



Deep learning-based application for indoor wayfinding assistance navigation

Mouna Afif¹ · Riadh Ayachi¹ · Yahia Said^{1,2} · Mohamed Atri^{1,3}

Received: 10 September 2020 / Revised: 30 March 2021 / Accepted: 4 May 2021 /

Published online: 13 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

There is an increasing need to develop new adaptive technologies and new wayfinding assistance systems for blind and visually impaired persons in order to improve their daily lives. To address this need, we propose in this paper to develop a new deep learning-based indoor wayfinding assistance system consisting of detecting landmark indoor signs. Assistive technologies used for blind and sighted persons used to support daily activities to improve social inclusion are developing very fast. Training and testing experiments were performed on the proposed indoor signage dataset. Through the experiments conducted, we demonstrated the efficiency of the proposed indoor wayfinding aid system. We obtained 93.45% as a mean average precision (mAP) of the proposed indoor wayfinding and signage detection system.

Keywords Indoor sign detection · Indoor wayfinding · Blind and visually impaired persons · Deep learning · Convolution neural network (CNN)

1 Introduction

According to the latest statistics of the World Health Organization (WHO) 188.5 million persons suffer from visual impairments, 217 million persons present moderate to severe visual impairment and 36 are blind [29].

Visual impairment present a real handicap for these persons as it reduces their mobility and their participation in the daily social life. A part of electronic and computer science-based systems,

✉ Mouna Afif
mouna.afif@outlook.fr

¹ Laboratory of Electronics and Microelectronics (EμE), Faculty of Sciences of Monastir, University of Monastir, Monastir, Tunisia

² Electrical Engineering Department, College of Engineering, Northern Border University, Arar, Saudi Arabia

³ College of Computer Science, King Khalid University, Abha, Saudi Arabia

many achievements including neurosciences, biology and medicine search to find new solutions for blind and sighted [21] to improve their quality of life and ensure a better integration in the social life. Recently, many researchers have been working on developing new assistive systems [13, 24] to improve the navigation for blind and impaired people and to reduce the dangers present in the surrounding environment. To satisfy the sighted person navigation assistance requirements, many assistive systems apply recognition technologies [25].

Developing an indoor navigation system for blind and sighted persons present an incredibly daunting challenge. The biggest complexity of this task lies on the complexity of the indoor environment as it presents many shapes and many object textures. In order to develop an application, used especially for indoor wayfinding, we have to develop an indoor sign detector as it presents the most reliable way because GPS and associated advances for outdoor environments do not apply. Indoor environments are very difficult environments to be understood. Visual information, provided in indoor environments, are difficult for people with low vision to follow. There is a great need to develop a low cost and a reliable assistance wayfinding system for blind and sighted persons. Generally, persons rely especially on perception and vision information to know their positions and to recognize objects, orientations and their directions. This challenging task is belonging to wayfinding task while the capability to recognize objects and avoid obstacles relies on mobility task.

Currently, there are few works done on navigational aids in large new unfamiliar environments for blind and visually impaired individuals. Indoor spaces present many decorations and are difficult to navigate on. Several works were done in this field by building and deploying wayfinding aid systems for blind and sighted persons. Vision impairments have a significant impact on impaired persons' lives and limit them from doing daily activities and from navigation in indoor environments. Our proposed work provides an assistive tool to improve blind and sighted persons quality of life and allow them a better integration into social life.

In our work, we will couple two main facts: (1) navigation, (2) object detection and recognition. By integrating both of navigation and object recognition we can dramatically improve the blind and VIP daily lives. In this paper, we propose new indoor wayfinding assistance based on detecting a set of indoor signs using the one-stage RetinaNet [20] neural network. In the regard of aiding blind and impaired persons, the proposed indoor wayfinding assistance system is presented in this work in order to provide blind and visually impaired persons (VIP) with information presented in its surrounding indoor environments. Nevertheless, recently many works were done in this regard to solve this big challenge, but generally they propose to solve this problem as a classification problem. In this study, the proposed system is able to recognize and localize new landmark indoor signs as (exit, wc, confidence zone and disabled exit) in order to perform an indoor wayfinding assistance navigation. We note that the proposed indoor wayfinding assistance system achieves very interesting detection results in terms of detection precision as well as detection speed. We also note that this work present the first work evaluating the deep learning-based architecture RetinaNet to build an indoor wayfinding assistance system. The proposed indoor wayfinding assistance system has been trained and evaluated in the proposed indoor signs dataset which consists of 4000 indoor images containing 4 landmark signs highly recommended for indoor wayfinding assistance navigation. This dataset is new and original as it contains various challenging conditions including: different lighting conditions, occlusion, different objects size and distances to camera and different objects positions and point view. Another main advantage presented in this work is that the proposed indoor wayfinding assistance system is able to detect new landmark indoor signs that were not studied before by state-of-the-art works.

The reminder of this paper is the following: Section 2 reviews the previous works performed on indoor signs detection and wayfinding assistance. Section 3 outlines the main contributions of the proposed work. In Section 4, we detail the proposed architecture used for indoor wayfinding assistance. Section 5 outlines the proposed experiments and discusses the obtained results and Section 6 concludes the paper.

