

Building Parts Classification using Neural Network

Miroslav Opiela¹, Viktória Mária Štedlová¹, Šimon Horvát¹, Ľubomír Antoni¹ and Lucia Hajduková¹

¹ Faculty of Science, Institute of Computer Science, Pavol Jozef Šafárik University in Košice, 04001 Košice, Slovakia.

Abstract

Indoor positioning methods vary, and recent studies suggest that combining multiple sources of information through proper fusion can improve the accuracy of positioning. In this context, machine learning and neural network approaches have gained prominence. The objective of this paper is to propose a neural network-based method specifically trained on a particular building. Magnetic field sensors and camera images are chosen as inputs for the proposed solution. An LSTM network is trained to classify building parts based on magnetic field values, while a CNN network is utilized to identify areas based on camera images. The outputs from both networks are merged to provide concise information about the user's location within the building. However, the merge of these networks is yet to be implemented and remains open as future work. The LSTM network achieves accuracy ranging from 73% to 95% on individual floors, and further analysis reveals its ability to compensate for the weaknesses of the positioning system across multiple floors, even with lower accuracy. The CNN classification using the VGG16 model with pretrained weights achieves an accuracy of 98%, with 80% or 60% of the individual images correctly classified on selected paths. This approach demonstrates its applicability in enhancing indoor positioning systems that require either rough identification of building parts or precise determination of corridor sections.

Keywords

magnetic field, camera, LSTM, CNN, indoor positioning, classification

1. Introduction

Indoor positioning does not constitute a singular domain with a standardized use-case, device type, and solution. Rather, the field of positioning within buildings, where satellite signals are limited or unavailable, encompasses a multitude of scenarios. Numerous methods [1] have been developed to address the challenge of accurately determining user or device location. A considerable number of these approaches are designed to be applied in smartphone-based systems, targeting a broad user base consisting of pedestrians rather than a specific individual or robot, e.g., Wi-Fi, BLE (Bluetooth Low Energy), PDR (Pedestrian Dead Reckoning), etc.


Recent approaches (e.g., IPIN competition [2]) suggest that a proper integration of multiple sources of information leads to increased accuracy and robustness of the positioning system, especially using smartphones with low-cost sensors. Different localization methods may report specific weaknesses. The positioning method [3] considered in this research is composed of the

Proceedings of the Work-in-Progress Papers at the 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN-WiP 2023), September 25 - 28, 2023, Nuremberg, Germany

✉ miroslav.opiela@upjs.sk; viktoria.maria.stedlova@student.upjs.sk; simon.horvat@student.upjs.sk; lubomir.antoni@upjs.sk; lucia.hajdukova@student.upjs.sk

ORCID iD 0000-0001-8802-4442 (Miroslav Opiela); 0000-0002-7526-8146 (Ľubomír Antoni)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Bayesian filtering which incorporates detected steps, map model, and floor transition detection method. The structure of the building with its junctions improve the PDR-based method using smartphone sensors. However, walking along a single corridor in one direction introduces an increasing error caused mostly by inaccurate step length estimation. This error is reduced when switching floors or changing walk direction. Nevertheless, the approach proposed in this paper is aimed at the task to substitute the missing relevant inputs from map and PDR with an output from another sensor. Moreover, the preference in this research was to use a neural network trained on the specific building where the navigation and positioning is performed.

Smartphones are equipped with sensors that are integrated into the device, and the measurements captured by these sensors can be accessed through platform-specific APIs. There are various challenges related to smartphones, including restrictions from the operating system or its specific versions, absence of some sensors in different devices models, etc. Multiple sensors are available to provide information for indoor positioning. Moreover, smartphone camera, Wi-Fi or Bluetooth receiver may be considered as sensors in terms that they provide data with potential for positioning system incorporation. Motion sensors (e.g., accelerometer, gyroscope) used for measuring acceleration and rotational forces are not considered in this approach as they provide relative information based on previous device state and are universal in terms of the environment where they are used. To provide an input for neural network trained on a specific building, two different types of data are selected. The magnetometer measures the ambient geomagnetic field, and camera images or video sequence deliver visual information about the position. Values from these two sources differ for distinct places along a single corridor to some extent and therefore seems to be suitable for supplementing the original approach.

