2021 The 21st International Conference on Control, Automation and Systems (ICCAS 2021)
Ramada Plaza Hotel, Jeju, Korea, Oct. 12∼15, 2021

A Novel Room Categorization Approach to Semantic Localization for Domestic Service Robots

Felix Yustian Setiono[1*], Armagan Elibol[2], Nak Young Chong[3*]

School of Information Science, Japan Advanced Institute of Science and Technology
Nomi, Ishikawa 923-1292, Japan
([1]felix.setiono@jaist.ac.jp, [2]aelibol@jaist.ac.jp, [3]nakyoung@jaist.ac.jp) * Corresponding authors

Abstract: Recently, room categorization as part of indoor robot localization has become a vital topic for semantic mapping. One approach is implemented via scene understanding by integrating available object information in the scene. In this paper, a novel room association approach is proposed based on the prior knowledge of the object appearance frequency in the specific room category inside the house. The front interface of the proposed technique employs a state-of-the-art YOLOv2-based object detection framework. Detected objects and their prior appearance frequency information form the input to the proposed room association through a novel scoring approach. This scoring function avoids any limit on the number of detected objects and is capable of operating with a low object detection confidence level. The experimental results of the novel proposed technique show significant improvement over the previously developed room categorization approach. On average, the correctness score increased up to 0.8387 while the indecisiveness level of the object detection framework decreases.

Keywords: simultaneous localization and mapping, semantic localization, room categorization, object detection, object-room information sharing

1. INTRODUCTION

We need to accelerate the development and deployment of domestic service robots, increasing their potential in response to changes or disturbances in the surrounding environment. Several robotics-related research domains (*e.g.*, robotic mapping and localization, human-robot interaction and understanding, and many others) have been established in past decades in order not only to create an improved capability of the robot but also to develop a safe working environment between humans and robots. One topic on these domains is called Simultaneous Localization and Mapping (SLAM). By far, SLAM has been an active topic of research in recent years due to the increasing demand of robotic exploration in unknown environments with improved sensor capabilities and computing resources. Diverse robotic platforms [1,2] also have become practical applications either in real life or simulation-based experimental tests, where the implementations have been further researched either on single individual agents or multi-robot agents famously called swarm topology.

Semantic localization as part of SLAM has been proposed as one solution to overcome the robot self-localization challenge by incorporating the prior or expert information of the environment. Semantics in terms of the robotics field can be described as the meaning of existing things of places, objects, and other entities occupying the environments, or even the languages used in communicating between robots and humans or between robots themselves [3]. Several previously related pieces of research on the semantics-based mapping and localization have been proposed based on the monocular visual-based perception [4, 5], multi-view perspectives [6], fusion approach of multiple sensor data [7, 8], and many more. Several kinds of research have also implemented semantic mapping for the application of room categorization [9, 10], gas-leaking source localization [11], and created an open-source library for semantic-based localization and mapping framework [12].

The Room categorization technique as part of the semantic-based mapping and localization is proposed and studied to improve the robot's intelligence and capability in indoor environments. One example of these advancements is the capability of the robot to assign the task given to them based on their location information or environmental condition. Another future progress is the ability of the robot to predict the future mission based on their location information without the needs of the preliminary task definition. Several previous types of research have further advanced the features of the room categorization approach by improved the robot's capability for an enhanced task and ability [13, 14]. Several strategies to solve the problem in the room categorization-based research have been studied through machine learning-based classifiers identifying the geometric or appearance features of the room [15, 16]. On the other side, there have been approaches utilizing object detection frameworks based on the understanding and application of the object and their location via their valuable information [17, 18].

Over the last few years, several frameworks for advanced object detection (*e.g.*, YOLOv2 [20], Faster R-CNN [23], and Mask R-CNN [19]) have been developed. Furthermore, the availability of the public datasets on the image categorization (*e.g.*, COCO dataset [22]) has greatly helped to improve the capability of the object detection frameworks. Integrating the object detection framework with the room categorization approach, it is expected that the correctness and confidence level can be further enhanced.

In this paper, we proposed an improved version of the previously developed semantic-based room categorization method [9]. Specifically, we designed a novel score function on the object-room association framework without limiting the number of the detected objects within a scene through the development of an ontology-like knowledge base of the object-room relation probability. This score function improves the room categorization approach and helps the robot to be able to define and locate itself inside the specific room category based on the probability scores of each room category coming from the appearances of detected objects in the room.

