SN Computer Science (2021) 2:424
https://doi.org/10.1007/s42979-021-00790-7

ORIGINAL RESEARCH

A Transfer Learning Approach for Indoor Object Identification

Mouna Afif[1] · **Riadh Ayachi**[1] · **Yahia Said**[1,2] · **Mohamed Atri**[3]

Received: 9 November 2020 / Accepted: 20 July 2021 / Published online: 23 August 2021

Abstract
Accessing new indoor environments is a well-known challenge for blind and visually impaired persons (VIP). In this work, we propose a new computer vision-based indoor object recognition system used especially for indoor wayfinding and indoor assistance navigation. The proposed recognition system is based on transfer learning techniques. This system is able to detect with a big performance three categories of indoor classes (door, stairs and sign). We developed an efficient and a robust indoor landmark identification system based on a lightweight deep convolutional neural network (DCNN). The proposed detection system is generic and performant enough to handle the large intra-class variation provided in the training and the testing sets. Experimental results have shown the big efficiency of the obtained systems by achieving high recognition rates.

Keywords Indoor object recognition · Blind and visually impaired people (VIP) · Deep convolutional neural network (DCNN) · Deep learning

Introduction

Blindness and visual impairments present a serious problem affecting a huge number of persons around the world. In 2018, the World Health Organization (WHO) has estimated that 1.3 billion persons suffer from visual impairments of whom 36 million are totally blind [1].

Robust and efficient indoor object identification system can widely help blind and persons with serve vision impairments to independently access new unfamiliar indoor environments. Computer vision-based systems for indoor object detection present a challenging problem due to many factors, such as (1) the large intra-class variation of appearance of objects of the same class categories and (2) high complexity of indoor images background and the high illumination variation.

To improve the life quality of blind and sighted person, we developed the proposed system to improve their ability to independently access, explore and navigate in new indoor environments.

We propose a new indoor object recognition system based on transfer learning techniques of deep convolutional neural networks.

Recognizing objects and estimating their categories and poses present a very powerful task in a wide range of applications including robotics [2], image segmentation [3], indoor object detection [4–6] and road sign detection [7]. This task become more challenging especially in indoor environments as they present cluttered spaces and very complex decorations. A very popular approach to tackle this problem is to employ deep learning techniques for indoor object recognition. However, to train and test the deep learning-based object detection and recognition systems requires a huge amount of annotated data.

In this study, we propose to develop a new indoor object recognition system based on transfer learning approach of deep convolutional neural networks (DCNN). Our proposed system can detect efficiently three types of indoor landmark objects classes (door, sign and stairs).

Robust object detection and recognition present a fundamental task including various fundamental aspects as the robot manipulation, human–robot interaction and augmented reality. However, various lighting conditions, heavy

✉ 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

occlusion, cluttered decorations remain the problem more challenging. Furthermore, indoor objects may appear with different scales, forms and sizes depending on the camera pose and calibration. Indoor object identification can widely help blind and sighted persons to more explore new environments and to more integrate in the daily life. For this fact, an accurate indoor object recognition system is crucial for practical interaction of the blind or the impaired person with the real-world indoor objects.

The performance of our proposed work was evaluated and benchmarked using the indoor object classification data set MCIndoor 20,000 [8].

Nowadays, life present a huge amount of visual data every minute split. This visual data should be understood and analyzed. The huge growth on visual data led to new challenges for the scientific community to facilitate daily life especially for blind and persons with serve impairments.

The visual system of human provides a great ability to extract important features and information from images and to find and categorize an immense number of classes and objects. This ability widely facilitates leaning and survival in everyday life. Indoor object recognition is a computer vision task dealing with recognizing and identifying instances of objects of specific class categories in input images or videos. This task has attracted a lot of researcher's attention, especially in the last few years. This strong interest is explained by the big importance of this task and also by the phenomenal advances provided since the arrival of deep learning methods.

To automatically recognizing and locate specific set of objects in input images and video present a very important task for computers to understand the surrounding environments. The paradigm of indoor object detection is one of the most tasks for blind and sighted persons to interact, work, communicate and survive. Solving the problem of indoor object recognition with all its challenges present a major component to solve the problem for blind and sighted persons.

Indoor object detection and recognition is one of the various computer vision tasks related to the inference and the recognition of the high-level information from videos and images.