2 Related work

Moving and exploring during traveling and indoor navigation is one of the most relevant and challenging problems faced by blind and sighted persons during their daily activities. In literature, there are many ways and techniques used to help blind and VIP to make their navigation easier as canes and guide dogs. However, these techniques are used by few sighted persons. Over the last few years, developing new indoor object detection systems for blind and visually impaired persons wayfinding assistance remained a very challenging problem for artificial intelligence and computer science community.

2.1 Vision based systems

Many ways are proposed in the literature to define new assistance systems. Among the very known systems, the vision-based assistive system which is based on using different cameras as: stereo-cameras, mono-cameras or RGB-D cameras in order to provide visual information of the real-world environment and to be used in a wide range of tasks as image segmentation, object detection and recognition or face detection. In [16], authors developed an assistive system used for detecting obstacles based on corner detection. In order to provide real world images, authors employed cameras and infrared sensors of Kinect.

Various systems, used for public safety, have been proposed in the literature as in [10]. RGB-D sensors are very used in visually impaired assistance systems and have stirred the interest of many researchers. Yang et al. [30] proposed a new approach which expands the detection traversable area based on an RGB-D camera sensor. This system is compatible with indoor and outdoor environments. A new robotic assistance navigation system is proposed in [32]. This system is called co-robotic cane (CRC). This system uses 3D cameras for pose estimation and for object recognition in unfamiliar environments. The indoor object detection system is made by employing an algorithm of gaussian-mixture model (GMM). A new object recognition method based on 3D input images is proposed in [31] in order to contribute to a safer navigation for blind and visually impaired persons. This assistive method is based on Gaussian mixture model (GMM) in order to classify planar patch belonging to a particular object model and then the detected objects will undergo a clustering procedure. Light weight devices are largely used in nowadays applications. These devices are excessively used for assistive applications dedicated for blind and sighted persons. An android application for smartphones especially used to assist these persons is proposed in [12]. The application is developed by using MEMS sensors. The primitive goal of assistive technologies is to help persons and to reduce the range of their limitations. In [26], authors proposed a wayfinding assistance system used to help blind people access unfamiliar buildings. As indoor signage detection plays an important role to find destinations, authors in [28] proposed a signage and doors detection system used for indoor navigation and wayfinding. Mobility presents a serious challenge for blind and persons with visual impairment. A sign-based system is proposed in [22] to help visually impaired to navigate in indoor buildings. The

authors proposed to develop a smartphone-based navigation aid system. As indoor navigation present a major challenge for blind and persons with visual impairments. In [14], authors proposed a real-time application aims to build a wayfinding ios app by detecting a set of indoor signs. In [18], authors provide a comprehensive survey on various indoor navigation and indoor positioning technologies. Also authors reviews computer vision-based indoor navigation assistance systems based on scene recognition techniques.

2.2 Deep learning based systems

Recently, deep learning algorithms became more and more ubiquitous it was involved in many computer vision and artificial intelligence tasks. This technique was applied to various computer vision tasks as indoor object recognition [1], traffic sign detection [7], pedestrian detection [8], indoor scene recognition [4], ship detection [11], data aggregation [6] and citywide traffic crowd flows [5]. Many computers vision-driven assistance systems were proposed in literature to help blind and sighted persons in indoor navigation assistance and to avoid surrounding dangers in indoor environments. In our previous works [2, 3], we proposed two indoor object detection systems based on one-stage object detection in order to assist blind and impaired persons to fully assist them during their indoor navigation and to explore more their surroundings. A new multi-model technique was used to assist visually impaired persons to recognize some specific indoor objects [27]. This system is developed by using a complex-valued neural network. It takes as an input an RGB image and as an output it recognizes some specific indoor objects present in the surrounding environment. Another assisting system was developed based on deep learning algorithm is proposed in [9]. This assistance system is used specially to assist blind and visually impaired persons in specific indoor environments as clinics, hospitals and urgent cares. This system detects especially some indoor objects as doors, stairs and some indoor signage. A simple smartphone-based guiding and assisting blind and sighted persons in indoor environments is proposed in [19]. This object detection system achieves 60% as recognition rate. Many new deep neural networks can be included to build new assistive applications dedicated for blind and sighted persons that reduce dangers in the surrounding indoor environments of these persons.

3 Contributions

- The proposed indoor sign detection system presents a new assistive system highly recommended for indoor wayfinding assistance of blind and sighted persons.
- The proposed detection system is based on a one-stage convolutional neural network.
- The proposed indoor wayfinding assistance system achieves high detection accuracies although it was trained and tested using different challenging conditions.
- This paper presents the first approach evaluating RetinaNet network in indoor wayfinding assistance.

4 Proposed indoor signage detection system

In order to improve the ability of blind and sighted persons to independently access and explore unfamiliar environments, we propose an indoor signage detection system based on deep learning models.

