Neural Processing Letters
https://doi.org/10.1007/s11063-020-10231-w

Deep Learning Based Application for Indoor Scene Recognition

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

Abstract

Recognizing indoor scene and objects and estimating their poses present a wide range of applications in robotic field. This task becomes more challenging especially in cluttered environments like the indoor scenery. Scaling up convnets presents a key component in achieving better accuracy results of deep convolutional neural networks. In this paper, we make use of the rethinked efficient neural networks and we fine-tune them in order to develop a new application used for indoor object and scene recognition system. This new application will be especially dedicated for blind and visually impaired persons to explore new indoor environments and to fully integrate in daily life. The proposed indoor object and scene recognition system achieves new state-of-the-art results in MIT 67 indoor dataset and in scene 15 dataset. We obtained 95.60% and 97% respectively as a recognition rate.

Keywords Indoor scene recognition · Visually impaired people (VIP) · Deep convolutional neural network (DCNN) · Deep learning

1 Introduction

In our modern life, a huge amount of data and information is created and available every split second. This big growth in visual-information data led to more challenges in processing and extracting the most relevant features that would facilitate various image processing and computer vision tasks. An image presents a matrix of numbers (pixels) for a computer. In order to develop an application for indoor environment perception and to discover all the geometric layout and the semantic clues, we propose in this work to develop an indoor scene recognition system based on deep learning methods. As an example, for a given RGB image the vision-based artificial intelligence agent will be able to recognize and to understand the

✉ 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

Published online: 20 March 2020

complete image layout. For a similar system, it is also required to know the relation-ship between the object and the input scenery. All these concepts are very fundamental for human in wayfinding and in environment interpretation.

Visual scene understanding and recognition present a challenging problem in artificial intelligence and computer-vision community. This problem can be divided into two-main parts and on two main inputs: static input (image) and dynamic input (video). Computer vision and artificial intelligence agents can be deployed to cover a huge variety of applications like autonomous robot navigation [1], autonomous navigation of vehicles [2], image retrieval [3].

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 impairments and 36 are blind [4].

Indoor scene understanding and recognition can widely help blind and visually impaired persons to navigate softly in indoor environments by detecting and avoiding obstacles and analyzing the surrounding environments [5].

In this paper, we will develop a cognitive system used to help blind, sighted and elderly persons to navigate in indoor environments. The number of elderly persons is expected to be 1.5 billion persons by 2050. As much as the number of works in scene recognition and understanding increases, indoor scene recognition presents a highly significant challenging problem due to the complex background of input images, complex indoor decoration, heavy occlusion, different viewpoints, scales and textures changes across different scenery and cluttered indoor environments.

Recent developments in deep learning and deep convolutional neural networks and the huge number of annotated datasets have sparkled a huge interest in addressing this challenging problem.

Image recognition and classification present one of the most fundamental tasks on image and visual contents understanding. Scene analysis and recognition present a very rich task with many challenges. Indoor environments provide very charged images with complex background different intra and inter-class variations. Artificial intelligence-based automated systems become a vital agent to perform very complex computer-vision tasks. Indoor scene understanding and recognition make robots or blind and visually impaired persons better aware of their indoor surroundings. These automated systems make achieving different tasks in a successful and safe way possible.

Scene recognition problem presents an image processing issue aiming to predict a scene category for an input image. In this paper, we propose a new scene recognition application based on deep convolutional neural network.

Existing methods still lack in indoor environments as they present very challenging environments. Indoor environments present rich and disordered decoration features.

Indoor navigation presents a challenging task for blind and sighted persons as well. In order to explore new unfamiliar indoor environments, blind, partially blind and impaired persons require safer assistive systems during their indoor navigation. We note that the developed system is highly recommended to be implemented on various embedded systems. This proposed system will highly help blind and sighted persons to explore new indoor environments in a secure way and they will further integrate in the daily life. The proposed system will accurately recognize and classify indoor scenery images. The system was trained and tested using two indoor scenery benchmarks: MIT 67 indoor dataset [6] and scene 15 dataset [7]. Our indoor scene classification system outperforms the state-of-the-art works in terms of accuracy obtained as it achieves high recognition rate, it was also trained using challenging images in order to increase its robustness. In this paper, we will take advantage of efficientNet

[8] to extract image features from input indoor images and to contribute for a new mobile implementation of indoor scene recognition system. We note that this work presents the first approach evaluating EfficientNet [8] model on indoor scene recognition.