In this paper, we propose a deep convolutional neural network-based method for indoor objects identification. This method is able to detect a specific set of indoor objects (sign, door, and stairs). This method is highly recommended for blind and sighted persons to fully interact with their surrounding environments.

Our proposed indoor object recognition system is based on a set of lightweight DCNN networks. We improved the performance of the proposed work by training the DCNN on challenging training data including cluttered decorations, different lighting conditions, different forms and sizes of object, high intra-class variation.

Blind and impaired persons facing many problems during their daily activities especially when accessing new indoor environments. The proposed work will be highly recommended for blind and sighted persons to navigate and to interact with objects around them.

Recently, a huge development was known on the computer vision and the artificial intelligence field for objects decorations based on deep learning algorithms and especially on DCNN networks. Deep learning methods present a family of machine learning algorithms the learning process can be supervised or unsupervised depending on the deep learning algorithm architectures. Deep CNN have achieved impressive results on many computer vision applications as image classification [9], image segmentation [10] and object detection [11].

To develop the proposed indoor object recognition system, we used transfer learning techniques. Therefore, the main contribution of our work can be summarized as follows:

- A new deep CNN solution for blind and sighted persons to recognize their surrounding environments.
- We used transfer learning techniques based on lightweight DCNN architectures.
- The proposed work achieved high recognition rates.
- The proposed work aims to recognize and identify specific set of indoor landmark objects from a large number of predefined class categories in natural images.

Related Work

The vision is the most arguably sense that human have to safely moving and interacting with the world. Object detection is a very important task due to its wide range of applications. Object detection and recognition in real scene present one of the most challenging computer-vision tasks as it deals with many difficulties, such as illumination variation, occlusion, view point changes and background complexity.

Many researchers associated a huge interest on indoor object detection tasks. Many classical works were proposed to solve the problem of indoor object recognition [12, 13]. Many of these works are based on designing statically models to understand the indoor scenery geometry [14], scene text detection [36], logo detection [37] and text detection [38].

Various classical works were proposed to solve the problem of indoor object recognition mainly based on machine learning techniques [13, 15]. However, these types of algorithms are based on a complex pipeline design which require

a huge computational resource. Many tasks rely mainly on indoor object recognition as the robotic navigation [16, 17].

The increasing interest on building new indoor object detection and recognition systems makes this task more challenging. When moving around the surrounding environments, human being use its sight more than its other senses to distinguish objects from their colors, shapes, sizes, positions and textures. In [18], authors proposed a new vision system used to detect objects in usual human environments. The proposed system is builded based on Support vector Machine (SVM) and the input of the system are RGB-D images.

Deep neural networks exhibit a big difference from traditional approaches as they provide very deep architectures and provide more complex tasks. In [19], authors propose an object detection system based on deep neural networks. Authors used the PascalVOC data set to evaluate the performance of their work. The indoor object recognition and classifications tasks can widely serve for indoor scene classification [20].

Recently, by employing deep neural networks-based algorithm, indoor object classification and recognition tasks are considerably improved. Using deep learning and especially deep convolutional neural networks-based architectures, computer vision tasks have been known a great improvements and achievements. Depp learning-based technique have employed for different applications including Alzheimer disease classification [30], gesture recognition [31], rice disease detection [32], facial expression recognition [33], places classification [34] and medical imaging [35].

Object recognition present a main task in various computer vision applications including medicine, augmented reality, robotics assistance systems and autonomous driving assistance systems. In [22], authors proposed an overview of the recent advances in 3D object recognition and its implementations using deep leaning and convolutional neural networks techniques. The focus of authors work was based on an industrial environment including different lighting conditions, occlusion, and incomplete data sets.

Deep learning algorithms have presented an immense success and presented a powerful component to treat challenging problems and have led to breakthroughs in artificial intelligence field. The goal of our work is to provide a comprehensive idea about the surrounding indoor environment for sighted and blind persons.

Building an accurate indoor object classification is very essential for blind and sighted persons to explore and navigate indoor environments. To towards help this category of persons, we propose in this work building new indoor object classification system.

The best choice to perform indoor object detection approach in computer vision is to exploit deep learning and especially deep convolutional neural networks architectures. However, training such algorithms requires huge amount of annotated data. In this work, we explore lightweight DCNN models to develop new indoor object identification application. To evaluate our work, we used two different lightweight DCNN architectures to develop a new application for indoor object classification.

