BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Artificial Intelligence Advances
https://ojs.bilpublishing.com/index.php/aia

ARTICLE

A Novel Dataset For Intelligent Indoor Object Detection Systems

Mouna Afif[1*] Riadh Ayachi[1] Yahia Said[2] Edwige Pissaloux[3] Mohamed Atri[1]

1.Laboratory of Electronics and Microelectronics (EμE), Faculty of Sciences of Monastir, University of Monastir, Tunisia
2. Electrical Engineering Department, College of Engineering, Northern Border University, Arar, Saudi Arabia
3. LITIS, EA4108& CNRS FR 3638, University of Rouen Normandy Rouen, France

ARTICLE INFO

Article history
Received: 26 February 2019
Accepted: 18 March 2019
Published Online: 30 April 2019

Keywords:
Indoor object detection and recognition
Indoor image dataset
Visually Impaired People (VIP)
Idoor navigation

ABSTRACT

Indoor Scene understanding and indoor objects detection is a complex high-level task for automated systems applied to natural environments. Indeed, such a task requires huge annotated indoor images to train and test intelligent computer vision applications. One of the challenging questions is to adopt and to enhance technologies to assist indoor navigation for visually impaired people (VIP) and thus improve their daily life quality. This paper presents a new labeled indoor object dataset elaborated with a goal of indoor object detection (useful for indoor localization and navigation tasks). This dataset consists of 8000 indoor images containing 16 different indoor landmark objects and classes. The originality of the annotations comes from two new facts taken into account: (1) the spatial relationships between objects present in the scene and (2) actions possible to apply to those objects (relationships between VIP and an object).This collected dataset presents many specifications and strengths as it presents various data under various lighting conditions and complex image background to ensure more robustness when training and testing objects detectors. The proposed dataset, ready for use, provides 16 vital indoor object classes in order to contribute for indoor assistance navigation for VIP.

1. Introduction

Indoor object detection and recognition is an important challenging task used in several autonomous and intelligent systems (e.g. in autonomous robots; humanoid robots; mobility assistive devices for people with visual impairments, VIP). However, the VIPs are not able to see the landmarks or indoor objects Therefore, an assistive device must indicate the presence of such information to VIP during indoor navigation.

A possible approach to indoor navigation passes through the integration of target objects recognition, thus using the appropriate data set and ad-hoc learning technology, in the assistive device intelligence.

The proposed annotated dataset was build using raw images of NAVIIS [4]. The selected images were manually annotated using the graphical image annotation tool LabelImg [1]. An object is identified with its bounding box (bbox) and the associated annotation is its class name and its coordinates in the image.

Autonomous indoor navigation, based on visual cues, is still a very challenging and open question. The indoor scene exploration cannot be assisted using satellites navigation (GPS, Galileo, etc.). Indoor object detection and

**Corresponding Author:*
Mouna Afif,
Laboratory of Electronics and Microelectronics (EμE), Faculty of Sciences of Monastir, University of Monastir, Tunisia;
*Email: *mouna.afif@outlook.fr*

DOI: https://doi.org/10.30564/aia.v1i1.925

recognition for wearable real-time systems specific characteristics which should be taken into account such as lighting conditions, the similarity between objects within the same category (having the same or similar shapes), image blur, etc. It must minimize the error, the recognition time (time complexity) and the volume of calculation (spatial complexity). Therefore, the training dataset is of paramount importance as it can be highly used to train object detectors in order to detect multiple indoor objects to efficiently assist the VIP's indoor navigation.

The carefully designed and the labeled indoor image dataset can be used for training and testing of Deep Convolutional Neural Networks (DCNN) thus to come up with new applications to help people with visual impairments to navigate freely in indoor scenes.

An automated system work will be more reliable and efficient if a (high-level) knowledge about the different class presented in the image scene is provided; the annotated dataset is a mean for such knowledge source.

However, the existing datasets do not contribute to the "indoor object" detection since they present either a total indoor scene or same indoor objects in order to perform their classification (e.g. TUW Object Instance Recognition Dataset [2] or MIT [3]).

This work focuses therefore on collecting different indoor images and their annotations give indoor object class its position in the images. Its goal is to present a new labeled indoor object dataset (or indoor landmarks).

Traditional approach of scene understanding and object recognition aims to provide a simple annotation of the object (object class and its position in the image), while the proposed annotation pay attention to the relationship between objects presented in the scene and includes actions possible to apply to objects, actions which can be performed by the VIP on these objects (object's affordances).