2 Related work

Indoor object detection and recognition present a fundamental aspect in human–robot interaction and augmented reality tasks. However, in cluttered indoor environments the high occlusion between objects, the various lighting conditions and the huge inter and intra-class variation remains the problem more challenging. Furthermore, objects may appear under different textures and forms depending on the camera viewpoint and its calibration. Therefore, developing a robust and an accurate scene understanding algorithm is an essential part for blind and visually impaired persons' practical-interactions in real world indoor scenery. Detecting and recognizing indoor environments in real-world images have attracted a lot of interest in computer vision community. One of the most specifically common problems in computer vision is the perception. Human visual system provides a huge capability to extract the most pertinent features from the surrounding environment. Recognizing indoor environments enables human to focus on just the most important features present in the scene and it contributes in facilitating the learning and the daily life survival.

Classifying and recognizing indoor environments is based on the human visual attention and on its ability to locate and differentiate between sceneries. Recognizing and classifying objects contents of input images can widely help and contributes to many computer vision tasks like: Object detection [9], image segmentation [10], place recognition [11] and traffic sign detection and recognition [12].

Indoor scene recognition task has attracted a lot of researchers' attention. Traditional scene classification method was based especially on hand-crafted features as SIFT (Scale Invariant Features Transform) [13] and SURF (Speed Up Robust Features) [14]. Numerous works have been proposed in scene classification issue. In [15, 16], authors proposed a scene classification method based on the relationship between the scenery and its objects' content (relations between objects). In [17] an indoor scene recognition method is proposed. In [18], authors proposed a new representation and a new modeling method used for indoor affordance areas based on CLM (CodeBookless Model). Experimental results have shown improvements about 20% compared with traditional codebook construction methods.

However, the mentioned methods are low-level features of images without rich semantic information. By using all these methods, it is very hard to get satisfying results on scene recognition task.

Recently, deep learning models have made a huge progress in image classification task [19], object detection [20, 21], face-pose estimation [22], image recognition [23], instance segmentation [24] and dimension reduction [25]. Deep convolutional neural networks provide the best solution for developing scene classification algorithms. Authors in [26] proved that by using transfer learning technique, improved the scene classification accuracy. Deep learning methods provide better results than the classical hand-crafted results. However, deep learning methods present some common problems like: huge amount of data to train models, if it is not the case, overfitting is around the corner.

Scene recognition is one of the most known issues in computer vision area. Whereas, the high progress in scene recognition and object recognition tasks is due to the availability of large datasets like: ImageNet [27], COCO [28], MIT indoor 67 dataset [6] and Scene 15 dataset [7] Sun database [29]. In [30], authors introduced a new scene-centric database called places.

Indoor scene recognition problem is especially based on the indoor object detection task. In our previous works [31–33] we developed applications based on deep learning models for indoor object recognition.

Scene recognition task presents a high-valuable computer vision task as it provides a highly perceptual ability for blind, sighted persons as well indoor domestic mobile robots especially, indoor scene recognition problem just like indoor environments, present a big appearance variability that is explained by the high intra and inter-class variation, the complexity of the background of the indoor scenery and the heavy occlusion. In order to ensure the blind and sighted persons a more comfortable life, we propose in this paper a new indoor scene recognition system based on powerful way for blind and impaired persons to explore new environments and to fully participate in the daily life and to better integrate in society. Our proposed indoor scene recognition system is based on using various versions of efficientNet and the system was evaluated on two benchmark datasets.

The remainder of the rest of this paper is the following.

Section 3 provides an overview of the proposed approach used for indoor scene recognition. Section 4 provides all the experiments conducted and the discussions made for the new indoor scene recognition system. Section 5 concludes the paper.

3 Proposed approach for indoor scene recognition

Human beings have a great ability to recognize and categorize complex scenes with a maximum of accuracy. Since convolutional neural networks achieved remarkable success in scene recognition, we employ a deep CNN architecture to develop a new application used for recognizing indoor scenery. Convolutional neural networks (convnets) are especially developed for a specific task with specific parameters and then scaled up to contribute for better performances.