4.1 Overview of the proposed method

The visual system of human being is very powerful and robust, extremely fast and can discriminate between thousands of different objects categories. Despite the complexity of the vision capabilities of human it is extremely hard to build a computer vision algorithm used for object detection. Motivated by this fact, in this paper we propose to develop an indoor signage detection system used for blind and sighted persons for wayfinding aiding them to independently navigate in unfamiliar indoor environments. In order to develop a very effective and robust indoor signage detection system we take into account the following issues:

- (1) Various lighting conditions
- (2) High intra and inter-class variations
- (3) Unsafe situations to avoid
- (4) New indoor sign to be detected
- (5) Occlusion, different shapes and different textures
- (6) Different points of view of the indoor sign

Among indoor objects, signage presents an important landmark component for indoor navigation and wayfinding. Therefore, reliable and efficient indoor signage detection is a key component for wayfinding aiding. In order to provide the maximum of information for the blind and the sighted person, our proposed detection has the ability to detect specific indoor signage that is considered for the first time.

Due to the rise of deep learning during the last few years, it made previously unsolvable tasks possible. Blindness and visions impairments cause a considerable isolation from society and community. The proposed work aims to provide the blind and the sighted person with an impressive assistive technology in order to help these persons to integrate more in the daily life and with another person. Our proposed sign detection system is built based on deep convolutional neural network RetinaNet [20]. By proposing this work, we ensure the blind and the VIP by an increasing level of independence and autonomy.

As presented in Figs. 1 and 2, our proposed indoor sign detection system is based on: an input image from RGB color space. The detection system is based on RetinaNet neural network. The output desired from our proposed work is that all the signs presented on the input image will be detected as well as their class names and coordinates. We also note that our proposed method does not undergo pre- or post-processing.

4.2 Proposed indoor detection system architecture details

Our proposed indoor signage detection system is used for blind and sighted persons for indoor wayfinding assistance. The proposed method is based on RetinaNet [20] deep convolutional neural network. Many neural networks suffer especially from the problem of class unbalance. In order to address this problem, we developed our proposed application based on RetinaNet

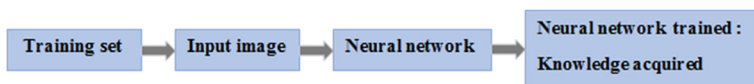


Fig. 1 Train process: knowledge acquisition

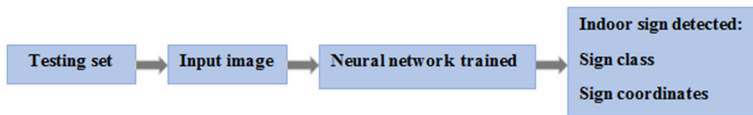


Fig. 2 Test Process: indoor sign detected

which provides a new loss function that contributes for more effectiveness compared to other models on the class imbalance problem.

Based on our literature research, RetinaNet network showed both: efficiency of detection and speed process compared to other one and two-stage detectors.

RetinaNet network is based especially on:

- A backbone built on the top of the feature extractor ResNet [15] called feature pyramid network FPN, used to calculate convolution feature map of the input image.
- A classification head: a subnetwork used to perform classification using the backbone's output.
- A regression head: a subnetwork responsible for performing bounding box coordinates using the backbone's output.

The use of FPN involves two pathways connected to the lateral connections:

- Bottom-up pathway: it chooses the last features map of successive layers that output features map of the same scales. These features map will be used as a foundation of the FPN. Figure 3 presents the bottom-up pathway movement introduced in the FPN architecture.
- Top-down pathway and lateral connections: the last feature map of the bottom-up pathway is expanded to present the same scale as the second to the last feature map. These features map will be merged to form a new features map. This process will be repeated until each features map of the bottom-up pathway has a corresponding feature map in the top-down pathway connected by lateral connection. The following Fig. 4 present the overall architecture with all the connections adopted in the RetinaNet architecture.

Generally, objects of the same class category may be presented under different scales in the input image. This fact may increase the detection accuracy of the neural network especially for