This approach requires to identify/recognize (via a physical parse) the relationship between objects present in an indoor scene, e.g. a relative (spatial) position of objects (table and chair) in a living room in order to be able to move around there or to apply an action (e.g. to sit down on the chair or move the chair spatial position or remove it from the scene, etc.).

Therefore, as far as VIP autonomous mobility, it is necessary to propose a specific indoor scene's objects annotation in order to distinguish those objects in the surrounding environments.

The object classes presented in this paper takes into account the scene global (spatial) coherence; moreover, the indoor landmarks specific for independent mobility of VIP are also considered (e.g. the confirmation of the current progress on the straight line).

The rest of the paper is organized as follows:

Section 2 outlines the current state of art on existing indoor datasets used for objects detection and classification.

Section 3 addresses the inter-class end intra-class relationships between objects of an indoor scene.

Section 4 overviews the indoor object detection and recognition (IODR) dataset.

Section 5provides the principle of images annotation using the software LabelImg.

Section 6 presents the dataset description and its possible uses while the section 7 concludes the paper.

2. Related work

There are several types of graphical tools for images annotation. LabeBox [16] is a software platform dedicated for enterprise to train machine learning applications. This software tool can be used with on-premise or hosted data. It is paying for more than 5000 images. Especially, it is used for image segmentation and classification for text, video and audio annotation. LabelMe [17] is an online software graphic tool used for image segmentation. RectaLabel [18] presents another platform for image annotation with polygons and bounding boxes but is used only for macOS. Collecting and annotating large-scale datasets present a challenging key contribution in order to train and test detectors to robustifies object detection tasks as they directly influence the quality and the performance of object recognition.

In [5] authors present a new large-scale synthetic dataset with 500K images physically reconstructed from realistic 3D indoor scenes. Different illuminations of scenes are considered. This dataset can be recommended for three computer vision tasks such as object boundary detection, semantic segmentation, and surface prediction.

In [6] McComarc *et al*. introduce an indoor dataset named SceneNet RGB-D which expands the previous dataset SceneNet [7] providing a different photo-realistic rendering of indoor scenes. This dataset offers perfect per-pixel labeling which helps in indoor scene understanding. It can be used in many computer vision tasks like depth estimation, optical flow calculation, 3D reconstruction, and image segmentation.

Indoor scene understanding presents a central interest to many computer vision applications including assistive human comparison, monitoring systems, and robotics. However, real-world data presents a default in the majority of these tasks. In [8] Silberman*et al*. present an image segmentation approach to interpret objects surfaces and to build relation support between objects present in the indoor RGBD scene. In their approach authors incorporate geometric shapes of objects to better define the indoor

 DOI: https://doi.org/10.30564/aia.v1i1.925

scene. Based on their results, the proposed 3D indoor scene approach leads to better objects segmentation.

A semantic scene completion is presented in [9]. Authors present a model that predicts object volume occupancy and the object category (scene labeling) from a single depth image of a 3D scene. It should be stressed that the knowledge of the objects' identities present in the scene helps to better identify the scene.

Song *et al*. [10] present a new deep CNN for semantic scene completion named "SSCNet". This deep convolutional model uses CNN for producing 3D voxel representation of scene object and their volumetric semantic labels.

In [11] authors propose a model used for repairing 3D shapes constructed from multi-view RGB dataset. These categories of techniques aim to obtain a semantic label for the object present in the scene [12, 13].

In [14] Zhang *et al*. show that the training model with a synthetic dataset improves the results obtained in the computer vision task as it better distinguishes the object boundaries and the object surface.

Qi *et al*. [15] present a human-centric method to synthesize 3D scene layouts (in particular they take the case of rooms). They propose an algorithm which generates 2D map of indoor images. The proposed algorithm can be included in many tasks such as 3D reconstruction, robot mapping, and 3D labeling.

Public benchmarks greatly support scientists by providing datasets that can be used for their algorithms.

A new indoor object dataset was introduced in [19]. Authors present a fully labeled indoor dataset to train and test deep learning models. All images presented in this dataset present one indoor object extracted from its surrounding environments which makes it recommended for classification tasks and not for indoor object detection problem.

The MC Indoor 20000 dataset is used for indoor object classification and recognition. It includes more than 20000indoor images containing 3 indoor objects landmark (door, sign, stairs). The MC Indoor 20000 dataset presents many challenging situations as images rotation, intra-class variation and images variation.

Xiao *et al*. [24] proposed an extensive database named "Scene UNderstanding" (SUN) containing 899 scene categories with over 130519 images. Their dataset presents various categories as indoor urban and nature categories.