In this paper, we will take advantage of different versions of efficientNet [8] which present a rethinked convolutional neural network scaling. Model scaling is based on carefully balancing the neural network depth, width and resolution. Network scaling contributes better performances and capacities of the convolutional neural networks.

EfficientNet provides new way to scale networks' dimensions by uniformly scaling all network dimensions of depth/width and resolution by using a highly effective agent named "the compound coefficient". EfficientNet provides a very powerful way to obtain accuracy improvement with more efficiency compared to other CNN models. EfficientNet family presents various versions scaled to different block layers. This deep convolutional neural network is scaled-up in multiple dimensions. Generally, most CNN architecture is scaled up by adding more layers as ResNet family [34]. Contrarily to other models, efficientNet scales-up all dimensions of depth, width and resolutions together. After many experiments and extensive grid search, the compound scaling presents the following coefficient:

$$\text{Depth} = 1.20$$
$$\text{Width} = 1.10$$
$$\text{Resolution} = 1.15$$

EfficientNet architecture presents a very powerful way to obtain a smaller number of parameters and computation requirements which make them the best choice for implementation in

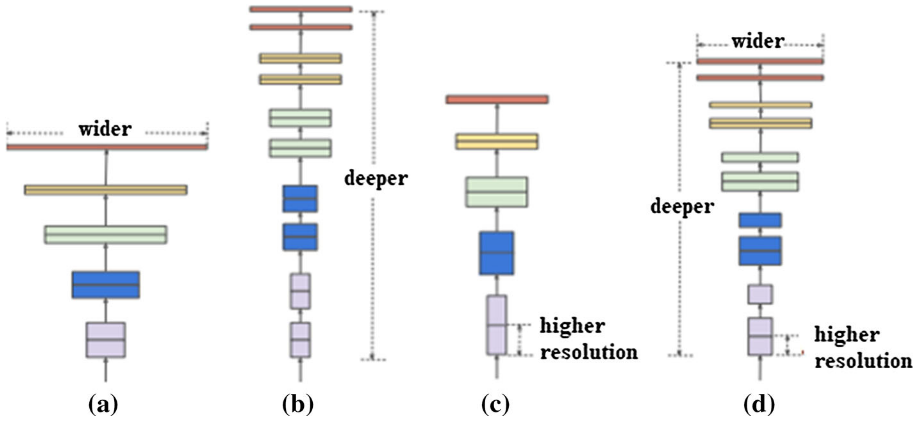

Fig. 1 Different model scaling: **a** width scaling, **b** depth scaling, **c** resolution scaling and **d** compound scaling [8]

mobile platform. The compound scaling method presents a very logical component because the input image is bigger, the network needs more layers to increase the receptive fields and provide more channels to capture more fine-grained features. The different scaling agents are independent. For this reason, for higher resolution images it is important to increase also the network depth, as a larger receptive field can simply captures similar features included in more pixels in bigger images. Correspondingly, it is also important to increase the depth of the network to capture more fine-grained patterns in bigger input image resolutions. This technique participates in highly increasing the neural network accuracy and efficiency. Figure 1 provides examples of different models scaling.

To balance all these scaling parameters, the compound scaling method is proposed in the new efficientNet architecture. Compound scaling method includes a compound coefficient Φ to uniformly scale the network width, depth and resolution by using the new parameters:

$$\begin{aligned}
&\text{Depth} = \text{d} = \alpha^{\Phi} \\
&\text{Width} = \text{w} = \beta^{\Phi} \\
&\text{Resolution} = \text{R} = \gamma^{\Phi} \\
&\text{While}\, \alpha\beta^2\gamma^2 \approx 2 \\
&\text{A} \geq 1, \beta \geq 1, \gamma \geq 1.
\end{aligned}$$

where α, β, γ are constant determined by the grid search.

For indoor navigation, it is very important to recognize objects' positions and orientations in indoor scenery. A very common approach to tackle this problem is by using deep learning methods for object recognition. In this study, the aim of our work is to develop an indoor scene recognition framework.

Indoor scene classification is a very important mission especially for robotics' indoor navigation and for including blind and sighted persons in the daily life.