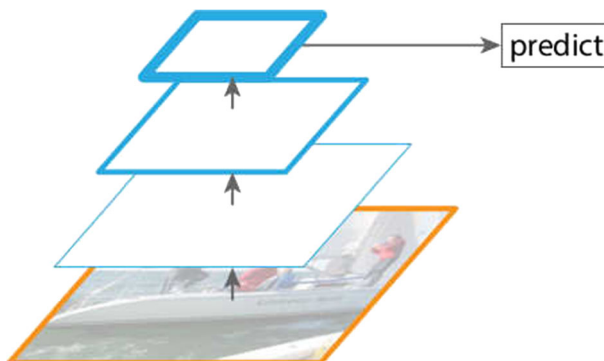


Fig. 3 Bottom-up pathway architecture

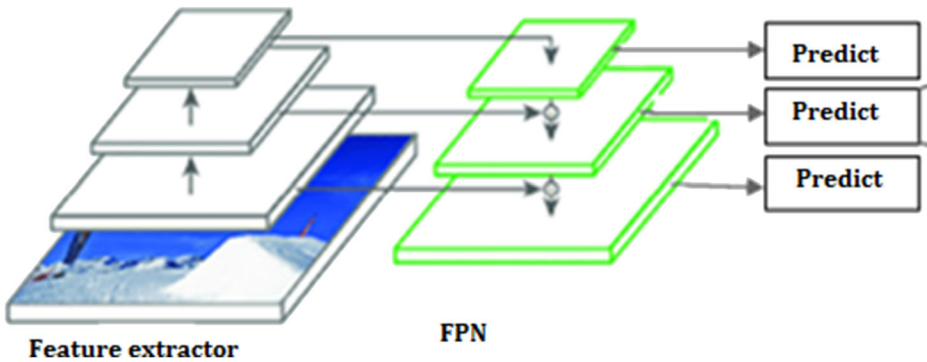


Fig. 4 Top-down pathway architecture

the small objects. Many neural networks use only the last feature map of the network to make detection prediction but this technique is not ideal.

FPN is basically a fully convolutional neural network which aims to take an image of an arbitrary size and outputs proportional features map sizes for multiple levels. High level features map contains grid cells covering large regions of the image which are very suitable for detecting large objects while grid cells of lower levels of features map are especially used for detecting small objects present in the input image. Figure 5 details the process of dividing the input image into cells in order to obtain low level and high level feature map.

Classification head The classification subnet is actually an FCN tied to each FPN level. Classification subnet consists of a 3×3 convolution layers with 256 filters. Each convolution layer is followed by an activation layer RELU. After that another 3×3 convolution layer of $K \times A$ filters is applied. Finally, the shape of the output is under the following shape (W,H,K,A) where W,H are the width and the height of the feature map respectively, K and A are the class objects number and the anchor box.

Regression head The regression head is attached to the FPN in parallel with the classification subnet. Regression subnet presents an identical design as the classification subnet. The

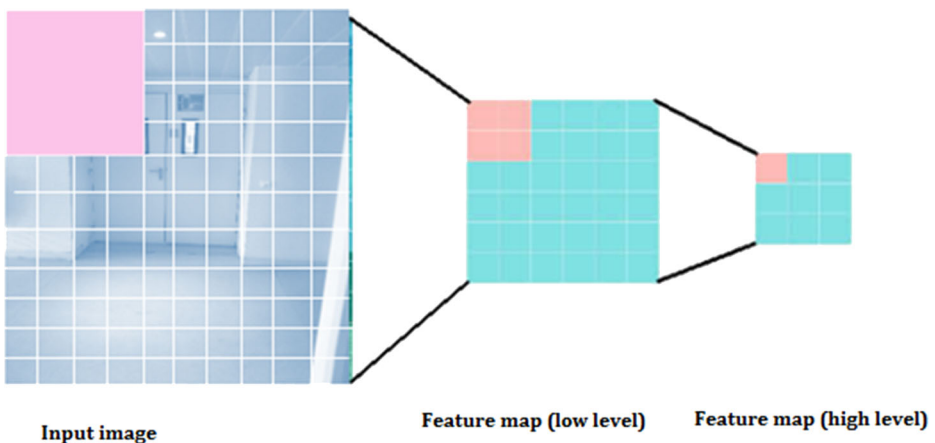


Fig. 5 Feature map extraction: low and high representations extraction

difference is that the last convolution layer is of the shape 3×3 , 4A filters. So the output shape is (W,H, 4A).

Let’s suppose that a feature map output of an FPN is 3×3 . So, for every grid cell of the nine, RetinaNet network define $A = 9$ anchor box. Each anchor box presents a shape and an aspect ratio. Each anchor box is responsible for detecting the existence or not of an object from K object class in the area that it covers. As each A anchor box can contain K object class, so the output shape of the classification head is KA channels.

For the regression subnet, it is responsible for detecting the existence of the object location and the object class. Also, it is responsible for providing the shape and the size of the detected object so the output shape of the regression subnet is 4A channels. The anchor box is matched to the ground truth of its intersection over union IOU is more than 0.5, the ground truth labels will be assigned to a tensor. The biggest originality of RetinaNet is its loss function. RetinaNet architecture provides two loss functions: one for localization and another for classification.

$$L = \lambda L_{loc} + L_{cls} \tag{1}$$

λ : is a balancing hyper-parameter to control imbalance between the two losses. The regression loss and the classification loss are calculated based on match between the anchors and the ground truth. So, we obtain the matching pairs as $(A^i, G^i)_{i=1..N}$: N are the matches, A are the anchors, G: ground truth. For every matching anchor on the regression head it predicts 4 parameters which can be presented as $P^i = (P_x^i, P_y^i, P_w^i, P_h^i)$. P_x and P_y specify the offset existing between the anchor center A^i and the ground truth G^i , P_w and P_h specify the offset existing between anchor’s width/height and the ground truth. So, for every prediction there is a recognition target T^i :

$$T_x^i = (G_x^i - A_x^i) / A_w^i \tag{2}$$

$$T_y^i = (G_y^i - A_y^i) / A_h^i \tag{3}$$

$$T_w^i = \log (G_w^i - A_w^i) \tag{4}$$

$$T_h^i = \log (G_h^i - A_h^i) \tag{5}$$

By considering all the above equations, the regression loss can be written as follows:

$$L_{loc} = \sum_{j \in \{x,y,w,h\}} smooth_{l1} (P_j^i - T_j^i) \tag{6}$$

$$smooth_{l1} = \begin{cases} 0.5 x^2 & , |x| < 1 \\ |x| - 0.5 & , |x| \geq 1 \end{cases} \tag{7}$$