2. METHODOLOGY

The primary motivation of this research is to improve the existing room categorization approach implemented in [9]. Specifically, the proposed research aims to claim and validate a general approach to room categorization for indoor semantic localization through the integration of the ontology as the prior knowledge base and the proposed room association score. The existing room categorization approach in [9] has several drawbacks such as the limitation of the numbers of the detected objects, high threshold of the confidence level on the object detection framework, and especially on the proposed Bayesian probability framework. Another approach of room categorization is developed via the deep learning algorithms [4–8]. They have shown significant improvements but suffer from some drawbacks mainly due to the availability and disparity of the datasets that tend to lead to overfitting and the need for high computing resources.

The overview of the proposed room association framework is given in Fig. 1 and **Algorithm 1**. The proposed procedure initially makes use of a pre-trained object detection network with input RGB images. Detected objects and their detection confidence scores are passed to the next step of the scene categorization method incorporating the novel association score function. On the other hand, using existing datasets [24], prior information on object appearance frequency according to the room categories is computed and used as semantic knowledge. The object-to-room probability information extracted from the knowledge base is utilized in the room association computation with the novel score function. In the end, the computational results of several association scores (as many as room categories) become the final measure where the robot most likely to be located.

The YOLOv2 algorithm [20] is selected and used as the object detection algorithm. The state-of-the-art YOLOv2 framework extends the capability of the previous algorithm through improvements on the recall and localization errors [20]. YOLOv2 improved real-time performance and capability of operating over a wide variety of object classes. In our experiments, the detection confidence level and label are utilized as the inputs of the prior information retrieval through the detected object's label and the proposed room association score calculation.

2.1 Semantic Knowledge Understanding

Semantic knowledge is one part of the knowledge representations and could be further enlarged as the prior or expert knowledge by developing a customary understanding of a specific topic or domain. One common approach to semantic knowledge representation can be implemented through the creation of ontology. Ontology is similar to the knowledge base where the required information can be acquired by the specific command or request based on the query languages. The applied ontology-like knowledge base of the relation of the object(s) and the possible room location of the object(s) is developed from the *Place365* dataset [24]. The knowledge base is used when calculating the appearance probability of the detected objects on specific room categories.

2.2 Proposed Room Association Scoring Approach

The probability of the possible room localization of the robot can be represented by the common understanding of the detected object's information. One of the applied object's information is the prior knowledge of the object-room knowledge base based on the relationship between the detected object and their room location. This knowledge base excerpt the object appearance probability in specific room categories. First, we defined the kinds or number of the room categories to be considered; for example: *Kitchen*, *Living Room*, *Bedroom*, *Bathroom* [1]. Using the *Place365* dataset [24], we run object detection framework on images belonging to the room categories. For each object category, we counted the total number of detection over images from room categories. Then we compute the probability of object categories appearing in a certain room by dividing the total number of its appearance in all room categories as given in Eq. 1.

$$P(R_j|O_i) = \frac{O_{ij}}{\sum_{j=1}^{n_{rc}}(O_{ij})} \tag{1}$$

where O_{ij} is the number of occurrences (or frequency) of the object category i in the room category j, while n_{rc} is the total number of room categories, which is set to 4 in this study.

In order to improve the accuracy of room association, we propose to use the following score function given below in Eq. 2. This function is designed to incorporate both the object-room appearance probability computed in Eq. 1 and the object detection confidence level information:

$$S_j = \frac{\sum_{i=1}^{n}\left(w_i \times e^{(P(R_j|O_i)-0.5)}\right)}{\sum_{i=1}^{n} w_i} - e^{(-0.5)} \tag{2}$$

where S_j is the association score computed for room category j over the total number of objects detected represented by n, while w_i represents the detection confidence level from the object detection framework. Multiple instances of the detected objects from the same category can be dealt with by computing their average detection

[1] Please be noted that we excluded the *Dressing Room* category since it uncommonly exists in datasets.

