFi fingerprint-based positioning methods [25], [26] calculate the location with the signal from the transmitter carried by the pedestrian. All the above methods require each person to carry a device, not the passive localization method, without carrying equipment that our target scenario pursues.

Video-based localization techniques do not support privacy protection for people in the given scenario. In extreme environments, such as explosions, fires, etc., where visibility is reduced, the location functionality of video-based techniques and visible light communication-based positioning techniques [27] will no longer be available.

There are floor vibration-based indoor localization methods which introduced TDOA [7], [28]. The TDOA method assumed that the speed of wave propagation is constant. Due to the ground's anisotropy, the ground substances do not satisfy this condition when a mechanical wave is conducted. In other words, the propagation velocity on the indoor floor is not constant. So if the TDOA method is used, the constant speed assumption should be eliminated. However, even if the speed is constant, the simple TDOA method cannot obtain satisfactory positioning performance. To improve the positioning accuracy, researchers introduced an ATDOA [29] method that combines the TDOA method and the AOA method by the weighted average method. Although this ATDOA method reduces the positioning error to a certain extent, the authors had to use a triaxial seismic sensor, which increases deployment costs, creating obstacles to commercial deployment from cost considerations. Furthermore, this method's localization performance on the concrete floor is far from satisfactory compared to the wooden floor.

Meanwhile, because the TDOA algorithm's essence is to find the coordinate of the intersection of more than three hyperbolae, this algorithm has the well known “no solu-

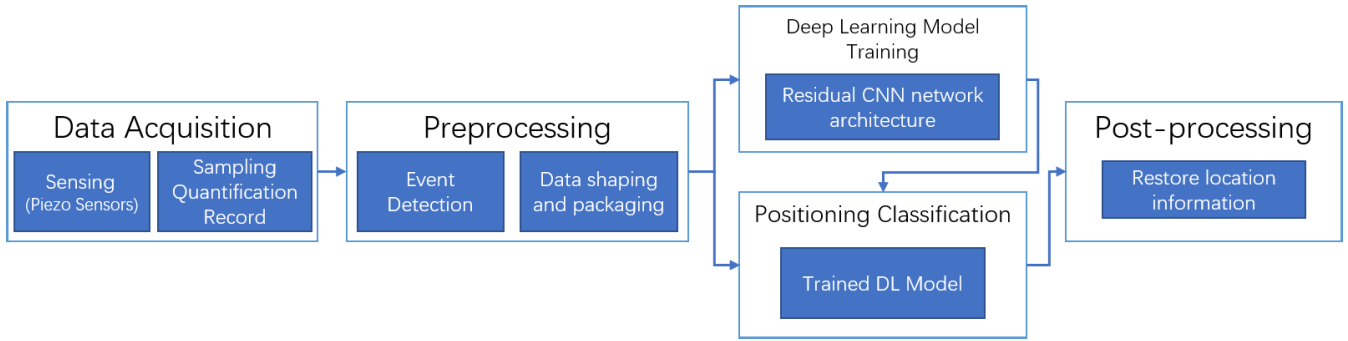


FIGURE 1. The system includes data acquisition, data preprocessing, deep learning model training, positioning classification, and post-processing.

tion” and “multiple solutions” problems. This has led to restrictions on where sensors can be deployed. To avoid these problems, all the above TDOA-based research deploys the sensors on the four vertices of the rectangle, and the area inside the rectangle is the positionable area. However, this is very unfavorable for large-scale deployment, especially in large areas, because the signal decays and thus the size of the supported area is strictly limited.

As a comparison, there is research using the cheap piezo sensor to conduct indoor localization [30]. Nevertheless, this method requires reference positioning objects with a known position, which can hardly be recognized as a real positioning system. The positioning accuracy depends on the density of the reference position objects. This method’s limitations are significant, and it can hardly be used in the most indoor localization scenario.

In contrast to the existing localization systems, our approach is a passive localization method that does not require persons to carry any device. This feature is an added benefit in chemical plants where, by default, mobile electronic devices are not allowed unless they are certified not to cause explosions in combination with potentially explosive gas. Furthermore, our approach does not require a clear line of sight and therefore works in heavily smoked rooms. Our approach uses cheap piezo sensors, which has a cost advantage, and there is no restriction on the relative position of the sensor position and the located area like the TDOA-based method. This makes it more feasible in large-scale deployments, especially in large areas. Our method still has good accuracy and robustness on the concrete floor.

III. METHODOLOGY AND SYSTEM

As shown in Fig. 1, the proposed system is a deep learning-based indoor and outdoor passive pedestrian localization system. In this section, the procedure is discussed.

A. PROBLEM FORMULATION AND ANALYSIS

In our system, four piezo sensors on the floor can detect footsteps in a square of 9m². In our experiments, we want to determine the position of individual footsteps in this square. This three meters by three meters square is divided into nine zones, each one meter by one meter in size (see Fig. 2).

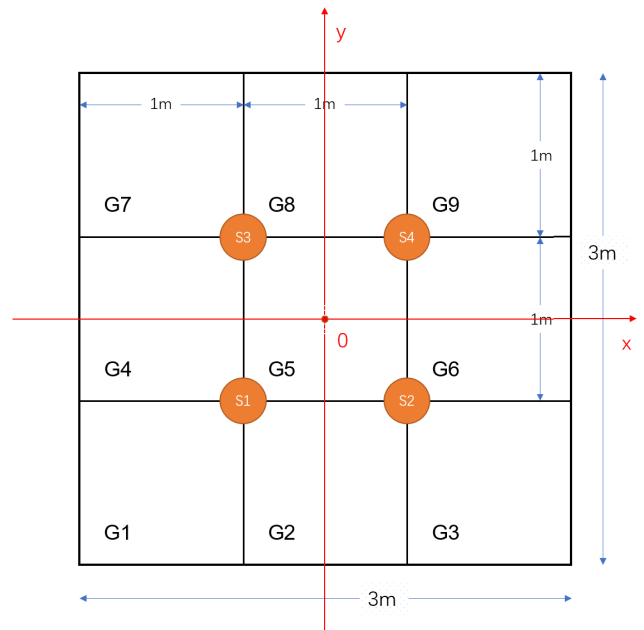


FIGURE 2. System model. A space of 3 meters by 3 meters is divided into nine grids named G1 to G9. Four sensors, S1 to S4, are deployed in the position points, as shown in this figure. When the sensor-matrix detects a vibration source in the observed zone, the zone name should be output.