The reminder of the rest of this paper is the following, "Proposed Approach for Indoor Object Recognition" overviews the proposed deep learning-based approach for indoor object classification. "Experiments and Results" details the experimental results and "Conclusion" concludes the paper. In Fig. 1, we presents the overall pipeline used in the proposed work.

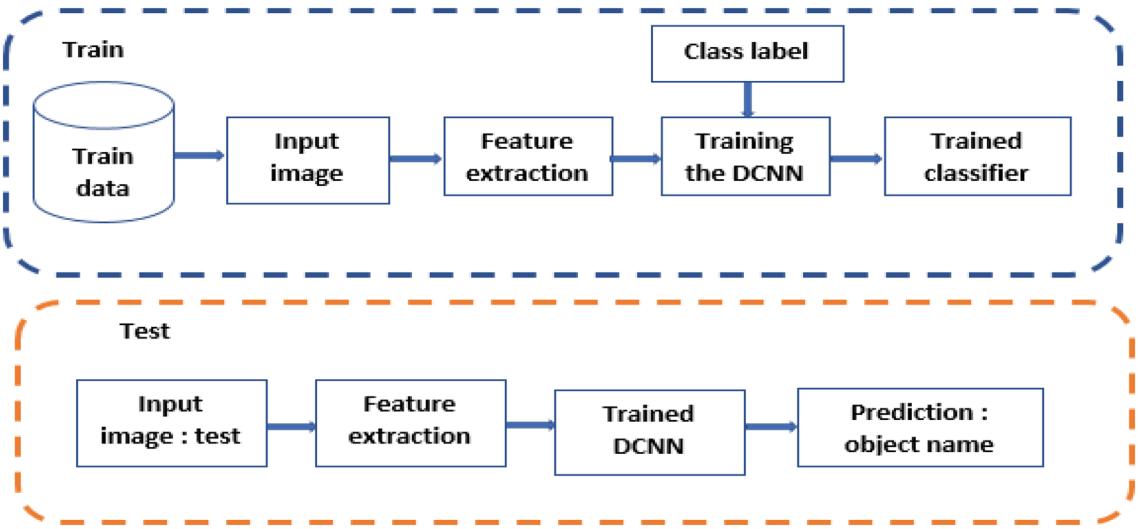

Fig. 1 Overall pipeline for indoor object recognition

Proposed Approach for Indoor Object Recognition

Training a DCNN require a huge amount of annotated data to train the huge number of parameters. The training and testing data cover various parameters, such as: different lighting conditions, various objects viewpoints, heavy occlusion, and complex indoor decorations. Our approach focusses on detecting and recognizing indoor landmark objects to develop new application for blind and sighted persons to fully participate and to explore their surroundings. During our experiments, we employ two set of state-of-the-art DCNN models: MobileNet v1 and v2 [23, 24] and inception v3 [25].

DCNNs have become essential and ubiquitous in computer vision and artificial intelligence field. In this paper, we make use of the efficient classification model MobileNet [23, 24]. This architecture is highly recommended for mobile-vision applications as they present a reduced number of parameters compared to other architectures. This neural network architecture provides a set of hyper-parameters which are highly matched to the implementation's requirements of mobile devices.

In this section, we will detail the proposed indoor architecture based on the lightweight DCNN architecture of MobileNet. The mobileNet classification-based architecture is especially designed for lightweight mobile implementations while ensuring powerful performances. MobileNet architecture introduce very powerful blocks called "depthwise separable convolutions". These blocks are responsible for the main idea besides the MobileNet lightweight model size. The depthwise separable convolution block is composed of two main components: the first part is composed of the depthwise convolution layer followed by the batch normalization and RELU6 [21] layers. The second part is the pointwise convolution followed by batch normalization and RELU 6 [21] layers. Figure 2 provides the depthwise separable convolution block used in MobileNet architecture.

The depthwise separable convolution are the key building block for the key efficiency of MobileNet architectures. The kernel of the depthwise separable convolution is applied to all the input image channels. Depthwise separable convolution present a factorized convolution which factorize the standard convolution to 1×1 convolution named pointwise convolution. The depthwise convolution applies one kernel filter to each input channel (3 channels). While the pointwise convolution applies 1×1 kernel filter and its output will be combined with the depthwise output. The output of the depthwise convolution is only one single channel. Figure 3 presents the difference between regular convolution, depthwise convolution and pointwise convolution.