The paper is organized as follows. Chapter 2 provides an insight into related methods using magnetometer and images as inputs with some remarks regarding the practical usage of these applications. In chapter 3, the proposed system is described. Separate LSTM network for magnetometer measurements and CNN for camera images are introduced supplemented by an approach proposal for merging outputs. Chapter 4 summarizes evaluation performed in a specific building. Observations based on experiments and ideas for future work conclude the paper.

2. Background and related work

Machine learning, neural networks, and deep learning approaches have emerged as recent trends across various domains. Indoor positioning, as a field, encompasses a diverse range of methods that leverage neural networks for accurate positioning. Furthermore, numerous solutions have been proposed that extend beyond general positioning, addressing associated tasks and challenges in indoor environments. These solutions aim to tackle not only the determination of user or device location but also other related aspects, e.g., to determine direction [4], to measure radio fingerprints similarity [5], to detect steps [6], etc.

Wei and Radu [7] trained a recurrent neural network for location tracking using smartphone sensors accelerometer, gyroscope, and magnetic sensor. More specifically, long-short term memory (LSTM) neural network was employed in their study. This type of network is capable of utilizing for various sensors, such as magnetic and light data [8], for indoor positioning and

related tasks.

Sahar and Han [9] employed the method using LSTM network on Wi-Fi fingerprints. The authors discussed two potential implementations of the LSTM network: bi-directional and deep forward. In their study, they opted for the bi-directional approach, which considers previous and next timestamps and is followed by a network layer predicting the current state. On the other hand, the deep forward method relies solely on previous timestamps, making it well-suited for real-time applications. In our proposed solution, which is presented in this paper, we specifically employ the deep forward method using magnetic field measurements.

Ashraf et al. [10] demonstrated the long term stability of magnetic field values in terms of variation in time (collected on various days in multiple years), presence of furniture and pedestrians, and various devices.

Approaches utilizing magnetometers often prioritize dynamic movement over static measurements due to the limited range of values, where identical values may occur in different locations [11]. Kuang et al. [12] proposed a magnetic field matching method based on PDR where relative trajectories are matched. Similar to other studies, our solution involves transforming the coordinate system from the device-relative frame of reference to the world coordinate system.

Ouyang et al. [13] highlighted the drawbacks of LSTM networks, particularly in terms of the time-consuming training process, and the degradation problem with increasing the number of layers. Temporal convolutional network was adapted for the corridor trajectories classification.

Convolutional neural network (CNN) [14] is a popular choice in computer vision domain or in scenarios with images as input. Various use cases are addressed with such approach in the field of indoor positioning, e.g., fusion of inputs from static and dynamic cameras [15], passive visual positioning by CNN-based pedestrian detection [16]. However, in our study, we focus on utilizing the smartphone camera of the device itself for the purpose of localization, rather than relying on fixed-mounted cameras within the building, similar to IPIN competition camera-based tracks [17]. Solutions based on CNN and camera images are capable of improving other positioning methods, e.g., Wi-Fi based solution in crowded building [18].

Zhang et al. [19] proposed a deep convolutional network for scene recognition on building and room level. The approach based on dividing and identification of building parts is introduced in this paper. Walch et al. [20] employed a combination of CNN and LSTM on pose regression for indoor and outdoor scenes. LSTM units are applied on CNN output achieving promising results.

3. System Overview

The proposed system is based on neural networks (Figure 1). In this case, two distinct networks are considered for two types of data (magnetic field measurements and camera images). Input for the neural network is acquired from the smartphone. Even though the user may be occasionally standing still, in the majority of data the user is walking. The smartphone is considered to be hand held facing upwards or with the screen in front of the user, especially when using camera images (Figure 2). The solution aims to be robust to various inclinations of the device.

The proposed neural network method is applied for classification to decide the part of the building where the user is located. Building parts are distinguished manually before the training