Our proposed method is divided into two main parts:

1. Train set: features extraction.
2. Test set: scene classification.

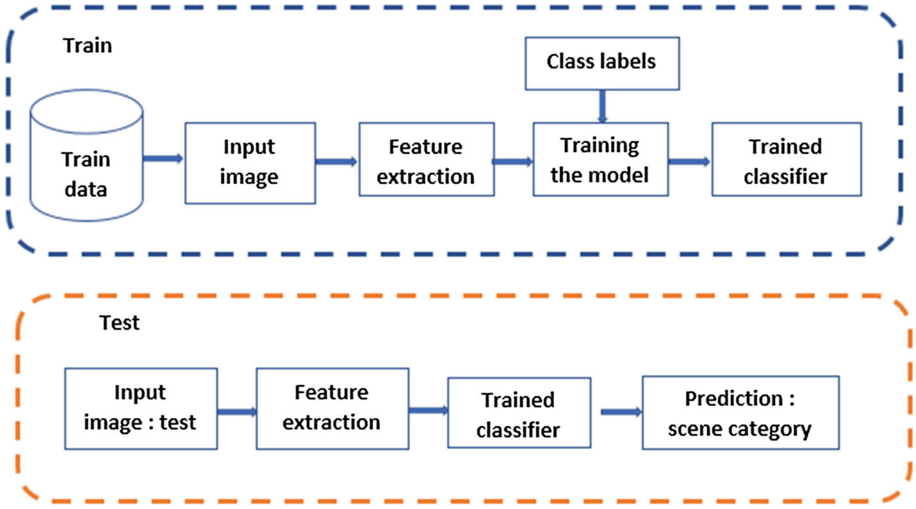

Fig. 2 Proposed pipeline for scene recognition

In order to extract features, we used the convolutional neural network efficientNet. Figure 2 provides the proposed pipeline used in the proposed method for indoor scene recognition. An indoor scene recognition system used for robot service navigation is proposed in this work. The overall framework is detailed in Fig. 2.

Our proposed framework is based on two parts.

The first part is used for training the network and knowledge acquisition. Features extraction is used by fine-tuning a pretrained model named efficientNet. The second part is used for the network test. A new test input scene image is put into the network to generate new features extraction.

These newly-extracted features are matched with the knowledge acquired in the first step and then will be matched with an indoor category feature. Scene classification process consists in calculating diverse scores for each scene category. Then, the scene category is named based on the highest score.

As mentioned in Table 1 EfficientNet [8] architecture is composed of a set of standard convolutions followed by the mobile bottleneck convolution layers (MBConv). The MBConv layers are used to encode features map in low-dimensional subspace in order to reduce the computation complexity. The pooling layers are used to reduce the features map dimensions and the fully connected layers are used for image classification.

Thanks to the huge development introduced since the appearance of deep learning algorithms and fine-tuned convolutional neural network (CNNs) many problems of scene classification were solved. However, by using this method we cannot obtain satisfying indoor scene recognition results because of overfitting problem. To tackle this problem, an indoor scene classification method is proposed in this work. This method, which is based on deep CNNs models is used to extract features from scene and generate the scene category and then classify it. The proposed method demonstrates its robustness and efficiency in recognizing indoor scene. In addition, in our work we eliminated the problem of overfitting as we evaluated our proposed method on two datasets benchmarks MIT 67 indoor dataset [6] and scene 15 dataset [7].

Table 1 EfficientNet-B0 architecture

Layer	Number of layers
Conv 3 × 3	1
MBConv 1, 3 × 3	1
MBConv 6, 3 × 3	2
MBConv 6, 5 × 5	2
MBConv 6, 3 × 3	3
MBConv 6, 5 × 5	3
MBConv 6, 5 × 5	4
MBConv 6, 3 × 3	1
Conv 1 × 1	1
Pooling	1
FC	1

4 Experiments and results

Our proposed indoor scene recognition system was evaluated using two benchmarks datasets: scene 15 dataset [7] and indoor MIT dataset 67 [6]. We selected 5 specific categories of the most common indoor scenery (kitchen, bedroom, bathroom, living room). We selected the most relevant four indoor scene categories that can widely participate in improving the daily life of blind and sighted persons as well. Our study relies on developing on indoor scene assistance navigation system for robot service and for blind and sighted persons. The proposed method was developed using python and TensorFlow framework.