For each footstep, we want to determine the zone in which the footstep occurred. These zones are labeled as G1 to G9. As the signal decays quickly in thick concrete floors, we can test by amplitude which arrangement of four sensors is closest to the pedestrian. Thus, we investigate how to detect a person inside these nine zones and the approach scales to arbitrarily larger floors nevertheless.

Concerning our machine learning approach, each zone is represented by a class. Thus, we designed a classifier that can determine the zone based on four channels of vibration signals. The sensors producing these signals are labeled as S1 to S4. In order to cover larger areas, the same arrangement of zones and sensors can be repeated in both directions.

The location information of the vibration source has a relationship with the maximum amplitude of the four channels and the relative peak position in time. The waveform of the signal of the channel nearest to the vibration source will start to rise first. Meanwhile, because the amplitude of the wave decays with the distance in the direction of the

wave propagation, the waveform of the sensor nearest to the vibration source will have the maximum amplitude among all four signals. To achieve localization, we used a deep residual convolutional neural network to learn the nonlinear relationship between the point location and the signal of the four sensors.

B. METHODOLOGY DESCRIPTION

In this research, the aim is to detect persons walking across the floors of large scale buildings, for example in a chemical plant.

Our methodology is to deploy an array of four sensors every 9 m^2 . When a person is stepping inside such a 9 m^2 area the four sensors will sense the vibration signal and we can such determine the 9 m^2 area in which a footstep occurred. Using an AI-approach we can then determine the position of each footstep inside the 9 m^2 area to get a precision of 1m in both directions. This way we can sense and track footsteps across the entire factory store.

An advantage of our approach is that we can concentrate on solving the positioning problem for the 9 m^2 area and our AI has to process 4 sensor signals only. However, our system can then be scaled out to larger areas.

Our methodology is privacy-friendly in contrast to video surveillance. Furthermore, our approach works in dark or heavily smoked areas, too. Thus, our approach is ideal to track the movements of persons especially in the case of an emergency.

C. DATA ACQUISITION

Fig. 4 shows our data acquisition devices setup. We used four EPZ-27MS44W piezoelectric sensors to detect the vibration signal. This sensor can detect a frequency band ranging from 0 Hz to 4400 Hz. As shown in Fig. 2, we established a plane rectangular coordinate system on the ground. The four sensors S1, S2, S3, S4 were deployed according to Fig.2. and the corresponding signals recorded are labeled as $s_1(t)$, $s_2(t)$, $s_3(t)$, $s_4(t)$.

A square steel plate of 10 cm by 10 cm is pressed onto each sensor to guarantee that ground waves cause pressure on the sensor from below due to the steel plate's inertia above. In the user's actual application environment, the sensor can be embedded in the floor tiles to ensure that the sensing device will not affect normal walking. The signal was amplified, sampled, quantized, and recorded with a R&S-RTB2000 oscilloscope. The sampling rate is 23.8 kHz, which is more than double the sensor's maximum bandwidth. Hence, a higher sampling rate would not improve the results anymore. Each sampling point's possible maximum value is manually set to 0.1 V, and waveforms higher than 0.1 V are cut off.

The same person with the same shoes generated the ground vibration by walking on points evenly distributed in each zone. We repeated this procedure for all nine zones. This way, we can easily label the data since we know how samples and zones are related. The recorded data includes four columns,

which refer to the S1, S2, S3, S4 signals, and the values in each row are the sampled data values of which the sampling period is 0.042 ms. Our approach is time-based, and therefore all signals must be synchronized. We achieved this by sampling all four signals with one oscilloscope.

D. PREPROCESSING

It is necessary to detect the starting point of the effective signal when there is vibration. After analyzing the existing threshold-based method [31], [32], a customized shift window with a grouped frame threshold-based method is used in this research. The sampling points are grouped into windows of 32 points, which means that each window covers 1.344 ms. When the maximum of the absolute value of the median of any channel in a group becomes bigger than the threshold, the first sampling point in the group before this group will be considered as the starting point of the signal segment. The group with the starting point is named the activated group. We chose the previous group of the activated group to guarantee all the relevant signal sampling points are taken into consideration by the system. The trigger threshold is $V_t = 25\text{mv}$. Starting from the activated group, four consecutive groups are grouped as a big group, which serves as a valid data frame. Furthermore, a valid data frame, totally with 128 sampling points, is served as one sample to the neural network.

One sample includes the waveform of 5.376 ms, which contains all the valid information for a one-time localization task for one shock. Considering that the step frequency time of elite professional sprinters is always longer than 200 ms [33], and the fact that the signal collected by the piezoelectric sensor will decay after 80ms, the data frame contains all the necessary information for one-time foot shock without any signal caused by adjacent steps. Meanwhile, after one data frame is detected, the next frame will be detected after 224 groups of about 300ms. There are no coincident points between the adjacent data frames.

After we got the data samples for an event, the valid data frame should be normalized before training the deep neural network. As the maximum value of the signal envelope is 0.1V, we scaled all the points from the four channels by dividing the values by 0.1 to guarantee all the values are between -1 and 1 . After scaling, we got the training samples. We do not conduct the denoising in the preprocessing procedure and still got good results, as shown in the experiment and evaluation section. Deep neural networks can, to some extent, tolerate noise. This result is consistent with the literature [15].