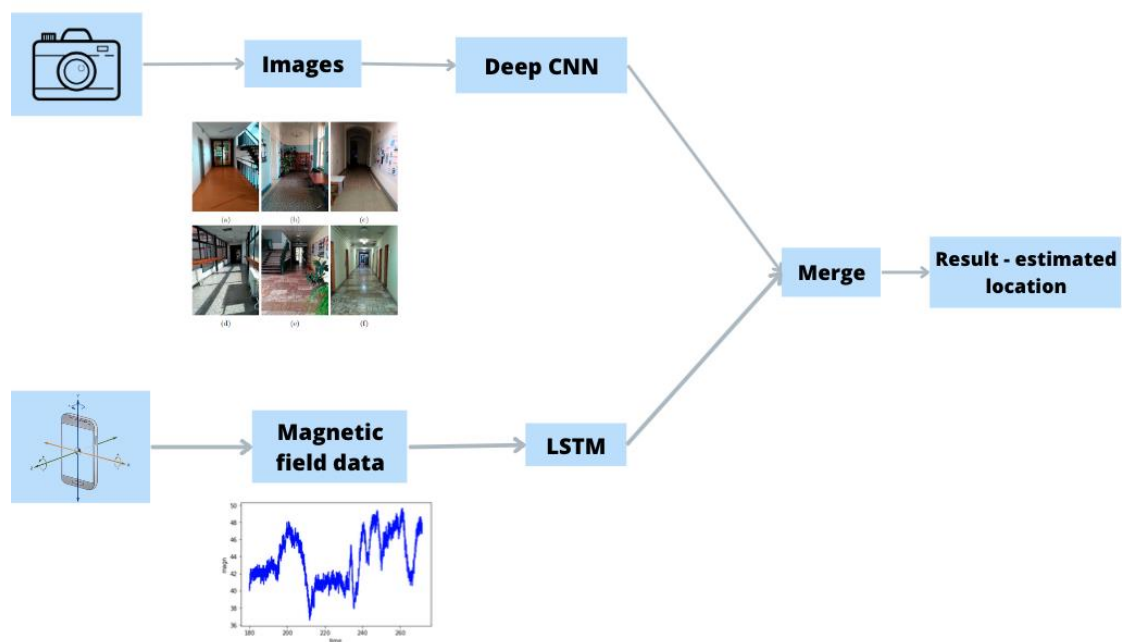


Figure 1: Overview of the system. Input images and magnetometer measurements are processed by separate neural networks. Outputs from these networks are merged to provide the final classification.

process.

The training process is performed with collected data from various devices and users. Measurements are collected when moving alongside the predefined trajectory. Data are labeled manually. Data augmentation for creating more diverse input may be performed.

3.1. LSTM using magnetometer data

Magnetometer measurements are obtained from smartphones using Android API. Calibrated magnetic field values along three axes are retrieved with 5Hz frequency. Measurements are transformed from device coordinate system to world coordinate system (same as in [13]):

$$\mathbf{m}_t^n = \mathbf{R}_t^{nb} \mathbf{m}_t^b$$

where $\mathbf{m}_t^b = (m_{x,t}^b, m_{y,t}^b, m_{z,t}^b) \in \mathbb{R}^{3 \times 1}$ is the measurement in the device coordinate system at time t . Rotation matrix $\mathbf{R}_t^{nb} \in \mathbb{R}^{3 \times 3}$ is provided by Android. The transformed value at time t is $\mathbf{m}_t^n = (0, m_{h,t}^n, m_{v,t}^n) \in \mathbb{R}^{3 \times 1}$ with horizontal ($m_{h,t}^n$) and vertical ($m_{v,t}^n$) component. Ideally, the first component should be zero after transformation, but in practice, it typically deviates from zero by a small value that does not cause any significant issues. After the transformation, horizontal and vertical components alongside with the magnetic-field intensity are used as feature vector $\mathbf{m} = (m_h, m_v, \sqrt{m_h^2 + m_v^2})$.

The LSTM neural network is employed for the purpose of classifying building parts. The inputs for the network are magnetic field vectors captured within a specified window of 10



Figure 2: A user collects data using a smartphone, where magnetic field values are captured and stored together with the corresponding camera image. The current position at a predefined checkpoint is labeled by pressing a designated button.

values, corresponding to a duration of 2 seconds. The proposed neural network architecture comprises four LSTM layers, each comprising 60 units, along with a dense layer. The number of units in the dense layer depends on the specific classes to be classified.

3.2. CNN classification of camera images

Dataset for image classification is prepared based on video recordings for easier labeling. Approximately 9 out of 10 frames are dropped. The image is scaled to a smaller resolution. The dataset is extended using data augmentation consisting of blurring, cropping, scaling, rotating, translating, shearing, and contrast or brightness changing. Images are presented to the neural network in batches.

Building is divided into visually distinguishable parts in advance. Three models are proposed for the evaluation.