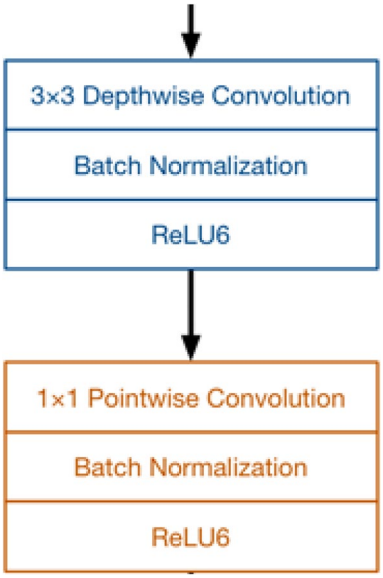

Fig. 2 Depthwise separable convolution block architecture

Both, the standard and depthwise convolution apply filters and kernel to create new features while the standard convolution requires more computational resources while the depthwise convolution require less computational resources and faster than the regular convolution. MobileNet v2 [24] present new powerful blocks than version 1. This updated architecture introduces three new powerful block layers:

- Linear bottleneck layers
- Shortcut connection
- Inverted residual blocks

Basically, the bottleneck layer contains fewer parameters nodes and parameters than its previous layer. In general, this type of layers is especially used to reduce the dimension of the input image. The MobileNet v2 architecture contain a set of activation layers which form a "manifold of interest". The value of this manifold is different from 0 after RELU 6 transformation. This operation corresponds to the linear transformation. Another key layer provided in the updated architecture is the inverted residual block. The inverted residual layer enables the feature map to e encoded in low-dimensional sub-space. The bottleneck layer appears very similar as the residual block where each block contains the input representation followed by the explanation layer. As bottleneck layers contains the necessary information and acts as an implementation of non-linear transformation, for this fact it introduces the shortcut connection directly between bottleneck layers. Table 1 provides the different layers used in mobileNet v2 architecture.

To summarize, the MobileNet v2 architecture is composed of a regular convolution followed by 11 bottleneck

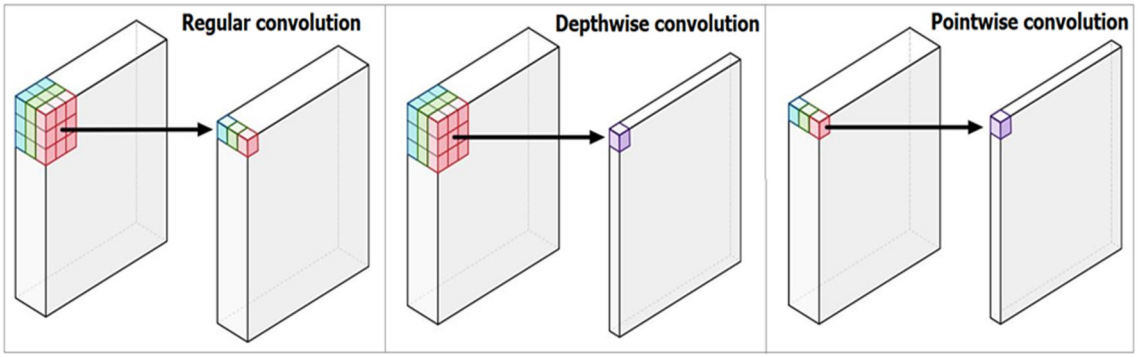

Fig. 3 Difference between regular, depthwise and pointwise convolution

Table 1 MobileNet v2 architecture

Input	Layer	*n*
224×224x3	Conv 2D	1
112×112x32	Bottleneck	1
112×112x16	Bottleneck	2
56×56x24	Bottleneck	2
28×28x32	Bottleneck	2
14×14x64	Bottleneck	1
14×14x96	Bottleneck	2
7×7x160	Bottleneck	1
7×7x320	Conv 2D 1×1	1
7×7x1280	Average pooling 7×7	–
1×1x1280	Conv 2D 1×1	–

layers and a pointwise convolution and an average pooling layer, another pointwise convolution layer then it ends with a fully connected layer and a softmax layer used to classify the objects categories.