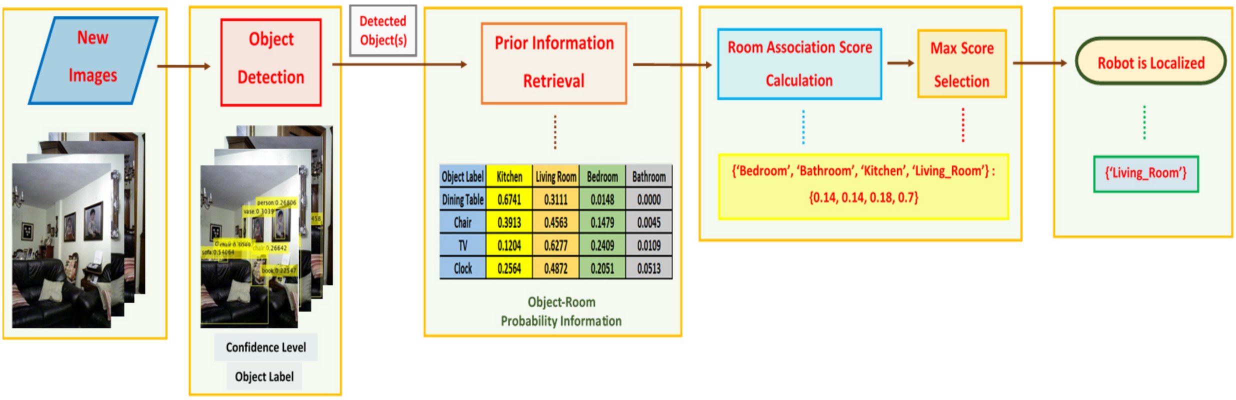

Fig. 1. The overview of the proposed room association approach

confidence level. The term $e^{(-0.5)}$ is used as a correction term to make the association score zero when the object-room appearance probability is equal to zero. This score function is designed in a way to put more emphasis on the objects with higher object-room probability (*e.g.*, ≥ 0.5) while acknowledging the contribution of all detected objects.

On each scene, the room association score values is computed and the room category with the maximum score value is selected as the final decision.

Algorithm 1 Proposed Room Association Approach

SET *frame rate* and *detection confidence level*
INPUT *RGB images*
LOAD *object detection framework*
for *every set of images with frame rate* **do**
    RUN *object detection framework*
    for *every detected object(s)* **do**
        RETRIEVE *detected object(s) information*: *object label(s)*, *confidence level(s)*
        EXTRACT *object-room probability*: Eq. 1
    end for
    CALCULATE *room association scores*: Eq. 2
    SORT *room association scores*
    DEFINE *final room location*: max *room association score*
end for

3. EXPERIMENTAL PROCEDURES

Experiments have been carried out using RGB images obtained from four houses available in the *Robot@Home* dataset [21], namely: *Alma*, *Anto*, *Pare*, and *RX2*. The YOLOv2 was employed as the object detection step and it was pre-trained using the COCO dataset [23] to retrieve the possible object labels. We used three different confidence scores in the ranges of *0.2*, *0.5*, and *0.7* based on the reason of the number of detected objects on the scene. The lower detection scores increase the number of the detected objects means more objects are found on the scene and improve the correctness of the possible room or scene location. The RGB images taken from each room on these houses were entered the object detection framework at first. The label and detection confidence level information of the detected objects were obtained as a result of the object detection framework.

Labels of detected objects are used to retrieve prior object-room probability information obtained previously using Eq. 1 over the *Places365* dataset [24]. Retrieved probabilities along with detection confidence values are later applied to calculate room association score via the proposed function in Eq. 2. The calculation of the room association score is performed on each room category. The room category with the maximum score is defined as the location of the robot.

During the experiments, we applied five different frame rates from 5 to 25. Therefore, a different number of time-consecutive images are fed to the object detection. There is no limit on the total number of detected objects in our framework different from its counterpart in [9], where the limit was set to 5. Also, our object-room probability prior information is retrieved via a data-driven approach using a large dataset in contrast to *ad hoc* definition in [9].

In order to make a comparison, we also carried out experiments with the score function based on multiplying the object-room probabilities with detection confidence levels similar to the one proposed in [9]. This score function is given in Eq. 3 and referred to as the baseline in the experimental results. This function was devised based on the Bayes rule.

$$S_j = \prod_{i=1}^{n} w_i \times P(R_j|O_i) \tag{3}$$

where $P(R_j|O_i)$ is the object-room association probability and w_i is detection confidence level, while n is the total number of the detected objects.

4. EXPERIMENTAL RESULTS

Experiments were carried out using five different frame rates with three different confidence levels of the object detection framework. The average of the correct and incorrect associated scene with the total number of the available set of images per room category were reported. We also showed the ratio of cases where there was no object detected in a set of input images. Such cases are referred to as *indecisive* since no score computation could be carried out.