- CNN without pretraining is a simple sequential model with four pairs of convolutional and pooling layers, followed by a flatten layer that converts the matrix to a vector. Finally, a classification layer is used to assign one of the selected classes.
- VGG16 model [21] without pretrained weights which is designed to improve the training time what was achieved by adding more convolutional layers. However, deepening the structure of the model means more computations and to avoid this, small convolution kernels were used and thus the number of parameters on the convolution layers was reduced.
- Pretrained model of VGG16 which was originally designed for classification into one of the 1000 classes and was trained on the ILSVRC-2012 dataset [21] consisting of 1.3M training images and 50K validation images scaled to size 224×224 . We load the pre-trained model from Keras library without top layer, that is without layer responsible for

classification so that we can adapt it for our problem. All its layers need to be frozen in order not to lose gained knowledge. The VGG16 model is connected with a flatten layer followed by two fully connected dense layers as in the original layer and finally a classification layer distinguishing between parts of the building. The training phase takes less time than without transfer learning because the network already knows a lot about the characteristics of the image and only needs to learn how to differentiate the individual parts of the building.

All three models are trained in 40 epochs using Adam optimizer [22], categorical crossentropy as the loss function, and accuracy metric.

3.3. Merging outputs of LSTM and CNN

Magnetic field measurements offer a means to leverage building knowledge for positioning purposes. However, this approach introduces several challenges, including low value discernibility, the need for calibration, device orientation issues, and the wide variety of devices used. On the other hand, camera-based approaches are susceptible to different environmental conditions, such as lighting variations, and are more sensitive to changes within the building, such as rearranged furniture or crowded corridors. Additionally, the current position may be in a different part of the building than the visible scenery captured by the camera.

Both approaches are suitable for positioning when utilizing a neural network specifically trained for the given building. The combination of these two inputs has the potential to complement each other. To address this, we propose the following merging technique:

- The output dense layers in both networks are eliminated from the architecture.
- The output vectors from both networks are combined and transformed into a unified vector representation of the inputs.
- A supplementary layer is introduced to perform the classification task. This layer takes the merged vector as input and generates the classification outcome.
- The classification process is triggered whenever a new input is detected, whether it is a recent camera image or a magnetic field measurement. This adaptive approach accommodates varying frequencies of input values.
- The division of the building into sections may differ between the LSTM and CNN networks. The final list of classes is obtained by uniting the respective sets of classes from both networks.

4. Evaluation

Independent experiments were conducted for each proposed method. The experiments took place in the faculty building located at Jesenná 5, 04001 Košice, Slovakia. This building served as a suitable evaluation environment due to its diverse characteristics and circumstances within a single venue. It consists of a historical section with high ceilings and small tiles on the floor, a newly reconstructed section with fewer windows, and a fully glazed connection corridor between these areas.