To more and deeper evaluate our proposed indoor object recognition system, we used another set of deep learning algorithm named inception model [25, 26].

Inception family [26] present an important step on building new DCNN classifiers. In this paper, we make use of inception v3 [25] version. This third version provides 42 powerful layers. This updated version requires fewer computational resources with less parameters than the previous versions. This version makes the lower error rate for the image classification challenge in ILSVRC [27]. Y reducing the number of parameters of the neural network, we ensure a mobile implementation of the proposed work.

To more reduce the parameters of the DCNN, in this pretrained architecture a 5×5 convolution is replaced by a 3×3 convolution. When using a 5×5 convolution, the number of parameters is set to 5×5=25, while when using two layers of 3×3 (2×3×3=18) parameters. Therefore, this technique reduces considerably the number of parameters and the network complexity. The 5×5 convolution layer is replaces y 2×3×3 contributes to a very powerful block named "inception module-A".

Another operation is presented in the updated version is that a 3x" convolution is replaced by 2×1×3 convolution. This operation contributes to the second powerful block named "inception module-B". Inception v3 architecture provides also the "inception module-C". All these operations highly contribute to reduce the parameters number of the neural network and to minimize the risk of the DCNN overfitting.

One auxiliary classifier is presented in the inception v3 architecture on the top of the last 17 layers. To minimize the feature map dimension, we make use of max pooling layers. For this fact, a grid size reduction block is proposed in the updated version of inception family. All the proposed work has been obtained by applying the transfer learning technique. This technique is a machine learning technique was the model is reused as a starting point for a second task. This approach is very effective as the data is trained on large scale data sets and reused as a second starting point for a second task totally different from the first task.

Experiments and Results

In this section, we will present all our experiments conducted in this work. As we mentioned before, the proposed work is especially dedicated for blind and sighted persons to more participate in the daily life.

To train and test our proposed indoor object recognition system, we used MCIndoor 20,000 [8] data set. The MCIndoor 20,000 data set consists of 20,000 indoor images containing three landmark objects (door, sign and stairs). To develop our indoor objects identification system, we make use of various DCNNs. To obtain more efficiency and robustness for the proposed work, we evaluated our work on MobileNet v1, v2 and inception v3 neural networks to

perform this work. Images provided in the MCIndoor data set are isolated form their surrounding environments which makes it a suitable data set for indoor object classification tasks. Indoor objects application proposed in this paper are crucial and landmark for indoor navigation system.

1700 images of the MCIndoor do not undergo any modifications or quality enhancements while the rest of the data set have undergone various modifications as applying salt and pepper filters, rotation, blurring and translation.

Table 2 provides the experiment settings provided on our proposed work.

The proposed indoor object recognition system is builded based on three powerful DCNNs models. To improve the effectiveness and the robustness of this work, we trained the proposed system using various versions of MobileNet architecture (MobileNet v1, v2) and inception v3. Training a neural network require huge amount of data to avoid the overfitting risks. We apply data augmentation to provide the neural network with much training data. Figure 4 presents the proposed methodology used in our work for indoor objects recognition.

The proposed work is highly relevant for blind and sighted persons to improve their daily life quality. The proposed method was implemented using python language and TensorFlow framework [28]. When training the DCNNs we make sure of using challenging images with challenging conditions to improve the system inference performances.

The proposed experiments were performed on desktop with an Intel Xeon ES-2683 v4 processor and equipped with a Quadro M4000 graphic processor with 8 GB of graphic memory.

The network implementation was performed using the deep learning-based framework TensorFlow [28]. In this work, we used transfer learning techniques to develop the proposed indoor object calcification application. We believe that these indoor objects (door, stairs, sign) are vital for blind and sighted persons to more explore their surroundings.it is extremely simple for a normal person to recognize objects and environments but, it is extremely hard for impaired persons to do such task. For this fact, we propose this work.

Our aim from this work is to train and test various architectures of neural networks to acquire deep knowledge to obtain new application with high perception capacities close to capacities of a normal person.

Avery known technique used in this work is the transfer learning. This technique is based on transferring knowledge learned for one task and transferred for a second task. It reuses weights of the first task and freezes specific layers and retrain the last layers on the new data set for the new task. As we treat a complex application for specific category of person, we used transfer learning technique to stimulate the recognition process.