E. CLASSIFIER DESIGN AND DEEP LEARNING MODEL TRAINING

Considering that existing works [15], [34], [35] suggest that deep learning is more suitable to analyze time-series of sensor data, our approach is based on CNNs as well. The works [34] compared CNN, LSTM, perceptron, and random forest and conclude that CNN provides the best results. The existing works [35] use the same sensors on a pattern recognition task and compare a deep learning approach with random forest

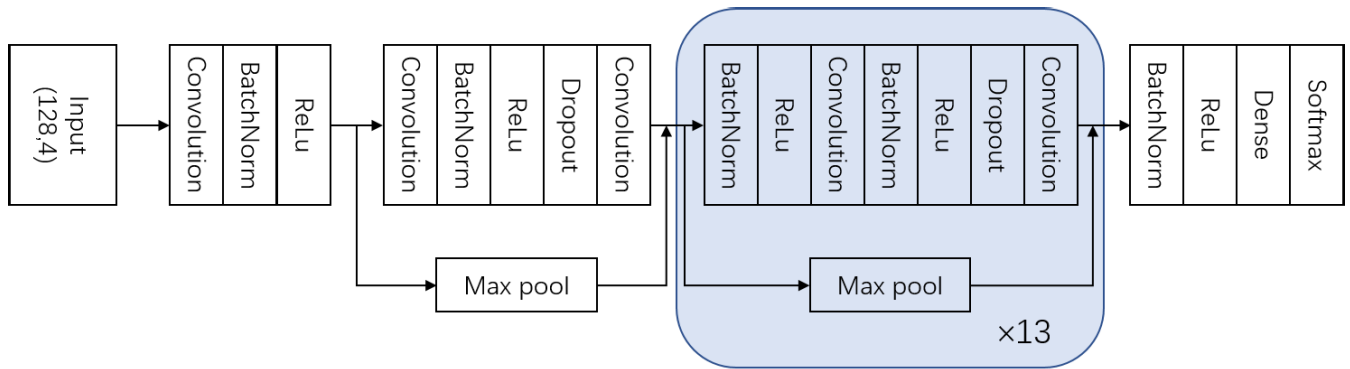


FIGURE 3. Deep neural network architecture. Our DNN model comprises 29 convolution layers, followed by a linear output layer that ended with a softmax. The network accepts packaged vibration signal as input (sampled at 23800Hz, or 23800 samples per second). The output of the DNN is a probability vector, which is the prediction of one of 9 possible location-related classes every consequent 128 points. The residual structure is utilized to gain accuracy from the increasing depth of the neural network.

classifiers in the experiment. The results also show that the deep learning approach works better than the classical classifier random forest. Also, deep residual networks show good performance in the fixed-length series pattern recognition task [15]. Therefore, we excluded classic classifiers and used a CNN-based deep learning approach instead.

We collected a data set, marked with “Data Part 1”, and split it into a training set for training model and validation set for model hyperparameters optimization. We finally determined the best model architecture and hyperparameters, as shown in this paper.

Fig. 3 shows the architecture of the deep neural network designed in this experiment, which presents a good performance in the classification task, which we discuss in the next section.

Our deep neural network consists of 29-dimensional convolution layers followed by a linear output layer into a softmax layer. The network accepts packaged effective vibration signal data as input and outputs a prediction of one out of 9 possible location-related classes every one input sample.

The first convolution layer in the deep neural network contains 48 filters, and the number of filters used in the convolution layer increases according to the increasing of the order of the layer. From the 2nd convolution layer, the filter number doubles every eight convolution layers, and finally, in the last convolution layer, 384 filters have been used. Due to the convolution operations and the max pooling operations, the data frame’s length becomes the half after the 3rd convolution layer. From the 4th convolution layer, the data frame length becomes half of the output length of the previous layer. Finally, the size of the data frame becomes 1 row and 384 columns. The dropout rate was set to 0.5, learning rate 0.0005, and batch size 32. The 1D convolution kernel size is 16 [36]. The initial stopping criteria are introduced to restrain the over-fitting issues, which are based on performance validation. After eight consequent training epochs have been conducted without improvement in the performance validation result, the training procedure has been stopped.

To objectively reflect the performance of our approach, four-fold cross-validation has been conducted on totally new

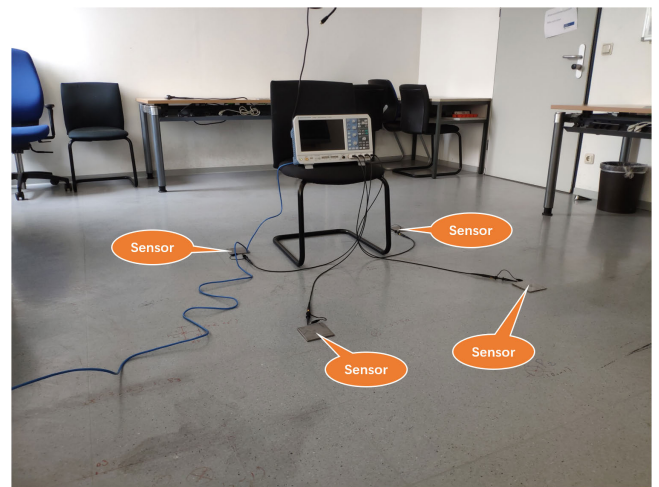


FIGURE 4. Experiment setup.

data, marked as data part 2, as shown in Fig. 5(a). We discuss this in detail in the next section.

IV. EXPERIMENT AND EVALUATION

In this section, firstly, we discuss how we organized the data for the evaluation. Secondly, we show the location-related classification performance. The ROC figure and the AUC index are used to represent the classification performance. Then the error analysis was conducted, and the results are shown as in Table. 2. In the error analysis part, Euler distance and Manhattan distance are introduced. We calculate the error distance with each fold, and the average values for the four folds evaluation procedures were calculated, as shown in Table. 2.

A. DATA ORGANIZATION

To evaluate our approach, model architecture, and hyperparameters, we conducted four experiments to get four data sets, marked as “Data Part 2”, as shown in Fig. 5(a). The “Data part 2”, including data set A, data set B, data set C, and data set D, a total of 30576 samples, are from four independent experiments A, B, C, and D. This “Data Part 2” has never been used in the model hyperparameter optimization procedure.