Several RGB-D datasets have been introduced for the last few years facilitating the implementation of computer vision applications. However, those datasets present a lack of comprehensive labels on these RGB-D datasets.

In [20] Hua *et al*. introduce a new RGB-D dataset containing 100 scenes named SeneNN. It's an RGB-D indoor dataset containing 100 indoor scenes. All images introduced presents a reconstruction into triangle meshes and having per-vertex, per-pixel annotations. When collecting and annotating the presented dataset, our aim is to perform an indoor detection system based on deep CNN model, while SeneNN dataset treats semantic segmentation problem.

This paper introduces a new indoor object detection and recognition approach: relationships between objects of a scene are also considered as possible labeling. Our aim from this work is to provide a ready annotated dataset that will be used for training a deep CNN model to perform a system used for indoor object detection for this we choose to use the graphical tool LabelImg [1] as an annotation tool.

3. Object Affordances as New Elements for CNN Efficient Classification and Detection

Indoor space presents an important difference comparing to other spaces as it is composed of several objects (e.g. doors, corridors, stairs, elevator, sign, etc) and the human being can directly interact with. Therefore, the possibility of interaction may be property important for their recognition and this property should conveniently annotate.

The affordances related to object are spatially and temporally invariant so it is very efficient to object location and classification by the CNN.

Figure 1 presents the wide intra-class variation between doors in the same dataset. Doors present many shapes, many poses, and many colors. Some doors are in the wood, some are on glass and others on iron. Annotations were done on different doors poses, some opened, and others are closed. All these figure case makes the presented dataset suitable and robust for building and training new indoor object applications based on deep learning techniques.

The biggest strength presented on this dataset is that it provides many challenging conditions in order to perform a robust model training to deal with different indoor environments belonging to various establishments.

Figure 1. intra-class variation

DOI: https://doi.org/10.30564/aia.v1i1.925

4. Indoor Object Detection and Recognition (IODR) Dataset Overview

4.1 Dataset Collection

The proposed dataset is composed of many categories and indoor object landmarks. The indoor object detection and recognition dataset is composed of 8000 indoor images captured under different light conditions (day, night, blurred images). Some examples of the collected images are presented in figure 2.

It should be stressed that the collected images come from the dataset of NAVIIS project [4]. We selected images presenting different lighting conditions to obtain a very robust dataset presenting many situations. Images resolutions of the dataset are various as 1616 x 1232 and 4592 x 3448.

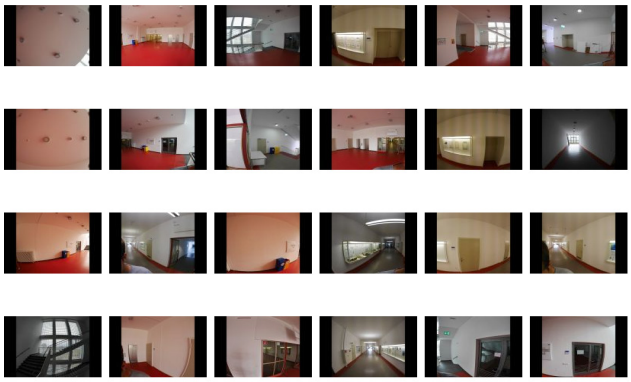


Figure 2. Dataset Images subset

4.2 Dataset Statistics

The dataset is composed of 8000 indoor images with indoor scenes taken under different conditions. This dataset presenting 16 indoor objects landmark. Classes presented in the defined dataset "Indoor object detection and recognition" (IODR) are (door, light, light switch, smoke detector, chair, fire extinguisher, sign, window, heating, electricity box, stairs, table, security button, trash can, elevator and notice table).

The present contribution provides new labeled indoor object dataset freely available for research community. This proposed annotated dataset can be highly recommended for training powerful deep learning models to perform accurate object detection and object recognition.

The proposed dataset is presented and labeled in order to help persons with visual impairments in their mobility in unfamiliar indoor environments like clinics, school, hospitals, and universities and so on. Our collected dataset provides many challenging cases as different lighting conditions, blurred images, and variable point of view of the same object class. Figures 3, 4 and 5 illustrate all these situations.

