Science in Information Technology Letters
Vol. 5., No. 1, May 2024, pp. 26-39
ISSN 2722-4139
http://pubs2.ascee.org/index.php/sitech

Determination of living quarters clutter for caregiver support

Stephen Karungaru [a,1,*]

[a] Computer Science Department, Tokushima University, 2-1 Minami Josanjima, Tokushima, 770-8506, Japan
[1] karungaru@tokushima-u.ac.jp
* Corresponding Author

ARTICLE INFO

Article history
Received May 28, 2024
Revised October 15, 2024
Accepted December 31, 2024

Keywords
Room Clutter
Yolov5
Caregiver support
Elderly living

ABSTRACT

Providing enough health caregivers due to an aging population has recently been challenging. To alleviate this problem, there's a growing demand for certain household monitoring tasks to be automated especially for elderly persons living independently to reduce the number of scheduled visits by caregivers. Moreover, gathering crucial data using AI technology about functional, cognitive, and social health status, is essential for monitoring daily physical activities at home. This paper proposes a system that determines a room's cleanliness (degree of clutter) to decide whether a caregiver visit is required. A Yolov5-based method is applied to recognize objects in the room including clothes, utensils, clothes, etc. However, due to background noise interference in the rooms and the insufficient feature extraction in YOLOv5, an improvement regime is proposed to improve the detection accuracy. The ECA (Efficient Channel Attention) is added to the network's backbone to focus on feature information, reducing the missed detection rate. The initial anchor box clustering algorithm is improved by replacing K-means with the K-means++ algorithm, enabling more effective adaptation to changing room views. The regression loss function EIoU (Enhanced Intersection over Union) is introduced to optimize the convergence speed and improve the accuracy. The room clutter is determined using set rules by comparing the detection results and prior information from the clean room using IOU. In 31 rooms, 9 subjects' evaluation was used to prove the effectiveness of the proposed system. Compared to the original Yolov5 algorithm, the method proposed in this paper achieved better performance.

1. Introduction

Artificial intelligence (AI) has significantly changed the elderly care market globally in recent years. Artificial Intelligence (AI) describes systems that analyze their surroundings and operate autonomously to accomplish particular goals [1]–[3]. AI-powered solutions, for instance, can help senior citizens live more independently by supporting caregivers in detecting falls, prescription administration, and navigation [4]. Moreover, AI-powered smart home devices can identify departures from typical behavior patterns and promptly notify users of impending emergencies [5]. Moreover, many elderly persons are living alone because their children are grown up and have moved out [6]. They are also adversely affected by many chronic diseases and other factors that require frequent caregiver visits. Although technology is supporting mitigate many health risks [7]–[9], there are not enough caregivers to cover all situations.

In Japan, society aging is progressing fast. Since 1950, the elderly population has risen, reaching 35.57 million in 2018 [10]. This trend is expected to continue, and there are concerns that the burden on the nursing care industry will increase. Additionally, there is a shortage of staff involved in nursing care, and the situation is worsening. Therefore, recently, nursing care prevention has become important. Nursing care prevention is an activity that improves motor function and nutritional status to minimize the need for care. Local comprehensive support centers in municipalities mainly carry out such support activities. However, the content of support, and whether to intervene is determined by interviews or staff visiting the elderly person's home directly. This process can be slow, tedious, and subjective. Therefore, building a system that automatically determines the cleanliness state of a room from captured images can help reduce the burden on staff and improve intervention timings.

In related works, some studies describe a method for estimating the degree of beauty of a room [11]. The problems in the system include multiple camera installations and removing areas that violate the resident's privacy. CNNs are being utilized extensively in object detection, particularly in human/object detection, due to recent developments in deep learning and CNN designs. As a result, numerous novel and highly accurate object identification techniques have been created, including Mask R-CNN [12], YOLO [13], RetinaNet [14], SSD [15], and Faster R-CNN [13]. In this paper, we leverage, the power of deep learning in object detection. A comprehensive summary of research on elderly support in smart homes is given in. Therefore, it is necessary to automatically evaluate the degree of clutter in a room to determine if cleaning/care is required especially for the elderly living alone. In other research, room cleanliness has been approached from a robotics perspective using autonomous robots [16], [17]. However, robots can only clean the floor and cannot distinguish between correctly and scattered stored items.

The structure of this paper is as follows. Initially, improvements to the original YOLOv5 are presented. Next, object detection is performed, an explanation of the storage determination of objects is offered and the room clutter is determined. Finally, the experiments, results, and discussion are presented.