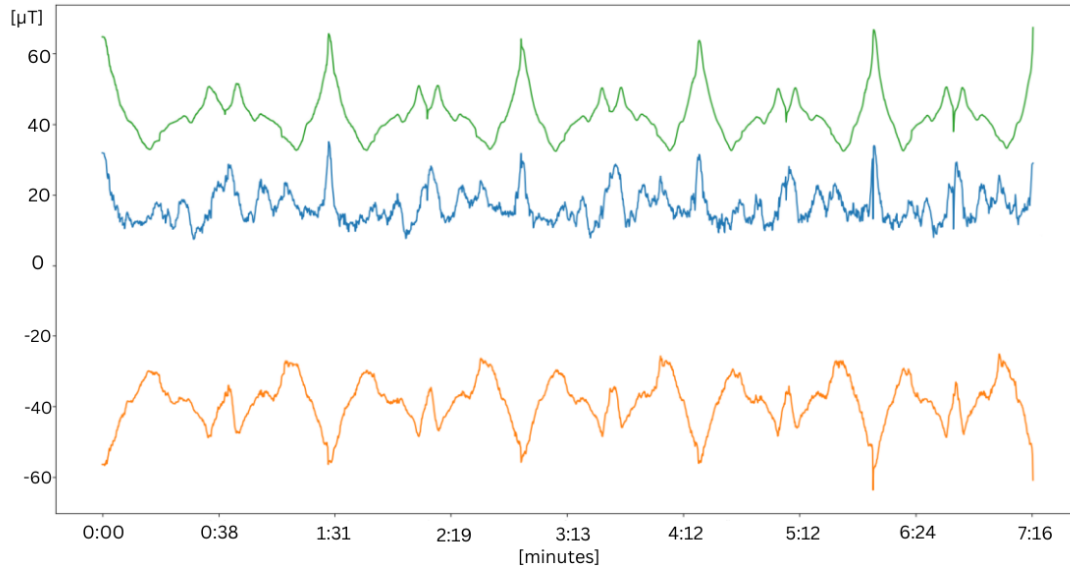


Figure 3: An example of collected data from the same corridors involves a user walking through the corridor five times in both directions. The collected data includes measurements of the magnetic field's horizontal component, represented by the blue line after being converted to the world coordinate system. The vertical component is denoted by the orange line, and the signal intensity is represented by the green line. These three values serve as inputs to the neural network.

The primary objective was to validate the feasibility of using these methods for positioning purposes. As such, the LSTM network was exclusively trained on a specific part of the building, which comprised similar corridors across multiple floors. These corridors presented greater visual challenges and were particularly relevant for magnetic field-based classification. Conversely, the CNN network was trained on the entire building but focused on a single floor, considering six visually distinct parts.

The merging of the outputs from these two networks is planned for future work. Additionally, a more comprehensive evaluation incorporating a larger number of classes and broader coverage of the building would be suitable for further refinement and precision.

4.1. Evaluation of magnetometer-based classification

Various experiments were performed in order to validate practical aspects of the magnetic field sensor. Even though the calibration process may induce sudden changes in data, these measurements were more stable compared to raw uncalibrated data obtained during multiple days. Figure 3 shows an example of measurements from the magnetometer.

Dataset for the evaluation was collected using 3 persons and 3 different smartphones Samsung Galaxy A52s 5G, Samsung Galaxy S8, and Xiaomi Mi 10 with no direct matchup between users and devices. Measurements were acquired on three floors. Data from various floors, devices, and users were not evenly distributed. Custom Android app was prepared for collecting magnetic field values. Users walked along predefined path manually indicating their position on selected checkpoints. The application recorded video from camera which forced users to prefer



Figure 4: Floor plan with four classes for classification based on magnetic field values. Each class is defined by a corridor segment and a walk direction. Every segment is approximately 17 meters long.

specific smartphone inclinations. Moreover, the collected dataset contains various scenarios, e.g., opening doors during walk, walking closer to a wall, different inclination, etc.

The dataset was divided into a training set (80%), a testing set (20%), and an additional subset of the training set (30%) was randomly selected as the validation set. The model was trained on prepared data in 100 epochs.

In the experiment, there were a total of 12 classes. The three floors, which featured corridors approximately 34 meters in length, were divided into two equal parts. The trajectories within each floor were classified separately for each direction, resulting in four classes for each individual floor. Figure 4 depicts the floor plan for the first floor, which closely resembles that of the second and third floors. The achieved classification accuracy of 35% for this particular task is relatively low. Nevertheless, a thorough analysis of the results revealed the underlying factors contributing to this outcome. While the model demonstrated proficiency in identifying the correct part of the corridor for the majority of cases, it frequently encountered misclassifications in terms of the floor affiliation (Figure 5).

Various alternatives were tested including changes in the network architecture and merge of two directions on the same path into a single class. The obtained results did not differ significantly.

An additional experiment was conducted, focusing on each floor individually. In this experiment, the corridor was divided into four segments of equal length, and both directions were taken into account for the classification task. As a result, a total of 8 classes were established for each floor. On test data, the model achieved the accuracy of 89.4% for the first floor, 94.4% for the second floor, and 73.5% for the third floor.

4.2. Evaluation of camera-based classification

While camera recordings were acquired simultaneously with the magnetic field values, a separate dataset was collected for evaluating the CNN model, covering a broader area of the building. The camera images were extracted from a video recorded over an extended period of time.