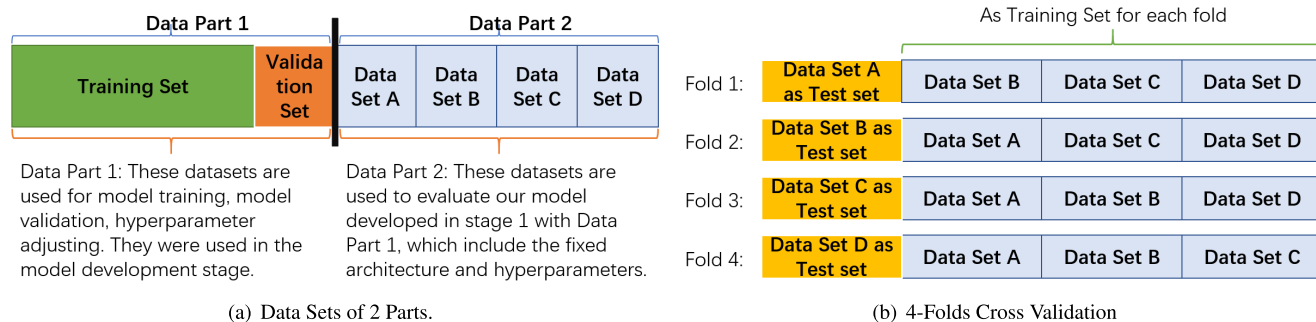


FIGURE 5. As shown in Fig. 5, we totally collected two parts of data sets. For data set part 1, we randomly shuffle the data samples and split them into a training set and a validation set according to a rate of 4:1. We developed our model with the data set part 1, finally fixing the architecture and the hyperparameters. We evaluate our approach, including the developed model architecture and hyperparameters with data set part 2. We conducted four independent experiments: experiment A, experiment B, experiment C, experiment D, and collected the data set A, data set B, data set C, data set D individually. We conducted a “4-folds cross-validation” evaluation, as shown in Fig. 5(b). The evaluation results are shown in Chapter IV.

In the evaluation, all data samples of “Data Part 2” were used for four-fold cross-validation. The typical k-folds cross-validation method [37], [38] splits the entire data set into k mutually exclusive subsets as folds with approximately equal size. Every time, one set of data is used as the test set and the rest as the training set. Our evaluation procedure made minor adjustments to the typical cross-validation method and implemented it as a variant. We independently collected four sets of data in four times. We did not merge all the 4 data sets or randomly split it into folds. Each test set in each fold can be considered as online data, and the training set as offline training data. In this way, it is even more challenging for our system to make good classification results, because no data point of the measurement session has been used for training. In comparison, imagine that we shuffled all data of “Data Part 2” and imagine we performed four-fold cross-validation on this shuffled data (which is the default procedure for cross-validation). In this case, each training set contains data points from all four measurement sessions. Thus, CNN is trained with data from all measurement sessions and would have an easier time to classify the test set, which contains data from all measurement sessions. In our paper, however, CNN is trained with three measurement sessions and must classify data from another measurement session that it has not seen during training. Thus, our cross-validation is even more rigorous than the default approach, and the evaluation experiments can better simulate the system’s situation in actual use, that it can objectively show that our approach works.

As shown in Fig. 5(b), for evaluating purpose, we trained the model four times and then tested each model instance. The “Data Part 2” has never been used in the model creation procedure. In the “four-fold evaluation”, all the data samples have been used for training and testing. The values of each evaluation metric were summed up, and the average values of them were calculated. Precision, recall, F1-score were used as evaluation metrics, as shown in Table. 1.

B. CLASSIFICATION PERFORMANCE ANALYSIS

Table. 1, Fig. 6 and Fig. 7 show the classification performance presented by our DNN model. Precision, recall, and

TABLE 1. Classification performance of the DNN.

	Precision	Recall	F1-score
G1	0.9465	0.9479	0.9471
G2	0.9066	0.9362	0.9209
G3	0.9202	0.9211	0.9206
G4	0.9684	0.9616	0.9650
G5	0.9609	0.9628	0.9617
G6	0.9457	0.9136	0.9291
G7	0.9741	0.9696	0.9718
G8	0.9644	0.9595	0.9618
G9	0.9371	0.9508	0.9437
Micro Average	0.9472	0.9472	0.9472
Macro Average	0.9471	0.9470	0.9469

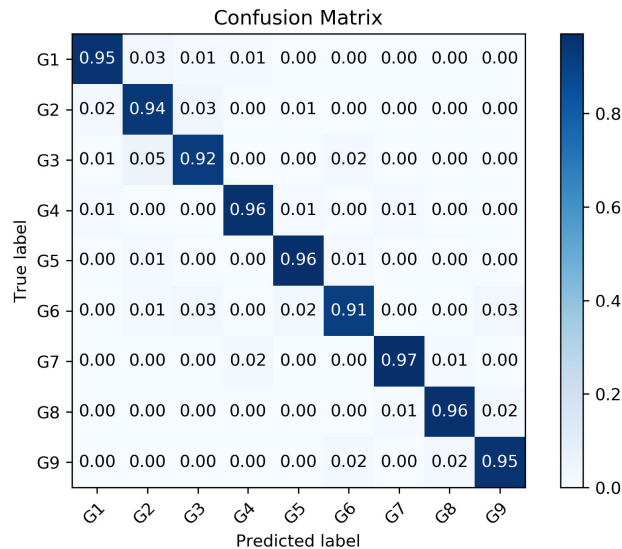
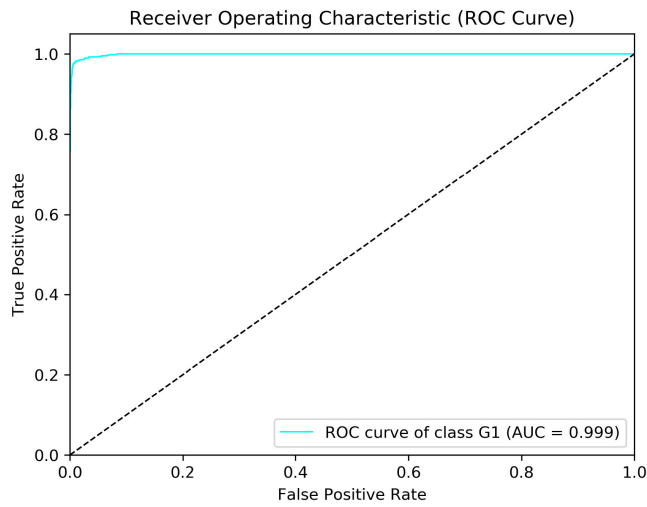
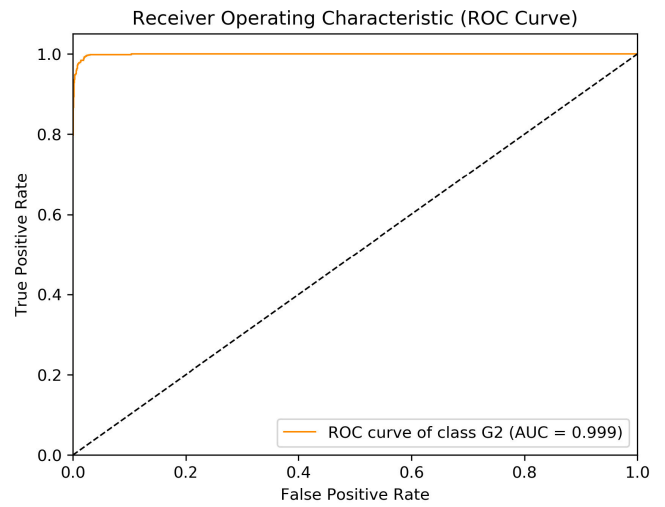