Figure 3. Different lighting conditions for the same object

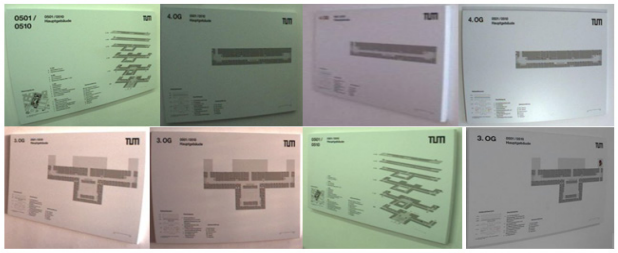


Figure 4. Different intra-class lighting conditions, poses, and point of view

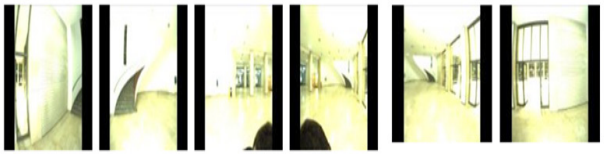


Figure 5. Blurred images presented in the dataset

5. labelingand Annotation via LabelImg Tool

LabelImg [1] is a software tool used for annotating images. LabelImg is a graphical software annotation tool. This software is written based on python background and uses Qt as a graphical interface. It is the widely used amongst other tools. LabelImg saves annotations in the. xml pascal voc [23] format. It is a software platform for developers to easily annotate images in order to train and test deep learning models. It labels objects by putting them in bounding boxes and by providing their x and y coordinates. During the detection and the recognition part, any deep learning model requires knowing the actual objects present in the indoor image. All the required information are present via the annotation.xml file which provides indoor object class ID, Indoor object class name, bounding box presented by their x and y coordinates, height, and width.

The indoor image labels present rich pixel with visual information in addition to the bounding box containing the indoor object. Image annotation is a labor-intensive and error-prone technique. As an example of the most popular annotated datasets are ImageNet [21], Ms COCO [22] and PASCAL VOC [23].

Figure 6 presents an example of an original image and

 DOI: https://doi.org/10.30564/aia.v1i1.925

its corresponding annotation.

The original images included in our data set are selected with respect of two main issues with respect of VIP mobility: (1) providing the most relevant indoor objects and landmarks, and (2) providing a good annotation to better understand the indoor scene. The second aspect is important as the relationship between the object in a given scene are paramount for establishing of the journey path.

Figure 6. Annotated Image Example

In table1 we present all the indoor object classes with all they class names and ID to ensure a better scene understanding of the indoor images present in the dataset. We present 16 main objects remarkable landmarks for indoor navigation that can be provided in any indoor scene. We note that in this work, we are introducing a new indoor multi-class dataset that not previously studied.

6. Dataset Value

6.1 Dataset Description

The paper describes a new indoor object dataset that will be used to evaluate performances of human-assistance navigation systems in indoor scenes. The data collected present various lighting conditions that include multi-level of objects brightness. Objects presented in this dataset are landmarks and vital for visually impaired people indoor navigation. All the images of the dataset are in .jpg format.

This indoor dataset is original as:

(1) It enhances the training and testing of deep learning models as it provides 16 indoor objects landmark.

(2) It includes objects specific for the VIP navigation.

(3) It provides various data under different lighting conditions and multiple images background to ensure robustness in indoor object detection when training.

(4) Provides a fully labeled data ready for use by the scientific community to develop their systems for indoor navigation systems.

(5) This dataset can be included with other indoor datasets in order to ensure robustness and efficiency of the developed deep learning models for indoor robotic navigation systems.

6.2 Recommendations

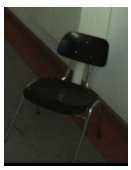
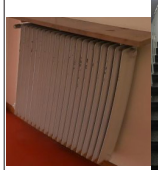
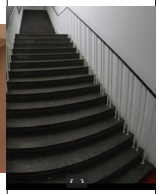
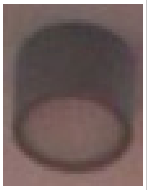
The indoor object dataset presented in this paper is a ready data that can be directly used for researchers in computer vision field to develop new deep convolutional neural networks (DCNN) that can be included in many indoor robotic navigation systems.

These database step-in new applications towards helping a large category of people who are partially sighted and blind persons.

7. Conclusion

This paper presents a new fully labeled dataset for indoor

Table 1. Indoor objects names and ID

Indoor Object								
Class Name	table	chair	smoke detector	fire extinguisher	security button	trash can	door	electricity box
Indoor Object								
Class Name	heating	stairs	notice table	Sign	window	light switch	light	Elevator

DOI: https://doi.org/10.30564/aia.v1i1.925

object detection and recognition. The proposed dataset is original and can be adopted for the design of autonomous robotic navigation systems and for visually impaired people (VIP) assistances. The originality of the proposed dataset comes from the inclusion of new characteristics of a 3D scene not considered so far, namely objects affordance. Such new training data will improve object recognition and may be used for autonomous navigation systems.