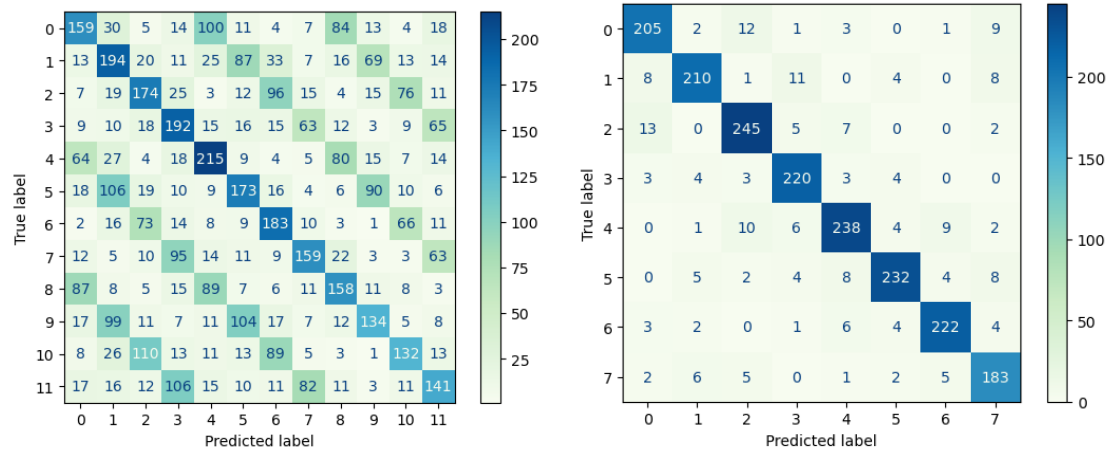


Figure 5: The first confusion matrix for the classification of three floors with 12 classes reveals that there are frequent misclassifications between corresponding classes on different floors. The second confusion matrix corresponds to the first floor, where the number of classes has been doubled compared to the first experiment, resulting in a total of 8 classes.

Consequently, the collected data are not restricted to a specific time of day, ensuring variations in light, weather conditions, and other circumstances.

The building was divided into six manually selected classes, with emphasis on visually distinguishable areas (Figure 6). For the evaluation, 30 videos were captured, comprising a total of 4000 images. The distribution of these images is not uniform. Three segments contain 900 images each, while the remaining three segments have a combined total of 1300 images. This distribution reflects the fact that some parts of the building are more complex, while others are smaller in area but may be easier to distinguish.

Accuracy and F1-score were calculated for the proposed models:

- CNN without pretraining - accuracy 93%, F1-score 0.85
- VGG16 without pretrained weights - accuracy 96%, F1-score 0.86
- VGG16 with pretrained weights 98%, F1-score 0.94

Upon closer examination of the results, no significant observations were found. As anticipated, misclassifications were more prevalent among similar classes, particularly within the larger sections of the historical part of the building (as depicted by the orange and green elements in Figure 6). The main distinguishing factors in these cases were the floor tiles and the presence of windows, while the walls and doors appeared similar.

Furthermore, an additional experiment was conducted using the VGG16 architecture with pretrained weights, which yielded the best results. Two new routes were recorded on a separate day, covering half of the building. Frames from the video were directly inputted into the model without any contextual information. The model achieved a classification accuracy of 80% for the frames from the first video and 60% for the frames from the second video. Upon analysis, it was observed that the majority of errors occurred in areas with poor lighting conditions.

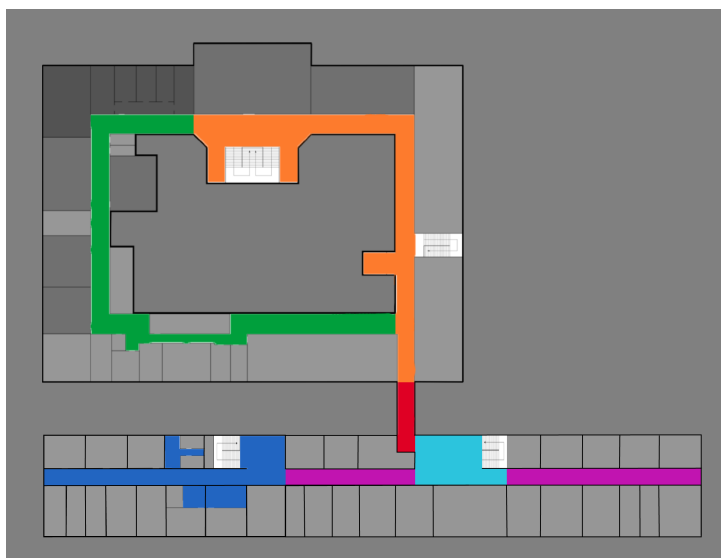


Figure 6: Floor plan of the faculty building with 6 visually distinguishable building parts labeled by various colors. Green and orange parts are in historical building with high ceiling and small tiles on the floor. Blue and purple colors denotes newer part of the building with less windows and more doors. Cyan color represents an open corridor with staircases and red color is for a connection between buildings which is fully glazed.