The testing step were performed on the same workstation computer with the same configurations as the training process. As an optimizer of the neural networks, we used the ADAM optimizer [29]. This optimizer updates in a very fast way the network parameters. The next table

Table 2 Experiment settings

Training steps	10,000
Learning rate	0.01
Validation set	30%
Testing set	30%
Training batch size	100
Validation batch size	100

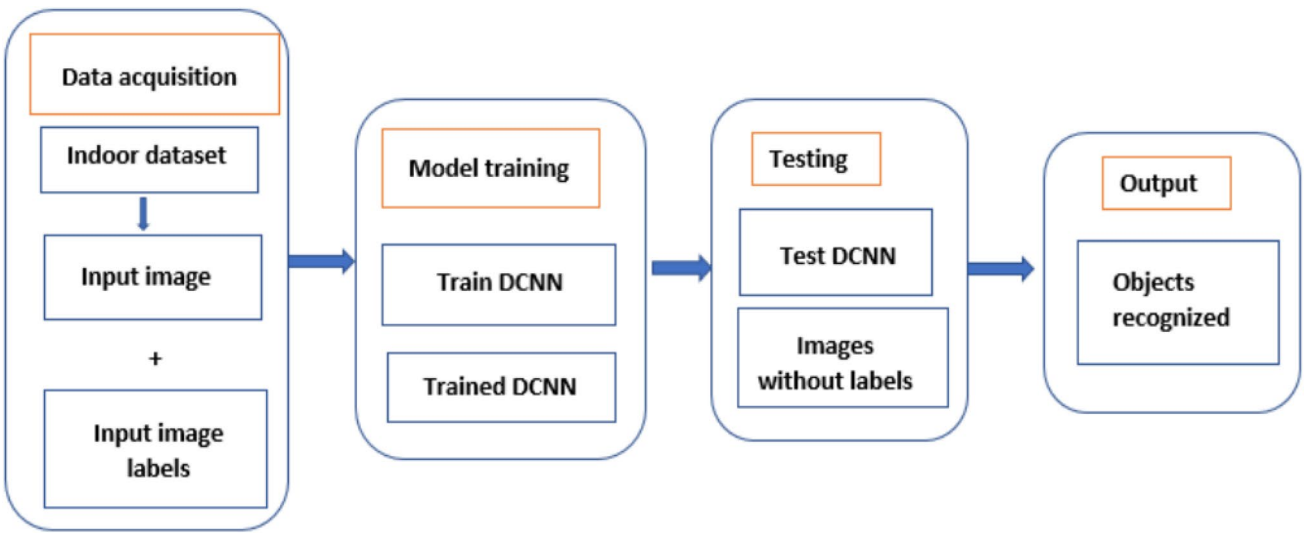

Fig. 4 Proposed methodology used for our work

Table 3 Comparison with the state-of-the art work on MCIndoor's original images

Approach (MCIndoor: original images)	Accuracy (%)
Method in [6]	93.7
Bashiri et al. method [8]	90.4
Ours (MobileNet v1)	99.6
Ours (MobileNet v2)	99.9
Ours (Inception v3)	99.7

Table 4 Per-class results when using MobileNet v1

Class name	Accuracy (%)
Door	99.7
Stairs	99.8
Sign	99.7
Mean	99.7

Table 5 Per-class results when using MobileNet v2

Class name	Accuracy (%)
Door	99.9
Stairs	99.9
Sign	99.9
Mean	99.9

provides a detailed comparison between inference rates obtained with state-of-the-art models. Table 3 provides a comparatives study with the state-of-the-art works.

As mentioned in Table 3, our proposed work outperforms our previous work made in [6] and the work in [8] in term of recognition rate. We obtained very encouraging results when using MobileNet family and inception v3 architectures. We note that results obtained when using MobileNet v2 outperform all the other architectures used. Table 4 provides the recognition rates per-class obtained in our experiments when using MobileNet v1 as a neural network.

As presented in Table 4, our proposed indoor landmark object classification system achieves very encouraging results coming up to 99.7 as inference classification accuracy when using MobileNet v1 as a neural network. Table 5 provides all the per-class results obtained when using MobileNet v2 as a DCNN.