FIGURE 6. Confusion matrix. Fig. 6 is the confusion matrix for the location-related class prediction of the classifier versus the sample labels. As part of the four-fold cross-validation, this confusion matrix’s values are the average of the corresponding values from the four tests.

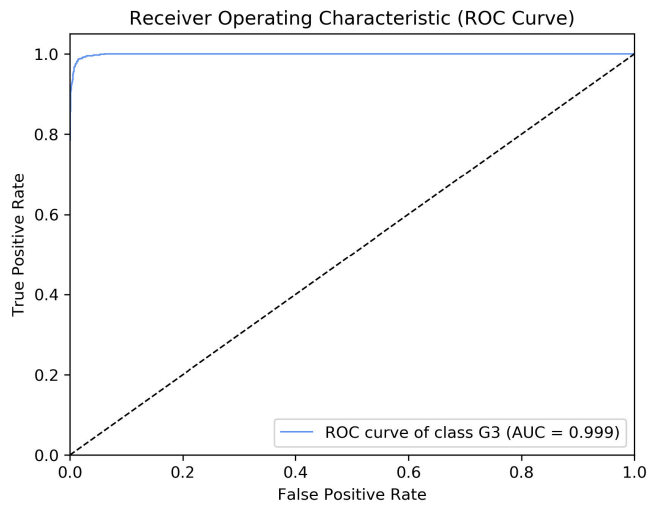
F1-score [40] are used as metrics for the evaluation and analysis approach. Intuitively, the precision represents the classifier’s ability not to label as positive a sample that is negative. The recall shows the ability of the classifier to find all the positive samples. The F1 score can be interpreted as a weighted harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and the worst score at 0.



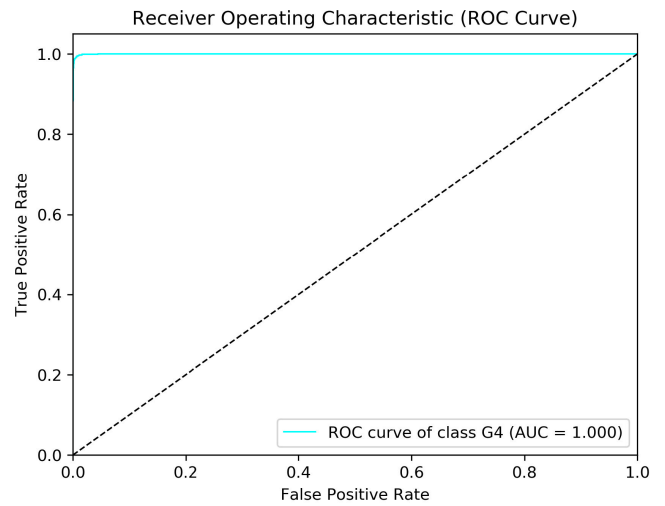
(a) ROC G1.



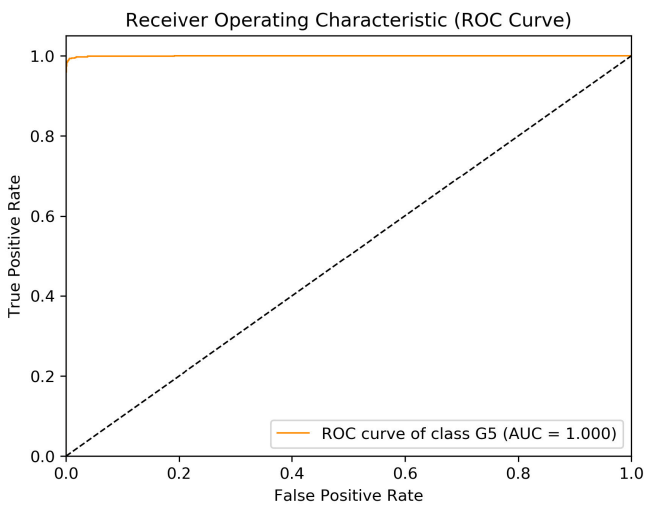
(b) ROC G2.



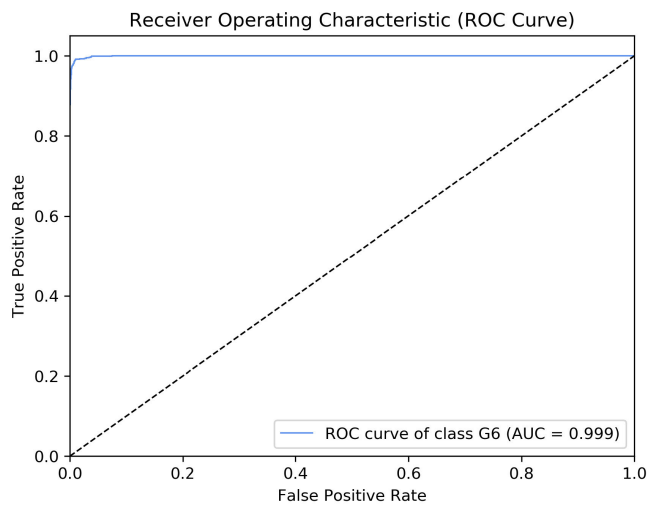
(c) ROC G3.