The classification was performed solely on individual image frames without considering their context in video. Introducing a network model capable of processing images chronologically could be advantageous in addressing this limitation. Additionally, a significant challenge in visual classification arises from the fact that the user's physical location may differ from the area visible through the camera. This issue is expected to become more prominent when dealing with a higher number of classes that represent smaller building parts.

4.3. Discussion

The proposed system underwent separate evaluations for the LSTM network using magnetometer measurements and CNN with images. However, the final step of merging outputs from these two networks remains pending and is planned for future implementation. Additionally, the magnetometer-based method was also examined in other areas of the building. The observations, in conjunction with references to prior research, suggest that the approach holds promise for universal applicability, despite being evaluated solely in a single building. To employ the proposed solution in different buildings, new data collection and training would be necessary.

Both methods are well-suited for real-time scenarios. The Android system provides magnetic field measurements at a higher frequency than utilized in this approach. Furthermore, it allows for adjusting the frequency or expanding the time window, currently set at 2 seconds of data span. Similarly, the frequency of input images is customizable, with the option to drop more frames if computational requirements demand it. While the positioning does not rely solely on this method, its incorporation aims to enhance the overall system accuracy.

The solution targets the weaknesses of the existing positioning system based on PDR, Bayesian filtering, and map constraints. By using two neural networks, inaccurate or ambiguous outputs from one network can potentially be corrected by the other. Additionally, the positioning system should be designed to utilize outputs from neural networks with a certain level of uncertainty, acknowledging the possibility of inaccuracies.

For full integration of the proposed method into the location determination solution, an extended evaluation is essential. The experiments in this paper primarily demonstrate key features and the ability to distinguish positions within a single corridor. To establish greater robustness, more comprehensive experiments are warranted, alongside observations of the positioning system using input from the proposed method.

Conclusion

The paper presents a model for classifying building parts using two neural networks, one for handling magnetic field measurements and the other for processing camera images. Separate datasets were collected and used to train the model to distinguish different building parts. The evaluation of merging the outputs from the two networks is planned as future work.

The LSTM network applied to magnetic field values achieved accuracies ranging from 73% to 95% on individual floors, and 35% when combining results from a challenging area. However, the analysis of the results reveals that even with relatively low accuracy, the model is still applicable as it helps compensate for the weaknesses of the indoor positioning method considered, which tends to introduce errors during long, straight walks along corridors.

The CNN evaluation yielded the best results using the VGG16 model with pretrained weights, achieving 98% accuracy and a 0.94 F1-score. This model correctly classified approximately 80% and 60% of frames in two trajectories. Taking video sequences into account is expected to improve the overall accuracy of the classification. Currently, the model processes image frames individually, without considering the temporal information present in video data. By incorporating the sequential nature of video sequences, the model can capture and leverage the contextual information and dependencies between frames.

Further complex evaluations should be considered in future studies. Increasing the number of classes may introduce new challenges, particularly in distinguishing similar parts of the building. An open problem to address is how to automatically divide the building into meaningful classes, with one possible approach being the automatic clustering of images to identify similar areas.

The motivation of this paper was to propose a neural network-based method specifically designed to leverage the characteristics of a particular building. In order to achieve this, magnetic field sensors and cameras were chosen as sources of information, as they can provide unique and specific output related to different building parts. By training the neural network on these specific inputs, the aim was to develop a model that can effectively classify and identify the various parts of the building based on their distinct characteristics captured by the magnetic field and camera images. The results from experiments utilizing both neural networks indicate that the proposed approaches are suitable and feasible for enriching indoor positioning with additional information.

Acknowledgments

This paper was supported in part by the Slovak Grant Agency, Ministry of Education and Academy of Science, Slovakia, under Grant 1/0177/21, and in part by the The Cultural and Education Grant Agency, under Grant 012UPJŠ-4/2021.