The classification loss can be defined as:

$$L_{cls} = -\sum_{i=1} (y_i \log(p_i) (1-p_i)^\gamma \alpha_i + (1-y_i) \log(1-p_i) p_i^\gamma (1-\alpha_i)) \tag{8}$$

$y = 1$ if ground truth belongs to i -th class and 0 otherwise.

P_i = prediction probability of the i -th class.

γ belongs to $(0, +\infty)$: focusing parameter.

$\alpha_i \in [0, 1]$: weighing parameter.

Figure 6 present the detailed architecture of RetinaNet used in this work in order to obtain and indoor wayfinding assistance system. In order to address the problem of the class imbalance which limits the detector’s performances, RetinaNet’s focal loss introduces a focusing parameter γ to down weigh loss of easily classified examples. Also, the balancing parameter α address as well the class imbalance problem. An object presenting an image can be detected by multiple anchor boxes.

To reduce the number of the anchor boxes detected, a non-maximum-suppression (NMS) is applied to select the anchor with the height confidence score. In this stage, every remaining anchor will be used as a bounding box prediction.

5 Experiment procedures and results

In order to evaluate the proposed indoor sign detection approach, multi-object datasets devoted for multi-class object detection are required to train and to test the proposed detection system. Due to the lack and the limitation of the datasets present, we propose to build our own dataset which covers very challenging and dangerous situations to avoid when blind or sighted person navigate in indoor scenery.

5.1 Data source and collection

Due to the lack of the indoor sign detection datasets, we proposed to create the new dataset in order to train and test our proposed indoor signs detection system. The proposed dataset contains 4000 indoor images composed of images under challenging conditions as different lighting conditions, different viewpoints of objects and occlusion. We note that the proposed dataset was divided into two main parts: one for training including 2600 indoor images and a test set containing 1400 images. The number of daylight images is around 2100, 900 night images and 1000 are taken under artificial lighting.

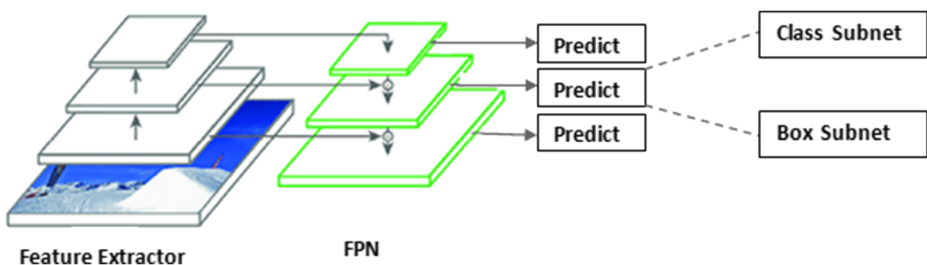


Fig. 6 RetinaNet Architecture Used for Indoor sign Detection System

The dataset provides 4 indoor signs (wc, exit, disabled exit and confidence zone) which are highly recommended for indoor assistance navigation for blind and sighted persons. The presented dataset is crated and labeled in order to train and test powerful deep learning algorithms in order to create new robust and accurate sings detection and recognition systems. The proposed indoor sign detection system can widely improve the life quality for blind persons by ensuring for them a safer navigation. In the following Fig. 7, we present an example of images from the proposed indoor dataset used in this work.

5.2 Experimental results

We constructed an indoor signage dataset providing four landmark indoor signage: exit, wc, disabled exit and confidence zone. The proposed dataset covers many situations as different lighting conditions (daylight, night), occlusion, many points of view of signs, different shapes and sizes and different textures.

To evaluate the robustness of our proposed indoor signage detection system, we trained and tested the detection algorithm on the proposed indoor landmark sign dataset. In this work, we investigated the problem of indoor navigation for blind and visually impaired individuals to reach their destinations. As some indoor objects are landmarks, we believe that it plays a central role for blind and sighted person's navigation and wayfinding assistance.

As RetinaNet achieves good results by employing a reshaped Cross entropy function named "focal loss" to address the problem of class imbalance, we use RetinaNet to realize an indoor signage detection system. During train process, images undergo a resizing to fit the deep convolution neural network. Although, our proposed deep CNN-object detection system can provide the indoor sign location and name and the bounding box (bbox) containing the



Fig. 7 Images example from the proposed dataset

indoor sign. Our proposed system presents a crucial way to help blind and visually impaired person and to improve the quality of their lives to help them to explore new unfamiliar indoor environments for better integration in daily life.