EfficientNet [8] architecture provide the mobile inverted bottleneck convolution layers. By using this type of layers, the DCNN became deeper. During the training process, we used cross entropy as a loss function. It generally calculates the difference between the probability of the indoor scene category and its target and minimize this difference.

When training the DCNN, we made sure of using challenging conditions. All our experiments were performed on a workstation computer equipped with intel Xeon E5-2683 V4 processor that was equipped with a tesla K40c graphic processor with 12 GB of graphic memory. The network implementation was done by using TensorFlow-framework. For more details, we used python TensorFlow-GPU 1.13, NVIDIA CUDA toolkit 10.0 and CUDNN 7.0.

In addition, in order to ensure the robustness of our proposed system, we evaluated our work by using various versions of EfficientNet neural network. The two datasets used were divided into two main parts: training and testing set respectively.

In this paper, we trained various versions of efficientNet on MIT indoor 67 dataset and scene 15 dataset. We took advantages of deep CNN in indoor scene recognition and classification. Thanks to the huge development in deep learning field, getting the best accuracies in scene recognition task is crucial. In this work, we used deep learning algorithm based-framework by using transfer learning [35] technique. It is very important to develop a robust indoor scene recognition system. By considering all the different challenges present in the indoor decorations as various lighting conditions, complex background and heavy occlusion. We note that it very difficult to recognize the current scenery, when the input image cannot cover the entire space of the room which makes incomplete information about the current scene.

In this work, we will focus only on recognizing specific indoor scenery as (kitchen, bedroom, bathroom and living room) for MIT 67 dataset and (Kitchen, bedroom, living

Table 2 Experiment parameters' settings

Training iterations	10,000
Learning rate	0.01
Validation set	30%
Testing set	20%
Train batch size	100
Validation batch size	100

Table 3 Comparison of results on MIT 67 indoor dataset

Approach	Accuracy (%)
Deep filter bank [36]	81.0
Places VGGnet + spectral features [36]	84.3
Feature level FOSNET-CCG [37]	90.37
Feature level FOSNET-mixed CCM-CCG [37]	90.30
Ours	93.97

room) categories on scene 15 dataset. We believe that these class categories are vital for blind or sighted person during the indoor navigation. It is extremely simple for a human to recognize the surrounding environment, but it is very difficult for a computer to recognize objects with the same level of perception. We note that the aim from training a neural network with a huge amount of data is to acquire knowledge and to develop an application with perception capacities close to that of a human being. In order to ensure more robustness for our proposed indoor scene detection system. in our work we use of transfer learning technique [35]. Transfer learning technique consists in transferring knowledge across tasks. The mind of transfer learning is that the knowledge learned for task one is reused to solve a second task. It consists on reusing weights of the first task and freezing the entire network and retraining the last layers on the new dataset for the new task. As the indoor scenery present a very complex data, we used transfer learning to stimulate the recognition process.

Since training a deep CNN from scratch requires time and memory consumption, we will employ transfer learning technique to accelerate the process. In order to ensure the robustness of our proposed indoor scene recognition system, we trained and tested the neural network using complex images with bad quality. The proposed indoor scene recognition system is built based on efficientNet deep neural network. To improve the accuracy of our application, we trained the proposed system using various versions of efficientNet. In order to avoid overfitting problem, we apply data augmentation technique to provide the network with a bigger amount of training data. Table 2 provides all the parameters settings during our experiments.

Table 3 presents the comparison of results obtained during our experiments and those obtained by the state-of-the-art models on MIT 67 indoor dataset.

As mentioned in Table 3, our proposed indoor scene recognition system outperforms the results obtained by the state-of-the-art models. Our proposed method achieved 93.97% as a recognition rate for the whole indoor MIT 67 dataset.

As presented in Table 4, our proposed indoor scene recognition system achieves very encouraging results coming up to 95.60% as a recognition rate.

Experimental results achieved outperforms the state-of-the-art in term of recognition rate and robustness. To furthermore robustifies our proposed work, we tested our method on new image that was not previously studied which present very challenging conditions.