Table 1. Room Categorization Results for All Houses on the Confidence Level of *0.2*

Room	House	New Score Function		Baseline Bayesian Function		Indecisive
		Correct	Incorrect	Correct	Incorrect	
Bedroom	Alma	0.7310	0.2670	0.3940	0.6040	0.0020
	Anto1	0.6920	0.3080	0.4100	0.5900	0.0000
	Anto2	0.4580	0.5420	0.3050	0.6950	0.0000
	Pare1	0.5440	0.4560	0.3370	0.6630	0.0000
	Pare2	0.7620	0.2380	0.5800	0.4200	0.0000
	RX2	0.9830	0.0170	0.8740	0.1100	0.0170
Bathroom	Alma	0.8870	0.1080	0.5160	0.4790	0.0500
	Anto1	0.9040	0.0960	0.4360	0.5640	0.0000
	Anto2	0.9970	0.0000	0.0540	0.9430	0.0030
	Pare1	0.9480	0.0520	0.5090	0.4910	0.0000
	Pare2	0.9960	0.0040	0.6480	0.3520	0.0000
	RX2	0.8360	0.1640	0.3890	0.6110	0.0000
Kitchen	Alma	0.9500	0.0500	0.9850	0.0150	0.0000
	Anto	0.8630	0.1370	0.7590	0.2410	0.0000
	Pare	0.8820	0.1180	0.7920	0.2080	0.0000
	RX2	0.7980	0.2020	0.9310	0.0690	0.0000
Living Room	Alma	0.8330	0.1670	0.7100	0.2900	0.0000
	Anto	0.8390	0.1610	0.8670	0.1330	0.0000
	Pare1	0.9960	0.0040	0.9890	0.0110	0.0000
	Pare2	0.9480	0.0520	0.9340	0.0660	0.0000
	RX2	0.8540	0.1460	0.0960	0.9040	0.0000

The experimental results on all houses for the proposed approach compared with the one in [9] using the object detection framework confidence level of *0.2* are presented in Table 1. The presented numbers are the average of the room categorization results of each room category over five different frame rates. From Table 1, it can be concluded that the proposed score function achieved better accuracy on almost all rooms compared with the baseline Bayesian score function in Eq. 3.

The results of Table 2 for the room categories of *Bedroom* and *Bathroom* in the house of *Anto*, and the same room categories plus *Living Room* in *Pare* are the averaged one since these houses are having more than one above-stated room categories. From the results reported in Table 2, the new score function could improve the correctness scores by around 22 percent with a low indecisiveness score. Meanwhile, from the experimental results with the object detection confidence level of *0.5*, the proposed approach could improve the correct association rates by 4 percent with an indecisiveness score of around $2 \sim 3$ percent. Using the confidence level of *0.7* on the object detection framework, we found that the correct association rates between the proposed score function and the baseline Bayesian probability function are on the same level. On the other hand, the total number of indecisive cases arise significantly to a value of around 19 percent. This is mainly due to the low total number of objects detected in the input images.

The confusion matrices for each of the tested confidence levels of the object detection framework on the new score function are provided in Table 3 and the baseline Bayesian probability function in Table 4, respectively. As can be noted, all the defined room categories have the highest classification rates on the confidence levels of *0.2*

Table 2. Average Correct, Incorrect, and Indecisive Ratios for the Experimental Results

Confidence Level *0.2*

House	New Score Function		Baseline Bayesian Function		Indecisive
	Correct	Incorrect	Correct	Incorrect	
Alma	0.8421	0.1562	0.6371	0.3612	0.0017
Anto	0.7974	0.2022	0.5289	0.4708	0.0004
Pare	0.8511	0.1489	0.6956	0.3044	0.0000
RX2	0.8748	0.1203	0.5808	0.4143	0.0049
Global Average	0.8387	0.1600	0.6167	0.3820	0.0013

Confidence Level *0.5*

House	New Score Function		Baseline Bayesian Function		Indecisive
	Correct	Incorrect	Correct	Incorrect	
Alma	0.7745	0.1839	0.7518	0.2066	0.0416
Anto	0.8538	0.1330	0.7628	0.2240	0.0132
Pare	0.8221	0.1455	0.8204	0.1472	0.0324
RX2	0.7888	0.2003	0.7207	0.2684	0.0109
Global Average	0.8162	0.1591	0.7732	0.2021	0.0247

Confidence Level *0.7*

House	New Score Function		Baseline Bayesian Function		Indecisive
	Correct	Incorrect	Correct	Incorrect	
Alma	0.5125	0.2355	0.5307	0.2172	0.2521
Anto	0.7143	0.1280	0.7047	0.1377	0.1576
Pare	0.6468	0.1050	0.6491	0.1027	0.2482
RX2	0.6701	0.2270	0.6707	0.2265	0.1029
Global Average	0.6458	0.1578	0.6473	0.1563	0.1965

and *0.5*, where the lowest indecisiveness scores appeared on *Kitchen* and *Living Room*. The highest indecisiveness score emerged on the room category of *Bedroom* as shown in the confusion matrices of all of the confidence levels of the object detection framework due to several common objects appeared and detected on the room.