Results conducted in Table 5 shows the interesting recognition rates of our proposed work on recognizing the three landmark objects proposed in the MCIndoor data set. We achieved 99.9% as a recognition rate when using MobileNet v2 architectures when trained and tested on the original images of the MCIndoor data set (1700 images). Table 6 reports the per-class experiments results obtained when using Inception v3 as a neural network.

Table 6 Per-class results when using Inception v3

Class name	Accuracy (%)
Door	99.8
Stairs	99.9
Sign	99.9
Mean	99.8

Table 7 Comparison with the state-of-the-art work on MCIndoor data set

Approach (MCIndoor: all the data set)	Accuracy (%)
Bashiri et al. method [8]	90.4
Ours (MobileNet v1)	99.7
Ours (MobileNet v2)	100
Ours (Inception v3)	99.8

As presented in Table 6, we obtained good results when using the Inception v3 architecture as a neural network for our proposed object classification system. Table 7 reports all the results obtained when training and testing the three DCNN models on the 20,000 images of MCIndoor data set.

The results mentioned in Table 7 shows the originality and the big efficiency of our proposed work. We achieved higher recognition rates then Bashiri et al. method [8]. We also note that the higher recognition rates obtained in our work are achieved when using MobileNet v2 architecture. We obtained almost 100% recognition rate. We also obtained very good results coming up to 99.7% and 99.8% when using MobileNet v1 and inception v3 architectures, respectively.

Conclusion

Indoor object recognition and classification present a crucial task in artificial intelligence to help blind and sighted persons during their daily activities.

In this paper, we propose a new indoor object identification system based on using three state-of-the-art neural networks architectures. We used transfer learning techniques using MobileNet v1, v2 and inception v3 DCNNs. The proposed application was evaluated on the MCIndoor 20,000 data set. Our work achieves high recognition rates. The proposed work provides more interest on using deep learning-based techniques to towards help and improve the daily life for blind and sighted persons y providing a comprehensive idea of their surrounding environments and to more interact with them. A powerful future work aims to develop a new indoor objects detection system used for indoor assistance navigation for blind and visually impaired persons and to be implemented on low-end devices.

Declarations

Conflict of Interest The authors declare that there is no conflict of interest.