2. Method

The proposed method consists of object detection using an improved Yolov5, followed by object location determination using the IoU (Intersection over Union), determining the level of the clutter an evaluation determination formula, and finally visualizing the results.

2.1. Yolov5

The YOLOv5 [18] marked a significant improvement in network architecture. Its network structure comprises four parts: the input, backbone network, Neck module, and output. The input image is processed with improvements including Mosaic data augmentation, adaptive anchor box calculations, and adaptive image scaling. Adaptive anchor box calculation automatically determines the optimal anchor boxes for various datasets before training, eliminating the need for separate calculations. Due to varying image sizes, adaptive image scaling minimizes the addition of black borders during scaling, reducing training time and speeding up inference. The backbone network introduces the Focus structure and

Cross Stage Partial Networks (CSPNet). The Focus structure expands input channels through slicing operations, where a 2×2×12 feature map is derived from a 4×4×3 image via slicing. CSPNet's advantage lies in its reduced parameter count, easing computational load while maintaining high performance.

2.2. Yolov5 improvements: Attention Mechanism

To acquire better detection information, the human visual system can focus on a target area, concentrating visual attention on areas of high- resolution while ignoring low-resolution areas. Drawing on this principle, attention mechanisms have been applied to neural network learning to avoid information overload. Attention mechanisms help the neural networks focus on key information relevant to the current task. In current developments in deep learning algorithms convolution and self-attention mechanisms [19], [20] have been continuously refined and widely applied in fields such as object detection and tracking. To enhance the Yolov5 algorithm's performance in reducing missed detection, the introduction of the ECA (Efficient Channel Attention) mechanism is considered. Therefore, a channel attention mechanism, the Squeeze-and-Excitation Network (SENet), is introduced. Through the "Squeeze-and-Excitation (SE) block", the network's representation power is significantly enhanced.

2.3. Yolov5 improvements: K-Means++

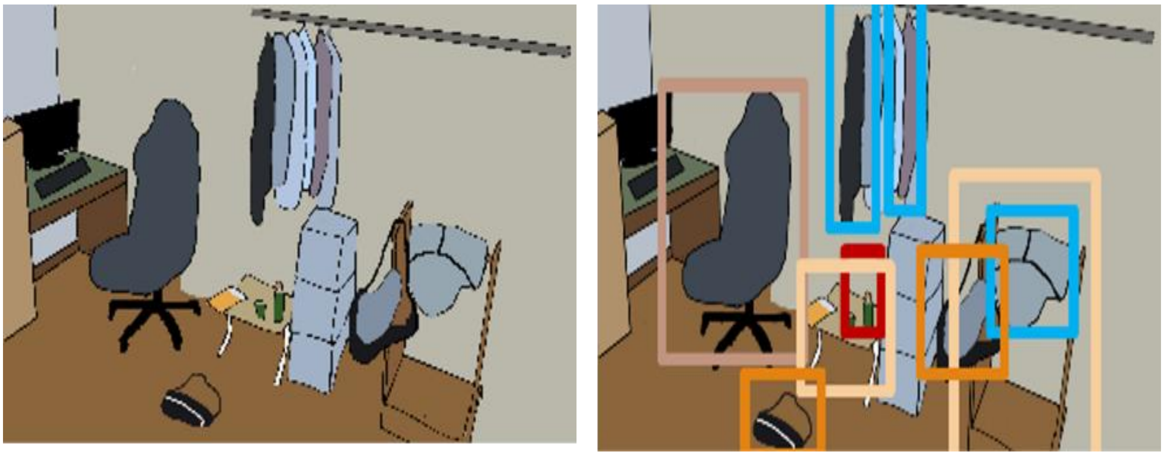
K-means [21] is a classic and efficient clustering algorithm widely used for anchor box clustering in object detection algorithms to adjust prediction boxes. It clusters samples by calculating their Euclidean distances, grouping closer samples into the same category. The K-means algorithm adjusts anchor box sizes to fit specific datasets, facilitating faster convergence during model training. However, the results of the K-means algorithm largely depend on the initial selection of cluster centers, often requiring multiple clustering operations to achieve stable results. To alleviate the problem, the K-means++ algorithm [22] for re-clustering analysis of the detected objects cluster center selection is proposed. The K-means++ algorithm ensures that initial clustering centers are more dispersed in distance, providing more stable and reliable clustering outcomes.