Based on the experimental results, the proposed score function improved the correct associations on the room categorization technique and reduced the indecisiveness levels. When the object detection framework ran on the confidence level of *0.7*, the indecisiveness level was significant and made several possible important or primary objects in the rooms unable to detect. In other words, there is less number of object categories on the images possible to be recognized on the confidence level of *0.7* compared to the number of the detected object on the other two confidence levels. In the end, this problem produced no difference in the correctness scores between the new score function and the baseline Bayesian probability function. It can be concluded that the new score function is still working on the high confidence level of the object detection framework but has not achieved its best performance.

When the confidence level of the object detection framework was selected as low as *0.2*, whereby more numbers of the object were detected, the new score function can produce better association accuracy due to its emphasis on the objects with higher object-room prob-

Table 3. Confusion Matrices and the Indecisiveness Levels for the New Score Function

Confidence Level *0.2*

	Kitchen	Living Room	Bedroom	Bathroom	Indecisive
Kitchen	**0.8694**	0.0670	0.0088	0.0548	0.0000
Living Room	0.0767	**0.8895**	0.0335	0.0003	0.0000
Bedroom	0.1129	0.1632	**0.7050**	0.0154	0.0034
Bathroom	0.0353	0.0353	0.0024	**0.9256**	0.0014

Confidence Level *0.5*

	Kitchen	Living Room	Bedroom	Bathroom	Indecisive
Kitchen	**0.7324**	0.0969	0.0391	0.1262	0.0054
Living Room	0.0446	**0.8529**	0.0910	0.0000	0.0115
Bedroom	0.0695	0.1191	**0.7598**	0.0062	0.0455
Bathroom	0.0143	0.0258	0.0010	**0.9266**	0.0324

Confidence Level *0.7*

	Kitchen	Living Room	Bedroom	Bathroom	Indecisive
Kitchen	**0.4927**	0.1918	0.0303	0.1282	0.1570
Living Room	0.0237	**0.6597**	0.1576	0.0000	0.1590
Bedroom	0.0209	0.0770	**0.6626**	0.0017	0.2379
Bathroom	0.0000	0.0191	0.0000	**0.7525**	0.2284

Table 4. Confusion Matrices and the Indecisiveness Levels for the Baseline Bayesian Function

Confidence Level *0.2*

	Kitchen	Living Room	Bedroom	Bathroom	Indecisive
Kitchen	**0.8464**	0.1531	0.0005	0.0000	0.0000
Living Room	0.1217	**0.7242**	0.1538	0.0003	0.0000
Bedroom	0.1964	0.3015	**0.4952**	0.0034	0.0034
Bathroom	0.5346	0.0415	0.0076	**0.4149**	0.0014

Confidence Level *0.5*

	Kitchen	Living Room	Bedroom	Bathroom	Indecisive
Kitchen	**0.7432**	0.1541	0.0475	0.0499	0.0054
Living Room	0.0338	**0.8354**	0.1192	0.0000	0.0115
Bedroom	0.1023	0.1588	**0.6889**	0.0044	0.0455
Bathroom	0.0792	0.0272	0.0262	**0.8350**	0.0324

Confidence Level *0.7*

	Kitchen	Living Room	Bedroom	Bathroom	Indecisive
Kitchen	**0.4868**	0.2192	0.0318	0.1052	0.1570
Living Room	0.0181	**0.6764**	0.1464	0.0000	0.1590
Bedroom	0.0195	0.0770	**0.6639**	0.0017	0.2379
Bathroom	0.0119	0.0191	0.0000	**0.7406**	0.2284

ability. The baseline Bayesian probability framework could not solve this situation, since its probability function will produce near-zero scores at the end, when the number of detected objects increase.

5. CONCLUSIONS AND FUTURE WORK

Localization plays an important role in SLAM. In this paper, we proposed a semantic localization framework for domestic service robots navigating freely in indoor environments. The proposed framework is composed of the state-of-the-art YOLOv2 object detection framework and the newly designed score function. The proposed approach has shown improved results shown in the experimental results without limiting the number of detected objects inside the scene while maintaining the correct association. An ontology-like knowledge base was employed in this study as the semantic-based information representation containing the preceding information on the probabilities of the occurrences of the detected objects that existed inside the specific room categories.