(d) ROC G4.



(e) ROC G5.



(f) ROC G6.

FIGURE 7. Fig. 7(a) to Fig. 7(k) are the ROC curves for deep neural network predictions on nine location-related classes. The macro-average ROC curve and micro-average ROC curve [39] reflect the multi-class classifier performance.

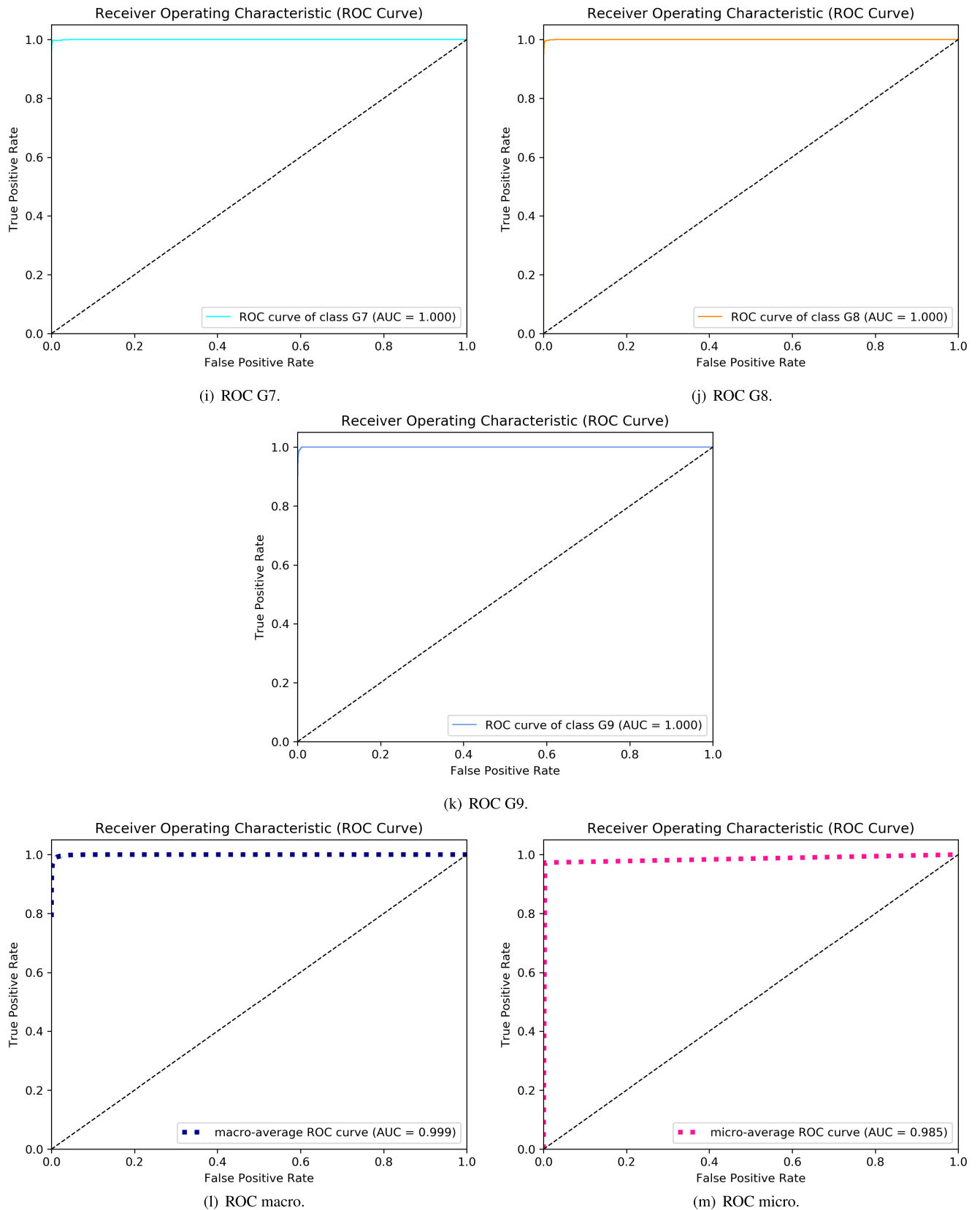


FIGURE 7. (Continued) Fig. 7(a) to Fig. 7(k) are the ROC curves for deep neural network predictions on nine location-related classes. The macro-average ROC curve and micro-average ROC curve [39] reflect the multi-class classifier performance.

The F1 score in our evaluation considered recall and precision as equally important.

$$\text{Micro - average Precision} = \frac{\sum_c TP_c}{\sum_c TP_c + \sum_c FP_c} \quad (1)$$

$$\text{Micro - average Recall} = \frac{\sum_c TP_c}{\sum_c TP_c + \sum_c FN_c} \quad (2)$$

The metrics used in Table. 1 include the macro average of the metrics and the micro average of the metrics. Table. 1 shows the metrics precision, recall, and F1-score. The macro average precision and the macro average recall are the arithmetic average of each class's precision and recall. The macro average F1-score is the harmonic mean of macro average precision and macro average recall. Micro average precision and micro average recall are as described in the formulas (1) and (2). In formula (1) and formula (2), TP means true positive; FP means false positive; FN means false negative; c is the class label. The micro average F1-score is the harmonic mean of micro-average of precision and micro-average of recall, which means averaging the unweighted mean per label [41], [42]. The micro and macro average performance metrics reflect the system's average level of performance concerning each class. In our evaluation setting, the number of instances in each class is similar. So the micro average figures are similar to the macro average figures.

Fig. 7 shows the ROC curves of the classifier. A ROC curve is a visual indicator of the trade-off between true-positive and false-positive cases. The area under the ROC curve (AUC) will always be between 0 and 1. According to the ROC curves in Fig. 7, our classifier shows outstanding performance.