We trained and tested our indoor signage detection system on our collected dataset. The dataset presents many challenging conditions in order to contribute for more robust detection system and to ensure for the blind or the sighted person a maximum of security when accessing new unfamiliar environments. Images present in the proposed dataset are taken under various challenging conditions as complex background, various illumination conditions (day, night, dark), different points of view of the indoor signage and the different distances between camera and the indoor object concerned. The proposed indoor signage detector presents four indoor landmark signs (exit, disabled exit, wc, confidence zone). There are indoor sign objects, that are considered on our proposed work, were not considered previously.

We trained the proposed detection system using ResNet 50 [15] backbone. We note also that the dataset was splatted into two-parts: one for train and the second for test respectively. The following Fig. 8 depicts the different layers used in ResNet 50 architecture used to extract features.

To evaluate the performance of the proposed indoor signage detection system we have used mean average precision (mAP). In order to obtain the maximum of performance, when training a deep convolutional neural network, it require a huge amount of data. For this fact, when training the deep learning model, we used data augmentation to increase the training data. In the data augmentation process, images undergo many processing as: translation, rotation, flipping and scaling. Experiments setup used in our implementation are the following: 50 epochs in which 10,000 iterations. We initialized the learning rate to 0.0001 and 100 as a batch size. Our experiments were performed on an HP workstation equipped with an intel Xeon E5-2683 V4 processor and with an NVIDIA Quadro M4000 as a graphic processor presenting 8 GB of graphic memory. To implement our proposed detection application, we used the Keras-TensorFlow framework. We also used python 3.6 TensorFlow-GPU 1.13 NVIDIA CUDA toolkit 10.0 and the deep neural network library CUDNN 7.0 version. Table 1 provides all the experiments' settings used during the proposed experiments.

As first step, we trained the neural network using the stochastic gradient decent (SGD). We obtained very interesting results for all the indoor sign classes as mentioned in Table 2. As presented in Table 2, we achieved very encouraging results coming up to 92.44% as a mean average precision of the four landmark indoor signs proposed in the proposed dataset. Average precision of the four indoor signs are almost presenting high detection accuracies more than 91%. All these results are obtained by using the stochastic gradient decent optimizer [23] during the training process.

The SGD optimizer updates the network parameters at each training epoch. This parameters updates update can cause high oscillations of the objective functions which enables the loss function to reach the local minimum. To solve the problems caused by SGD optimizer, we proposed to retrain the network with ADAM optimizer. The ADAM optimizer computes the learning rate for each parameter and updates parameters for each training step. As presented in Table 3, we obtained very encouraging results, we noted that by using ADAM as a network optimizer we improved the detection accuracy for each class sign by around 1% from the results obtained with SGD optimizer.

As mentioned in Table 3, we used the Adam optimizer [17] for our second implementation of our proposed indoor sign detection system. By using Adam as an optimizer, thus, the accuracy of detection and we obtained 93,45% mAP. We note that when using ADAM as a