Our proposed novel room association approach for indoor localization implementation has been tested and evaluated by applying the *Robot@Home* dataset using the available RGB images under specific room categories. The proposed approach has shown to be effective in recognizing the representative room categories with the correctness score up to around 0.8387 on average. Besides that, the proposed room association approach could decrease the indecisiveness score setting the lower confidence level of around *0.2* on the object detection framework.

For future work and the improvement of the capability of the proposed approach, we plan to integrate it into a more advanced framework. The future framework will attempt to utilize the progressive scene understanding framework combining the object detection framework and feature-based scene recognition.

REFERENCES

[1] Aguilar, Wilbert G., et al. ”Visual SLAM with a RGB-D camera on a quadrotor UAV using on-board processing.” in *International Work-Conference on Artificial Neural Networks*, 2017, Springer, Cham.

[2] Shen, Dong, et al. ”Research and implementation of SLAM based on LIDAR for four-wheeled mobile robot.” in *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, IEEE, 2018.

[3] Garg, Sourav, et al. ”Semantics for robotic mapping, perception and interaction: A survey.” in *arXiv preprint*, arXiv:2101.00443, 2021.

[4] Xiao, Linhui, et al. ”Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment.” *Robotics and Autonomous Systems*, 117, pp. 1-16, 2019. 1

[5] Zeng, Zhen, et al. ”Semantic mapping with simultaneous object detection and localization.” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018.

[6] Gawel, Abel, et al. ”X-view: Graph-based semantic multi-view localization.” *IEEE Robotics and Amation Letters*, 3.3, pp. 1687-1694, 2018.

[7] Wang, Peng, et al. ”Dels-3d: Deep localization and segmentation with a 3d semantic map.” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[8] Qi, Xianyu, et al. ”Object semantic grid mapping with 2D LiDAR and RGB-D camera for domestic robot navigation.” *Applied Sciences*, 10.17, 5782, 2020.

[9] Fernandez-Chaves, David, et al. ”From object detection to room categorization in robotics.” in *Proceedings of the 3rd International Conference on Applications of Intelligent Systems*, 2020.

[10] Qian, Zhentian, et al. ”Semantic SLAM with Autonomous Object-Level Data Association.” *arXiv preprint*, arXiv:2011.10625, 2020.

[11] Monroy, Javier, et al. ”A semantic-based gas source

localization with a mobile robot combining vision and chemical sensing." *Sensors*, 18.12, pp 4174, 2018.

[12] Rosinol, Antoni, et al. "Kimera: an open-source library for real-time metric-semantic localization and mapping." in *Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, IEEE.

[13] Uršič, Peter, Aleš Leonardis, and Matej Kristan. "Part-based room categorization for household service robots." in *Proceeding of 2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, IEEE.

[14] Chaves, D., et al. "Integration of CNN into a robotic architecture to build semantic maps of indoor environments." in *Proceeding of International Work-Conference on Artificial Neural Networks*, 2019, Springer, Cham.

[15] Pronobis, Andrzej, et al. "Multi-modal semantic place classification." *The International Journal of Robotics Research*, 29.2-3, pp. 298-320, 2010.

[16] Quattoni, Ariadna, and Antonio Torralba. "Recognizing indoor scenes." in *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, IEEE.

[17] Pronobis, Andrzej, and Patric Jensfelt. "Hierarchical Multi-Modal Place Categorization." *ECMR*. 2011.

[18] Pereira, Ricardo, et al. "Deep-Learning based Global and Semantic Feature Fusion for Indoor Scene Classification." in *Proceedings of 2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2020, IEEE.

[19] He, Kaiming, et al. "Mask r-cnn." in *Proceedings of the IEEE international conference on computer vision*, 2017.

[20] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[21] Ruiz-Sarmiento, José Raúl, Cipriano Galindo, and Javier González-Jiménez. "Robot@ home, a robotic dataset for semantic mapping of home environments." in *The International Journal of Robotics Research*, 36.2, pp.131-141, 2017.

[22] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." in *European conference on computer vision*, 2014, Springer, Cham.

[23] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." in *arXiv preprint*, arXiv:1506.01497, 2015.

[24] Zhou, Bolei, et al. "Places: A 10 million image database for scene recognition." *IEEE transactions on pattern analysis and machine intelligence*, 40.6, pp. 1452-1464, 2017.