The IODR presents 8000 images of different resolution with 16 indoor landmark objects.

Future work on VIP mobility assistance will use the proposed dataset for training and testing of deep convolutional neural networks (DCNN) which will be integrated into an embedded platform.

Conflicts of Interest:

The authors declare no conflict of interest.

References

[1] https://github.com/tzutalin/labelImg accessed: 23-08-2018

[2] https://repo.acin.tuwien.ac.at/tmp/permanent/dataset_index.php

[3] Quattoni, A., &Torralba, A. Recognizing indoor scenes.In 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009: 413-420.

[4] http://www.navvis.lmt.ei.tum.de/dataset/accessed: 21-07-2018

[5] Yinda Zhang, Shuran Song, ErsinYumer, ManolisSavva, Joon-Young Lee, HailinJin, and Thomas Funkhouser.Physically-based rendering for indoor scene understanding using convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5057 – 5065.

[6] John McCormac, AnkurHanda, Stefan Leutenegger, and Andrew J Davison.SceneNet RGB-D: 5M photorealistic images of synthetic indoor trajectories with ground truth. In International Conference on Computer Vision (ICCV): 2697–2706.

[7] A. Handa, V. Patr ˇ aucean, V. Badrinarayanan, S. Stent, ˇ and R. Cipolla. SceneNet: Understanding Real World Indoor Scenes With Synthetic Data. arXiv preprint arXiv:1511.07041, 2015.

[8] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In European Conference on Computer Vision, Springer, 2012: 746–760.

[9] A. Z. A. X. C. M. S. T. F. Shuran Song, Fisher Yu. Semantic Scene Completion from a Single Depth Image.InarXiv, 2016.

[10] S. Song and J. Xiao. Deep sliding shapes for amodal 3D object detection in rgb-d images. In CVPR, 2016.

[11] D. Thanh Nguyen, B.-S. Hua, K. Tran, Q.-H. Pham, and S.- K. Yeung. A field model for repairing 3D shapes. In CVPR, 2016.

[12] S. Gupta, P. Arbelaez, and J. Malik. Perceptual organization and recognition of indoor scenes from RGB-D images.In CVPR, 2013.

[13] K. Lai, L. Bo, and D. Fox. Unsupervised feature learning for 3D scene labeling.In ICRA.IEEE, 2014.

[14] Yi Zhang, WeichaoQiu, Qi Chen, Xiaolin Hu, and Alan Yuille.Unrealstereo: A synthetic dataset for analyzing stereo vision. arXiv preprint arXiv:1612.04647, 2016.

[15] Siyuan Qi, Yixin Zhu, Siyuan Huang, Chenfanfu Jiang, and Song-Chun Zhu.Humancentric indoor scene synthesis using stochastic grammar.In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 5899–5908.

[16] https://github.com/Labelbox/Labelbox

[17] https://github.com/wkentaro/labelme

[18] https://rectlabel.com/

[19] Bashiri, F. S., LaRose, E., Peissig, P., &Tafti, A. P.. MCIndoor20000: A fully-labeled image dataset to advance indoor objects detection. Data in brief, 2018, 17: 71-75.

[20] Hua, B. S., Pham, Q. H., Nguyen, D. T., Tran, M. K., Yu, L. F., &Yeung, S. K.. Scenenn: A scene meshes dataset with annotations. In 3D Vision (3DV), IEEE, Fourth International Conference on 2016: 92-101.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," CoRR, vol. abs/1409.0575, 2014. [Online].
Available: http://arxiv.org/abs/1409.0575

[22] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. ´ Zitnick, "Microsoft COCO: common objects in context," Computing Research Repository, vol. abs/1405.0312, 2014. [Online].
Available: http://arxiv.org/abs/1405.0312

[23] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," International Journal of Computer Vision, 2015, 111: 98–136. [Online].
Available: https://doi.org/10.1007/s11263-014-0733-5

[24] XIAO, Jianxiong, HAYS, James, EHINGER, Krista

 DOI: https://doi.org/10.30564/aia.v1i1.925

A., et al. Sun database: Large-scale scene recognition from abbey to zoo. In : 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 3485-3492.

[25] "Who: Vision impairment and blindness," http://www.who.int/ mediacentre/factsheets/fs282/en/, accessed: 2017-12-08.

DOI: https://doi.org/10.30564/aia.v1i1.925