Fig. 6 shows that most of the evaluation samples were correctly classified. Considering the practical walking people location scenario, the positioning-related grid can be predicted correctly in most cases. 5% of G3 samples have been classified as G2, which is the highest classification error. Meanwhile, 3% G1 were predicted as G2; 3% G2 samples were predicted as G3; 3% G6 samples are predicted as G3; 3% G3 samples were predicted as G6; 3% G6 samples were predicted as G9. However, all the above mentioned misclassified samples are location-adjacent. As G8 is next to G5, G3 is next to G2. That is to say, although some samples are wrongly classified, in most cases, the wrongly classified results are still with practical meaning. The remaining misclassified samples are less than 2% of the total sample number of its class.

C. ERROR DISTANCE ANALYSIS

From the perspective of walking people positioning, the error distance analysis has been conducted, as shown in Table. 2. For error distance analysis, Euler distance and Manhattan distance [43] were counted.

In Table. 2, the error Euler distance and the error Manhattan distance of a sample, which wrongly classified G_x as G_y , is the Euler distance and Manhattan distance of the geometric center point of G_x to the Euler distance and the Manhattan

TABLE 2. Error distance description of misclassified samples.

	Error Euler Distance	Error Manhattan Distance
Mean	0.059	0.062
Standard Deviation	0.255	0.280
Min	0	0
Median	0	0
Max	2.532	3.500

distance of the geometric center point of G_y , respectively. The unit is meter. The average error Euler distance and the average error Manhattan distance are 0.0560 0.0596, respectively, and the medians of that respectively are 0 and 0. The error distance evaluation results show that the proposed system can implement localization function with good performance and low error.

V. CONCLUSION

We presented a novel approach to determine the walking persons' position by analyzing ground vibrations caused by individual footsteps. Due to the anisotropic nature of the ground waves, we used a machine learning approach to determine a footstep position. We divided a $9m^2$ area into nine squares of $1m^2$ size. The learned classifier can assign a ground vibration signal to one of these nine areas. The mean Euler error of our approach, as defined in our paper, is 0.059m.

Our approach can identify the position of an individual step. Therefore, for a walking person, we can identify their position and the direction of their movement. This is especially useful in production plants, where safety demands that firefighters know where people are and in which direction they are running. In contrast to video surveillance, our approach works in smoked areas. Furthermore, our approach is less privacy-invasive than cameras. Since many companies are not allowed to install always-on cameras for privacy reasons, our approach can deliver more safety without invading workers' privacy.

ACKNOWLEDGMENT

The author would like to thank Evonik Digital to access their production plant for ground wave measurements.

REFERENCES

- [1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 3rd Quart., 2019.
- [2] A. Serra, D. Carboni, and V. Marotto, "Indoor pedestrian navigation system using a modern smartphone," in *Proc. 12th Int. Conf. Hum. Comput. Interact. With Mobile Devices Services - MobileHCI*, 2010, pp. 397–398.
- [3] A. E. Fano, "Shopper's eye: Using location-based filtering for a shopping agent in the physical world," in *Proc. 2nd Int. Conf. Auton. Agents*. New York, NY, USA: Association for Computing Machinery, 1998.
- [4] K.-H. Park, Z. Bien, J.-J. Lee, B. K. Kim, J.-T. Lim, J.-O. Kim, H. Lee, D. H. Stefanov, D.-J. Kim, J.-W. Jung, J.-H. Do, K.-H. Seo, C. H. Kim, W.-G. Song, and W.-J. Lee, "Robotic smart house to assist people with movement disabilities," *Auto. Robots*, vol. 22, no. 2, pp. 183–198, Jan. 2007.
- [5] Y. Tajika, T. Saito, K. Teramoto, N. Oosaka, and M. Isshiki, "Networked home appliance system using Bluetooth technology integrating appliance control/monitoring with Internet service," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 1043–1048, Nov. 2003.