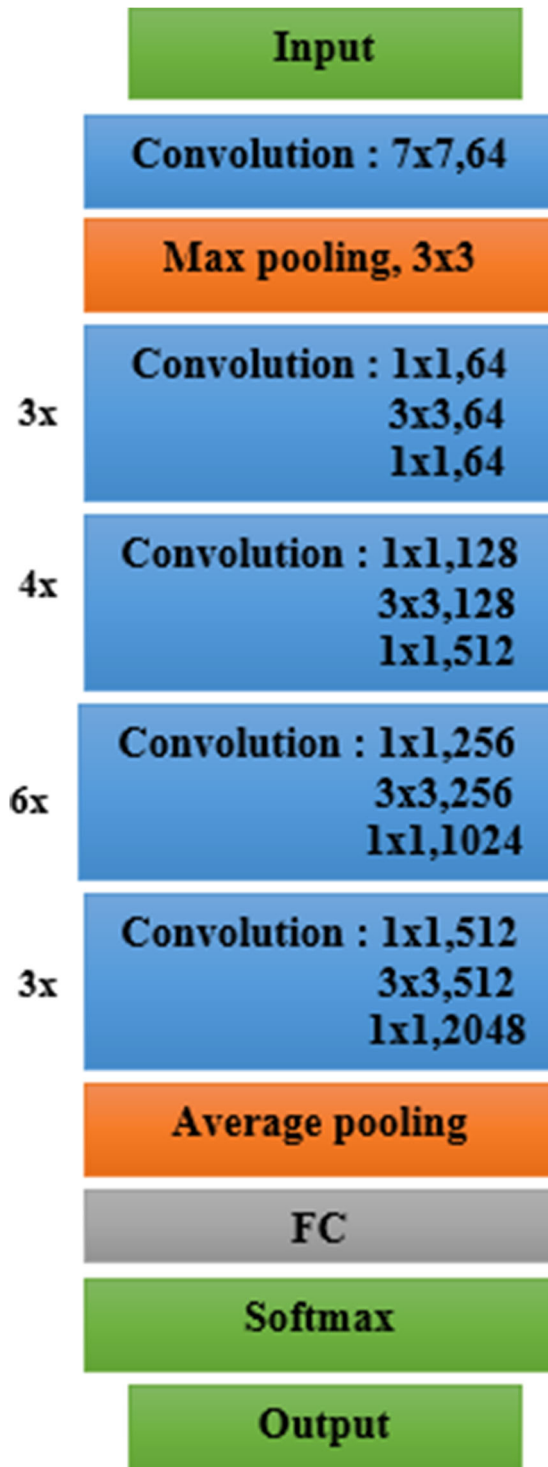


Fig. 8 ResNet 50 architecture

Table 1 Experiments setup

Epochs	50
Steps	10,000
Learning rate	$1 \cdot e^{-5}$
Min-image rescale	800
Max-image rescale	1333
Optimizer	ADAM/SGD
Batch size	16
Regression loss	Smooth-l1
Classification loss	Focal loss
Training images	2600
Testing images	1400

Table 2 Results obtained using stochastic gradient decent optimizer (SGD)

Indoor sign name	Accuracy (% mAP)
Exit	92.93%
Wc	92.77%
Disabled exit	92.83%
Confidence zone	91.24%
Mean	92.44%

network and a loss function optimizer, we succeeded to increase the detection performances for the four indoor signs.

In order to acquire more information about the results obtained with the proposed indoor signage detection system, we calculated true positive (TP), false positive (FP) and false negative (FN) in order to analyze different metrics as precision, recall and F1-score. Table 4 presents all the experiment metrics used in our work.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F1\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

Table 3 Results obtained using Adam optimizer

Indoor sign name	Accuracy (% mAP)
exit	93.52%
wc	93.48%
Disabled exit	93.83%
Confidence zone	92.97%
mean	93.45%

Table 4 Evaluation metrics

Precision	90.82%
True Positive	9250
False Positive	813
False Negative	934
F1-score	91.56%
Recall	90.08%
Accuracy	93.45%

Table 5 Comparison results

Indoor sign name	Ours (%)	[28] method (%)	[2] method (%)	[3] method (%)
exit	93.52	//	63.79	80.33
Disabled exit	93.83	90	//	//

As presented in Table 5, our proposed indoor signage detection system achieves better results compared to the results obtained in [2, 3, 28] for exit and disabled exit classes. We achieved higher detection accuracies for the two classes. We have outperformed the state-of-the-art results for exit class by obtaining 93.52% compared to 63.79% obtained in [2] and 80.33% obtained in [3]. We have also improved the precision detection for disabled exit class by obtaining 93.83% as a detection accuracy compared to 90% obtained in [28]. The following Fig. 9 present a detection example. As presented in this figure the indoor sign is detected accurately and with a high detection rate. We note that the proposed wayfinding assistance system is able to detect a set of four landmark indoor signs however its size, color, texture and its view point.

As far as the indoor mobility of blind and impaired persons is addressed, the proposed indoor wayfinding assistance navigation should provide the user information about the up-

**Fig. 9** Detection example

coming indoor sign. In order to ensure an efficient indoor signage detection system, the developed system must present a contribution between the detection precision as well as the detection speed. We note that the proposed indoor wayfinding assistance system achieves 40 ms per image which corresponds to 25 FPS which matches the blind and the impaired person's mobility.

6 Conclusion

By using the newest technologies of deep learning, we proposed an efficient system for wayfinding assistance navigation for both blind and sighted persons in indoor environments. Unlike, common technologies present in literature which are limited to recognize unique category. In this paper, we have defined a new indoor signage detection system. It presents the first work using a one stage-based deep CNN and that enables the blind or the visually impaired person to recognize a set of indoor signs present in their indoor surroundings. To train and test our proposed indoor object detection system, we make use of RGB images presenting very challenging conditions as various illumination conditions, complex background and textured objects. We conducted our proposed experiments on our proposed challenging indoor signage dataset. The results showed that our work outperforms existing methods. It has achieved 93.45% as precision of detection.

As a future work, we will focus on implementing and adopting the proposed indoor sign detection system on an embedded system with limited resources. We will also extend our work for recognizing and locating more indoor objects.