References

1. World Health Organization. Vision Impairment and Blindness. Available online: https://www.who.int/newsroom/fact-sheets/detail/blindness-and-visual-impairment. Accessed 19 Sep 2019.
2. Martinez-Martin E, et Del Pobil AP. Object detection and recognition for assistive robots: experimentation and implementation. IEEE Robot Autom Mag. 2017;24(3):123–38.
3. Wang L, Shi J, Song G, et al. Object detection combining recognition and segmentation. In: Asian conference on computer vision. Berlin: Springer; 2007. p. 189–99.
4. Afif M, Ayachi R, Said Y, et al. An evaluation of RetinaNet on indoor object detection for blind and visually impaired persons assistance navigation. Neural Process Lett. 2020. https://doi.org/10.1007/s11063-020-10197-9.
5. Afif M, Ayachi R, Said Y et al. Indoor object classification for autonomous navigation assistance based on deep CNN model. In: 2019 IEEE international symposium on measurements & networking (M&N). IEEE; 2019. p. 1–4.
6. Afif M, Ayachi R, Said Y et al. 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; 2018. p. 364–371.
7. Ayachi R, Afif M, Said Y, et al. Traffic signs detection for real-world application of an advanced driving assisting system using deep learning. Neural Process Lett. 2020;51(1):837–51.
8. Bashiri FS, Larose E, Peissig P, et al. MCIndoor20000: a fully-labeled image dataset to advance indoor objects detection. Data Brief. 2018;17:71–5.
9. Sultana F, Suflan A, et Dutta P. Advancements in image classification using convolutional neural network. In: 2018 Fourth international conference on research in computational intelligence and communication networks (ICRCICN). IEEE; 2018. p. 122–129.
10. Yuheng S, et Hao Y. Image segmentation algorithms overview. arXiv preprint arXiv:1707.02051 (2017).
11. Zhao Z-Q, Zheng P, Xu S-T, et al. Object detection with deep learning: a review. IEEE Trans Neural Networks Learn Syst. 2019;30(11):3212–32.
12. Mei S, Yang H, Yin ZP. Discriminative feature representation for image classification via multimodal multitask deep neural networks. J Electron Imaging. 2017;26(1): 013023.
13. Nan LL, Xie K, Sharf A. A search-classify approach for cluttered indoor scene understanding. ACM Trans Graph. 2012;31(6):1–10 (**Article no. 137**).
14. Wang HY, Gould S, Roller D. Discriminative learning with latent variables for cluttered indoor scene under-standing. Commun ACM. 2013;56(4):92–9.
15. Ranzato M, Huang FJ, Boureau YL, LeCun Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: Computer vision and pattern recognition, 2007. CVPR '07. IEEE Conference on; 2007. p. 1–8.
16. Srinivasa SS, Ferguson D, Helfrich CJ, Berenson D, Collet A, Diankov R, et al. HERB: a home exploring robotic butler. Auton Robot. 2010;28(1):5–20.
17. Ramisa A, Alenyà G, Moreno-Noguer F, Torras C. Learning RGB-D descriptors of garment parts for informed robot grasping. Eng Appl Artif Intell. 2014;35:246–58.
18. Hhernandez AC, Gomez C, Crespo J, et al. Object detection applied to indoor environments for mobile robot navigation. Sensors. 2016;16(8):1180.
19. SzegEedy C, Toshev A, et Erhan D. Deep neural networks for object detection. In: Advances in neural information processing systems. 2013. pp. 2553–2561.
20. Afif M, Ayachi R, Said Y, et al. Deep learning based application for indoor scene recognition. Neural Process Lett. 2020. https://doi.org/10.1007/s11063-020-10231-w.
21. Agarap AF. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375 (2018).
22. Salem B, Stjepandic J, et Stobrawa S. Assessment of methods for industrial indoor object recognition. In: Transdisciplinary engineering for complex socio-technical systems: Proceedings of the 26th ISTE international conference on transdisciplinary engineering, July 30–August 1, 2019. IOS Press; 2019. p. 390.
23. Howard AG, Zhu M, Chen B et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017).
24. Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 4510–4520.
25. Szegzegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 2818–2826.
26. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 1–9.
27. Deng J, Su H, Krause J, et al. Imagenet large scale visual recognition challenge. arXiv preprint arXiv:1409.0575 (2014).
28. Abadi M, Braham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on operating systems design and implementation ({OSDI} 16). 2016. pp. 265–283.
29. Kingma DP, et Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
30. Prakash D, Madusanka N, Bhattacharjee S, et al. A comparative study of Alzheimer's disease classification using multiple transfer learning models. J Multimed Inf Syst. 2019;6(4):209–16.
31. Kim J-H, Hong G-S, Kim B-G, et al. deepGesture: Deep learning-based gesture recognition scheme using motion sensors. Displays. 2018;55:38–45.
32. Sharma V, Mir AA, et Sarwr A. Detection of rice disease using bayes' classifier and minimum distance classifier. J Multimed Inf Syst. 2020;7(1):17–24.
33. Jeong D, Kim B-G, et Dong S-Y. Deep Joint Spatiotemporal Network (DJSTN) for efficient facial expression recognition. Sensors. 2020;20(7):1936.
34. Yeo W-H, Heo Y-J, Choi Y-J, et al. Place classification algorithm based on semantic segmented objects. Appl Sci. 2020;10(24):9069.
35. Fradi M, Afif M, Machhout M. Deep learning based approach for bone diagnosis classification in ultrasonic computed tomographic images. Int J Adv Comput Sci Appl (IJACSA). 2020;11(12). https://doi.org/10.14569/IJACSA.2020.0111210.
36. Keserwani P, Dhankhar A, Saini R, Roy PP. Quadbox: quadrilateral bounding box based scene text detection using vector regression. IEEE Access. 2021;9:36802–18. https://doi.org/10.1109/ACCESS.2021.3063030.
37. Su H, Zhu X, et Gong S. Deep learning logo detection with data expansion by synthesising context. In: 2017 IEEE winter

conference on applications of computer vision (WACV). IEEE; 2017. pp. 530–539.
38. Jabnoun H, Benzarti F, et Amiri H. A new method for text detection and recognition in indoor scene for assisting blind people. In: Ninth International Conference on Machine Vision (ICMV 2016). International Society for Optics and Photonics, 2017. p. 1034123.