- [6] J. Scott, A. J. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar, "PreHeat: Controlling home heating using occupancy prediction," in *Proc. 13th Int. Conf. Ubiquitous Comput. - UbiComp*, 2011, pp. 281–290.
- [7] M. Mirshekari et al., "Characterizing wave propagation to improve indoor step-level person localization using floor vibration," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems* (International Society for Optics and Photonics), vol. 9803. Las Vegas, NV, USA: SPIE, 2016. Accessed: Jan. 6, 2020. [Online]. Available: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9803/980305/Characterizing-wave-propagation-to-improve-indoor-step-level-person-localization/10.1117/12.2222136.short>
- [8] X. Tang, M.-C. Huang, and S. Mandal, "An 'Internet of ears' for crowd-aware smart buildings based on sparse sensor networks," in *Proc. IEEE Sensors*, Oct. 2017, pp. 1–3.
- [9] W. Chen, M. Guan, L. Wang, R. Ruby, and K. Wu, "FLoc: Device-free passive indoor localization in complex environments," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [10] J. D. Poston, R. M. Buehrer, and V. Tech, "Vibration sensing in smart buildings," in *Proc. Int. Conf. Indoor Position. Indoor Navig. (IPIN)*, 2015, p. 4.
- [11] P. Blank, T. Kautz, and B. M. Eskofier, "Ball impact localization on table tennis rackets using piezo-electric sensors," in *Proc. ACM Int. Symp. Wearable Comput.*, New York, NY, USA: Association for Computing Machinery, 2016, pp. 72–79, doi: [10.1145/2971763.2971778](https://doi.org/10.1145/2971763.2971778).
- [12] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, Jan. 1989.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [15] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, p. 65, 2019.
- [16] Z. Wang, B. Du, and Y. Guo, "Domain adaptation with neural embedding matching," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–11, Sep. 2019.
- [17] L. Song, C. Wang, L. Zhang, B. Du, Q. Zhang, C. Huang, and X. Wang, "Unsupervised domain adaptive re-identification: Theory and practice," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107173.
- [18] F. Luo, H. Huang, Y. Duan, J. Liu, and Y. Liao, "Local geometric structure feature for dimensionality reduction of hyperspectral imagery," *Remote Sens.*, vol. 9, no. 8, p. 790, Aug. 2017. [Online]. Available: <https://www.mdpi.com/2072-4292/9/8/790>
- [19] G. Shi, H. Huang, and L. Wang, "Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1425–1429, Aug. 2020.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [21] F. Capezio, A. Sgorbissa, and R. Zaccaria, "GPS-based localization of a surveillance UGV in outdoor areas," in *Proc. 5th Int. Workshop Robot Motion Control RoMoCo*, Jun. 2005, pp. 157–162.
- [22] W. Kang and Y. Han, "SmartPDR: Smartphone-based pedestrian dead reckoning for indoor localization," *IEEE Sensors J.*, vol. 15, no. 5, pp. 2906–2916, May 2015.
- [23] C. Fischer, K. Muthukrishnan, M. Hazas, and H. Gellersen, "Ultrasound-aided pedestrian dead reckoning for indoor navigation," in *Proc. 1st ACM Int. Workshop Mobile Entity Localization Tracking GPS-Less Environments MELT*, 2008, pp. 31–36.
- [24] S. Beauregard and H. Haas, "Pedestrian dead reckoning: A basis for personal positioning," in *Proc. 3rd Workshop Positioning, Navigat. Commun.*, 2006, pp. 27–35.
- [25] B. Wang, X. Liu, B. Yu, R. Jia, and X. Gan, "An improved WiFi positioning method based on fingerprint clustering and signal weighted Euclidean distance," *Sensors*, vol. 19, no. 10, p. 2300, May 2019.
- [26] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on Wi-Fi: An overview," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 5, p. 135, Apr. 2017.
- [27] H.-S. Kim, D.-R. Kim, S.-H. Yang, Y.-H. Son, and S.-K. Han, "An indoor visible light communication positioning system using a RF carrier allocation technique," *J. Lightw. Technol.*, vol. 31, no. 1, pp. 134–144, Jan. 2013.
- [28] J. Clemente, W. Song, M. Valero, F. Li, and X. Liy, "Indoor person identification and fall detection through non-intrusive floor seismic sensing," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Jun. 2019, pp. 417–424.
- [29] F. Li, J. Clemente, M. Valero, Z. Tse, S. Li, and W. Song, "Smart home monitoring system via footstep-induced vibrations," *IEEE J. Mag.*, vol. 14, no. 3, pp. 3383–3389, Sep. 2020, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8840889/versions>
- [30] Y. Kashimoto, M. Fujimoto, H. Suwa, Y. Arakawa, and K. Yasumoto, "Floor vibration type estimation with piezo sensor toward indoor positioning system," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2016, pp. 1–6.
- [31] S. Pan, A. Bonde, J. Jing, L. Zhang, P. Zhang, and H. Y. Noh, "BOES: Building occupancy estimation system using sparse ambient vibration monitoring," *Proc. SPIE*, vol. 9061, Apr. 2014, Art. no. 90611O.
- [32] M. Lam, M. Mirshekari, S. Pan, P. Zhang, and H. Y. Noh, "Robust occupant detection through step-induced floor vibration by incorporating structural characteristics," in *Dynamics of Coupled Structures*, vol. 4. New York, NY, USA: Springer, 2016, pp. 357–367.
- [33] R. F. Chapman, A. S. Laymon, D. P. Wilhite, J. M. Mckenzie, D. A. Tanner, and J. M. Stager, "Ground contact time as an indicator of metabolic cost in elite distance runners," *Med. Sci. Sports Exercise*, vol. 44, no. 5, pp. 917–925, May 2012.
- [34] V. Matkovic, M. Waltereit, P. Zdankin, M. Uphoff, and T. Weis, "Bike type identification using smartphone sensors," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2019, pp. 145–148.
- [35] Y. Yu and T. Weis, "A privacy-protecting indoor emergency monitoring system based on floor vibration," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2020, pp. 164–167.
- [36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, Nov. 2016.
- [37] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. London, U.K.: Pearson, 2009.
- [38] G. James et al., *An Introduction to Statistical Learning*, vol. 112. New York, NY, USA: Springer, 2013.
- [39] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [40] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. - ICML*, 2006, pp. 233–240.
- [41] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2004.
- [42] H. Narasimhan, W. Pan, P. Kar, P. Protopapas, and H. G. Ramaswamy, "Optimizing the multiclass F-Measure via biconcave programming," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 1101–1106.
- [43] S. Craw, "Manhattan distance," in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2017, pp. 790–791.



YANG YU received the B.Eng. degree in communication engineering from Zhengzhou University and the M.Sc. degree in embedded system engineering from the University of Duisburg-Essen, Duisburg, Germany, where he is currently pursuing the Ph.D. degree with the Distributed Systems Research Group. His research interests include AI-based context recognition, AI-based context reasoning, machine learning, and explainable AI.



MARIAN WALTEREIT received the M.Sc. degree in computer science from the University of Duisburg-Essen, Duisburg, Germany, where he is currently pursuing the Ph.D. degree with the Distributed Systems Research Group. His research interest includes privacy and digital forensics in automotive systems.



WEIYANG HOU received the bachelor's and master's degrees, in China, in 1986 and 1998, respectively, and the Ph.D. degree from Shanghai University, in a Sandwich Ph.D. Program between Germany and China, in 2004. He is currently a Professor with the School of Information Engineering, Zhengzhou University, China. His research interests include network control, time performance modeling of the industrial IoT, and infrastructure of wired/wireless communication integration.



VIKTOR MATKOVIC received the M.Sc. degree in computer science from the University of Duisburg-Essen, Duisburg, Germany, where he is currently pursuing the Ph.D. degree with the Distributed Systems Research Group. His research interest includes bike-aware pervasive applications, such as bike type-related navigation services.



TORBEN WEIS received the Ph.D. degree in computer science from the Technical University of Berlin, Berlin, Germany. He is currently a Professor with the University Duisburg-Essen, Duisburg, Germany, where he leads the Distributed Systems Research Group. His research interests include cyber-physical systems, cloud computing, and security and privacy in distributed systems.

...

DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

ub | universitäts
bibliothek

This text is made available via DuEPublico, the institutional repository of the University of Duisburg-Essen. This version may eventually differ from another version distributed by a commercial publisher.

DOI: 10.1109/ACCESS.2020.3044497

URN: urn:nbn:de:hbz:464-20210319-121949-6



This work may be used under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 License (CC BY-NC-ND 4.0).