References

1. Afif M, Ayachi R, Said Y, et al. (2018) Indoor image recognition and classification via deep convolutional neural network. In : International conference on the Sciences of Electronics, Technologies of Information and Telecommunications. Springer, Cham, p. 364–371
2. Afif M, Ayachi R, Pissaloux E et al (2020) Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people. *Multimed Tools Appl*:1–18
3. Afif M, Ayachi R, Said Y et al (2020) An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation. *Neural Process Lett*:1–15
4. Afif M, Ayachi R, Said Y, et al. (2020) Deep Learning Based Application for Indoor Scene Recognition. *Neural Process Lett*, p. 1–11
5. Ali A, Zhu Y, Chen Q, Yu J, Cai H (2019) Leveraging Spatio-Temporal Patterns for Predicting Citywide Traffic Crowd Flows Using Deep Hybrid Neural Networks. 2019 IEEE 25th international conference on parallel and distributed systems (ICPADS), Tianjin, China:125–132. <https://doi.org/10.1109/ICPADS47876.2019.00025>
6. Ali A, Zhu Y, Zakarya M (2021) A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-020-10486-4>
7. Ayachi R, Afif M, Said Y et al (2020) Traffic signs detection for real-world application of an advanced driving assisting system using deep learning. *Neural Process Lett* 51(1):837–851
8. Ayachi R, Said Y, Abdelaali AB (2020) Pedestrian Detection Based on Light-Weighted Separable Convolution for Advanced Driver Assistance Systems. *Neural Process Lett*, 1–14
9. Bashiri FS, LaRose E, Badger JC, D'Souza RM, Yu Z, Peissig P (2018) Object detection to assist visually impaired people: A deep neural network adventure. In *International Symposium on Visual Computing*; Springer: Cham, Switzerland; pp. 500–510

10. Chen Z, Cai H, Zhang Y et al (2019) A novel sparse representation model for pedestrian abnormal trajectory understanding. *Expert Syst Appl* 138:112753
11. Chen Z, Chen D, Zhang Y et al (2020) Deep learning for autonomous ship-oriented small ship detection. *Saf Sci* 130:104812
12. Epelea L, Gavrilu TI, Gacsádi A (2017) Smartphone application to assist visually impaired people. In *Proceedings of the 2017 14th International Conference on Engineering of Modern Electric Systems (EMES)*, Oradea, Romania, pp. 228–231
13. Fei Z, Yang E, Hu H, Zhou H (2017) Review of machine vision-based electronic travel aids. In *Proceedings of the 2017 23rd International Conference on Automation and Computing (ICAC)*, Huddersfield, UK, 7–8, pp. 1–7
14. Fusco G, Cheraghi SA, Neat L et al (2020) An Indoor Navigation App using Computer Vision and Sign Recognition. In: *International Conference on Computers Helping People with Special Needs*. Springer, pp 485–494
15. He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 770–778
16. Kanwal N, Bostanci E, Currie K, Clark AF (2015) A navigation system for the visually impaired: a fusion of vision and depth sensor. *Appl Bionics Biomech* 2015:479857
17. Kingma DP (2014) et BA, J. L. Adam: A method for stochastic optimization. arXiv 2014. arXiv preprint arXiv:1412.6980
18. Kuntho J, Karkar AG, Al-maadeed S et al (2020) Indoor positioning and wayfinding systems: a survey. *Human-centric Computing and Information Sciences* 10:1–41
19. Lin B, Lee C, Chiang P (2017) Simple smartphone-based guiding system for visually impaired people. *Sensors* 17:1371
20. Lin T, Goyal P, Girshick R, et al (2017) Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. p. 2980–2988
21. Mekhalfi ML, Melgani F, Zeggada A, De Natale FG, Salem MAM, Khamis A (2016) Recovering the sight to blind people in indoor environments with smart technologies. *Expert Syst Appl* 46:129–138
22. Rituerto A, Fusco G, Coughlan JM (2016) Towards a sign-based indoor navigation system for people with visual impairments. In: *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. p. 287–288
23. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747
24. Sivan S, Darsan G (2016) Computer vision based assistive technology for blind and visually impaired people. In *Proceedings of the International Conference on Computing Communication and Networking Technologies*, Dallas, TX, USA, 6–8
25. Tian Y (2014) RGB-D sensor-based computer vision assistive technology for visually impaired persons. In *Computer Vision and Machine Learning with RGB-D Sensors*; Shao, L., Han, J., Kohli, P., Zhang, Z., Eds.; Springer: Cham, Switzerland, 2014; pp. 173–194.
26. Tian YL, Chucai YI, et al (2010) ARDITI, Aries. Improving computer vision-based indoor wayfinding for blind persons with context information. In: *International Conference on Computers for Handicapped Persons*. Springer, Berlin, Heidelberg, p. 255–262
27. Trabelsi R, Jabri I, Melgani F, Smach F, Conci N, Bouallegue A (2019) Indoor object recognition in rgb-d images with complex-valued neural networks for visually-impaired people. *Neurocomputing* 330:94–103
28. Wang S, Yang X, Tian Y (2013) Detecting signage and doors for blind navigation and wayfinding. *Network Modeling Analysis in Health Informatics and Bioinformatics* 2(2):81–93
29. WHO (n.d.) Available online: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
30. Yang K, Wang K, Hu W, Bai J (1954) Expanding the detection of traversable area with real sense for the visually impaired. *Sensors* 2016:16
31. Ye C, Qian X (2018) 3-D object recognition of a robotic navigation aid for the visually impaired. *IEEE Trans Neural Syst Rehabil Eng* 26:441–450
32. Ye C, Hong S, Qian X, Wu W (2016) Co-robotic cane: a new robotic navigation aid for the visually impaired. *IEEE Syst Man Cybern Mag* 2:33–42