References

- [1] G. M. Mendoza-Silva, J. Torres-Sospedra, J. Huerta, A meta-review of indoor positioning systems, *Sensors* 19 (2019) 4507.
- [2] F. Potortì, J. Torres-Sospedra, D. Quezada-Gaibor, A. R. Jiménez, F. Seco, A. Pérez-Navarro, M. Ortiz, N. Zhu, V. Renaudin, R. Ichikari, et al., Off-line evaluation of indoor positioning systems in different scenarios: The experiences from ipin 2020 competition, *IEEE Sensors Journal* 22 (2021) 5011–5054.
- [3] M. Opiela, F. Galčík, Grid-based bayesian filtering methods for pedestrian dead reckoning indoor positioning using smartphones, *Sensors* 20 (2020) 5343.
- [4] P. Babakhani, T. Merk, M. Mahlig, I. Sarris, D. Kalogiros, P. Karlsson, Bluetooth direction finding using recurrent neural network, in: *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2021, pp. 1–7.
- [5] B. Dong, T. Burgess, H.-B. Neuner, S. Fercher, Neural network based radio fingerprint similarity measure, in: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2018, pp. 1–8.
- [6] N. Al Abiad, Y. Kone, V. Renaudin, T. Robert, Smartphone inertial sensors based step detection driven by human gait learning, in: *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2021, pp. 1–8.
- [7] X. Wei, V. Radu, Calibrating recurrent neural networks on smartphone inertial sensors for location tracking, in: *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2019, pp. 1–8.
- [8] X. Wang, Z. Yu, S. Mao, Deepml: Deep lstm for indoor localization with smartphone magnetic and light sensors, in: *2018 IEEE international conference on communications (ICC)*, IEEE, 2018, pp. 1–6.
- [9] A. Sahar, D. Han, An lstm-based indoor positioning method using wi-fi signals, in: *Proceedings of the 2nd International Conference on Vision, Image and Signal Processing*, 2018, pp. 1–5.
- [10] I. Ashraf, Y. B. Zikria, S. Hur, Y. Park, A comprehensive analysis of magnetic field based indoor positioning with smartphones: Opportunities, challenges and practical limitations, *IEEE Access* 8 (2020) 228548–228571.
- [11] G. Ouyang, K. Abed-Meraim, A survey of magnetic-field-based indoor localization, *Electronics* 11 (2022) 864.
- [12] J. Kuang, T. Li, X. Niu, Magnetometer bias insensitive magnetic field matching based on pedestrian dead reckoning for smartphone indoor positioning, *IEEE Sensors Journal* 22 (2021) 4790–4799.

- [13] G. Ouyang, K. Abed-Meraim, Z. Ouyang, Magnetic-field-based indoor positioning using temporal convolutional networks, *Sensors* 23 (2023) 1514.
- [14] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, *IEEE transactions on neural networks and learning systems* (2021).
- [15] C. Kim, C. Bhatt, M. Patel, D. Kimber, Y. Tjahjadi, Info: indoor localization using fusion of visual information from static and dynamic cameras, in: *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2019, pp. 1–8.
- [16] D. Wu, R. Chen, Y. Yu, X. Zheng, Y. Xu, Z. Liu, Indoor passive visual positioning by cnn-based pedestrian detection, *Micromachines* 13 (2022) 1413.
- [17] V. Renaudin, M. Ortiz, J. Perul, J. Torres-Sospedra, A. R. Jiménez, A. Perez-Navarro, G. M. Mendoza-Silva, F. Seco, Y. Landau, R. Marbel, et al., Evaluating indoor positioning systems in a shopping mall: The lessons learned from the ipin 2018 competition, *IEEE Access* 7 (2019) 148594–148628.
- [18] J. Jiao, F. Li, Z. Deng, W. Ma, A smartphone camera-based indoor positioning algorithm of crowded scenarios with the assistance of deep cnn, *Sensors* 17 (2017) 704.
- [19] F. Zhang, F. Duarte, R. Ma, D. Milioris, H. Lin, C. Ratti, Indoor space recognition using deep convolutional neural network: a case study at mit campus, *arXiv preprint arXiv:1610.02414* (2016).
- [20] F. Walch, C. Hazirbas, L. Leal-Taixe, T. Sattler, S. Hilsenbeck, D. Cremers, Image-based localization using lstms for structured feature correlation, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 627–637.
- [21] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [22] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: *ICLR*, 2014.