Table 4 Indoor scene categories recognition rates on MIT 67 dataset

Class name	Accuracy (%)
Kitchen	95.13
Bathroom	95.45
Bedroom	95.94
Living room	95.91
Mean	95.60

Table 5 Comparison of results obtained on scene 15 dataset

Class name	Method in [18] accuracy	Ours accuracy
Bedroom	0.95	0.98
Kitchen	0.80	0.95
Living room	0.95	0.99
Mean	0.90	0.97

Table 6 Evaluation of efficientNet on MIT 67 indoor scene categories

EfficientNet version	Results accuracy (%)
EfficientNet-B0	92.35
EfficientNet-B3	93.57
EfficientNet-B5	94.35
EfficientNet-B7	95.60

Table 7 Evaluation of efficientNet on scene 15 indoor scene categories

EfficientNet version	Results accuracy (%)
EfficientNet-B0	95.05
EfficientNet-B3	95.93
EfficientNet-B5	96.35
EfficientNet-B7	97

As presented in Table 5, our proposed indoor scene recognition system achieves high accuracies compared to the results obtained by the method provided in [18].

Table 6 provides the evaluation of the results obtained by using various versions of efficientNet deep convolutional neural network on the indoor scene categories of MIT 67 dataset.

As presented in Table 6, we evaluated four versions of efficientNet on four indoor scene categories of MIT 67 dataset (kitchen, bathroom, bedroom, living room). As mentioned in Table 5, efficientNet-B7 achieves the best results compared to other versions of efficientNet.

Table 7 provides the evaluation of the results obtained by using various versions of efficientNet deep convolutional neural network on the indoor scene categories of scene 15 dataset.

As mentioned in Table 7, we evaluated our proposed indoor scene recognition system on four versions of efficientNet and on three indoor scene categories of scene 15 dataset (bedroom, kitchen, living room). The results obtained demonstrate the efficiency of efficientNet-B7 compared to other versions.

As far as the blind or the sighted person mobility is considered, the proposed indoor scene recognition system should provide ahead information about the upcoming sceneries. As well as we propose an indoor scene classification for blind and impaired persons, we have to optimize the contribution between recognition accuracy and the processing speed of the inference. The proposed system achieves a processing speed of 12 ms per image which corresponding to 83 FPS, therefore the obtained results matches the needs of blind and VIP mobility.

5 Conclusion

Indoor scene recognition was and still is a crucial task in computer vision task specially to help blind and sighted persons in their indoor daily navigation. Indoor scene recognition plays an essential role and present a vital part in the robot environment cognition. In this paper, we propose a recognition methodology of highly variable indoor scene categories using transfer learning concept using TensorFlow with GPU support. For our implementation, we used different versions of efficientNet to develop the indoor scene recognition application. In order to increase the robustness of our proposed method, we trained and tested the network on very challenging conditions including various illumination conditions, complex background and heavy occlusions. This paper, presets the first work based on an evaluation of efficientNet to develop a robust indoor scene recognition system to help blind and visually impaired persons to explore new indoor environments and to fully participate in the daily life. The results achieved outperform the state-of-the-art models. We obtained 95.60% for the indoor MIT 67 dataset and 97% for scene 15 dataset.

References

1. Breuer T, Macedo GRG, Hartanto R, Hochgeschwender N, Holz D, Hegger F, Jin Z, Muller C, Paulus J, Reckhaus M et al (2012) Johnny: an autonomous service robot for domestic environments. J Intell Robot Syst 66(1–2):245–272
2. Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? The kitti vision benchmark suite. In: 2012 IEEE Conference on computer vision and pattern recognition (CVPR). IEEE, pp 3354–3361
3. Yu J, Tao D, Wang M et al (2014) Learning to rank using user clicks and visual features for image retrieval. IEEE Trans Cybern 45(4):767–779
4. Who: Vision impairment and blindness. http://www.who.int/mediacentre/factsheets/fs282/en/. Accessed 8 Jan 2020
5. Rodrıguez A, Bergasa LM, Alcantarilla PF, Yebes J, Cela A (2012) Obstacle avoidance system for assisting visually impaired people. In: Proceedings of the IEEE intelligent vehicles symposium workshops, vol 35. Madrid, Spain, p 16
6. Quattoni A, Torralba A (2009) Recognizing indoor scenes. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 413–420
7. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural categories. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR’06), New York, USA, pp 2169–2178
8. Tan M, Le QV (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946
9. Song S, Xiao J (2016) Deep sliding shapes for amodal 3d object detection in rgb-d images. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 808–816
10. Couprie C, Farabet C, Najman L, LeCun Y (2014) Toward realtime indoor semantic segmentation using depth information. J Mach Learn Res 1:1–48
11. Yu J, Zhu C, Zhang J et al (2019) Spatial pyramid-enhanced NetVLAD with weighted triplet loss for place recognition. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2019.2908982
12. Riadh A, Mouna A, Said Y et al (2019) Traffic signs detection for real-world application of an advanced driving assisting system using deep learning. Neural Process Lett. https://doi.org/10.1007/s11063-019-10115-8

13. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition, CVPR2016, pp 779–788
14. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: Proceedings of the 30th IEEE conference on computer vision and pattern recognition, (CVPR'17), Honolulu, Hawaii, USA pp 6517–6525
15. Li LJ, Socher R, Fei-Fei L (2009) Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In: CVPR, pp 2036–2043
16. Sudderth EB, Torralba A, Freeman WT (2005) Learning hierarchical models of scenes, objects, and parts. In: ICCV, pp 1331–1338
17. Espinace P, Kollar T, Soto A, et al (2010) Indoor scene recognition through object detection. In : 2010 IEEE international conference on robotics and automation. IEEE, pp 1406–1413
18. Wu P, Li Y, Yang F, Kong L, Hou Z (2018) A CLM-based method of indoor affordance areas classification for service robots. Jiqiren/Robot 40(2):188–194
19. Abu MA, Indra NH, Rahman AHA et al (2019) A study on image classification based on deep learning and tensorflow. Int J Eng Res Technol 12(4):563–569
20. Zhao Z-Q, Zheng P, Xu S-T et al (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30(11):3212–3232
21. Jiao L, Zhang F, Liu F et al (2019) A survey of deep learning-based object detection. IEEE Access 7:128837–128868
22. Hong C, Yu J, Zhang J et al (2018) Multimodal face-pose estimation with multitask manifold deep learning. IEEE Trans Ind Inf 15(7):3952–3961
23. Yu J, Tan M, Zhang H et al (2019) Hierarchical deep click feature prediction for fine-grained image recognition. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2019.2932058
24. Yao J, Yu Z, Yu J, et al (2019) Single pixel reconstruction for one-stage instance segmentation. arXiv preprint arXiv:1904.07426
25. Zhang J, Yu J, Tao D (2018) Local deep-feature alignment for unsupervised dimension reduction. IEEE Trans Image Process 27(5):2420–2432
26. Akilan T, Wu QMJ, Safaei A, Jiang W (2017) A late fusion approach for harnessing multi-CNN model high-level features. In: Proceedings of the 2017 IEEE international conference on systems, man, and cybernetics, SMC 2017, Windsor, Canada, pp 566–571
27. Deng J, Dong W, Socher R, et al (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 248–255
28. Lin T, Maire M, Belongie SJ et al (2014) Microsoft COCO: common objects in context. CoRR abs/1405.0312 (2014). arXiv preprint arXiv:1405.0312
29. Xiao J, Hays J, Ehinger KA, Oliva A, Torralba A (2010) Sun database: large-scale scene recognition from abbey to zoo. In: Proceedings of CVPR
30. Zhou B, Lapedriza A, Xiao J et al (2014) Learning deep features for scene recognition using places database. In: Advances in neural information processing systems, pp 487–495
31. Afif M, Ayachi R, Said Y et al (2019) Indoor object classification for autonomous navigation assistance based on deep CNN model. In: 2019 IEEE international symposium on measurements and networking (M&N). IEEE, pp 1–4
32. 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, pp 364–371
33. 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. https://doi.org/10.1007/s11063-020-10197-9
34. 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, pp 770–778
35. Tan C, Sun F, Kong T et al (2018) A survey on deep transfer learning. In: International conference on artificial neural networks. Springer, Cham, pp 270–279
36. Khan SH, Hayat M, Porikli F (2017) Scene categorization with spectral features. In: Proceedings of the IEEE international conference on computer vision, pp 5638–5648
37. Seong H, Hyun J, Kim E (2019) FOSNet: an end-to-end trainable deep neural network for scene recognition. arXiv preprint arXiv:1907.07570