2.4. Yolov5 improvements: Introduction of EIoU Loss Function

In YOLOv5, the CIOU loss function [23] is employed for the localization loss. CIOU, an improved version of the IoU (Intersection over Union) loss function [24], is specifically designed for bounding box localization in object detection. It enhances model accuracy in localization by adding considerations for the distance between center points and the aspect ratio, based on the standard IoU. However, CIOU has limitations in addressing the aspect ratio differences between the predicted and actual bounding boxes, mainly because it relies on relative values to adjust the aspect ratio without directly considering the real differences in width and height. To address this issue, the EIOU (Efficient Intersection over Union) loss function [25] is proposed, a further optimization of the CIOU loss function. The EIOU loss function specifically emphasizes the aspect ratio differences between the predicted and actual bounding boxes by independently calculating width and height, effectively compensating for the shortcomings of CIOU in this respect. The EIOU loss function offers a more detailed and comprehensive approach to bounding box recognition. Compared to traditional IoU loss functions, EIOU provides a more precise and comprehensive assessment in these areas. Furthermore, EIOU loss function is crucial for enhancing the performance of YOLOv5 in various scenarios. Especially when dealing with objects of different sizes and shapes, EIOU can better adjust the predicted bounding boxes, thereby improving detection accuracy.

2.5. Household items detection

CNN is used as a classifier of the ripeness of oil palm fruit. The essential instinct behind these frameworks is that a processing architecture based on a huge number of layered and massively interconnected simple units may be fit than sophisticated algorithms to handle complex issues. The fundamental processing unit, the neuron, is exceptionally basic. It calculates the output activation by looking at the weighted entirety of its contribution with a threshold and applying a suitable nonlinearity, below is Fig.1.



(a) Room Image (b) Result

Fig. 1. Object detection

2.6. Object location Determination

The IOU loss is used to determine the location of objects and by extension their storage status. The IOU used here should not be confused with the EIoU used to improve Yolov5 algorithm. Initially, the correct storage locations are known. By calculating the IOU loss, and comparing the center of gravity of the targets, a difference of ±10% of the median value is determined to have been stored correctly. Fig. 2. shows the clothes storage determination results.

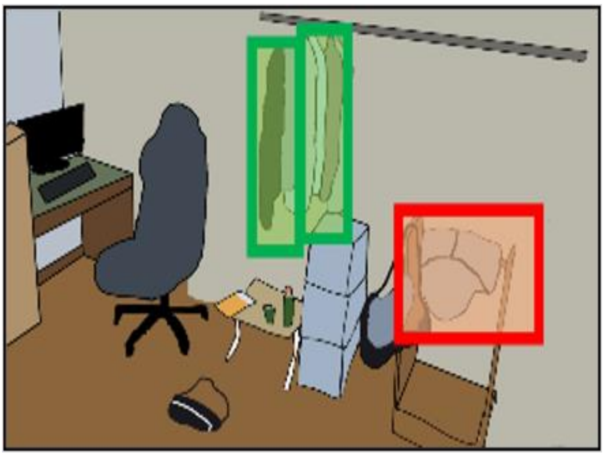
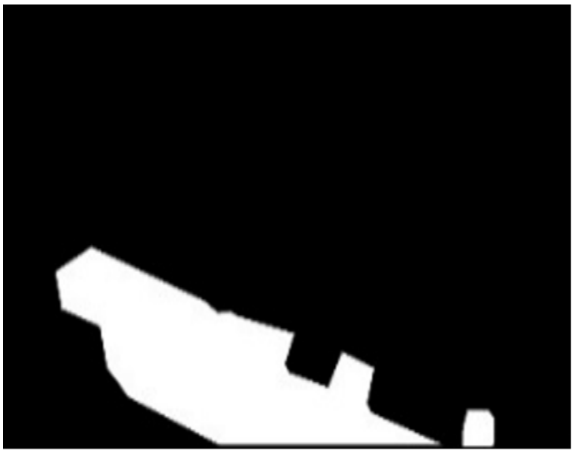


Fig. 2. Objects storage determination

2.7. Object-on-floor Determination

If the detected object is not correctly stored, it is necessary to determine where they are located. Initially, the floor area mask created beforehand is compared with the object bounding box to determine if the object is on the floor. If the IOU loss is above 70%, it is concluded that the object is on the floor. Fig. 3(a) shows the mask image, and Fig. 3(b) shows the determination results.



(a) Mask Image

(b) Result

Fig. 3. Floor Area Determination

2.8. Room Furniture Determination

The room contains several types of furniture and other objects. There are three types of storage furniture: clothes racks, shelves, and desks. By evaluating the IOU loss the storage status can be determined. A high loss signifies proper storage. However, if the object is garbage, and the other object is not a dustbin, then the object is not properly stored. Fig. 4 shows the storage determination results.

Fig. 4. Storage determination

2.9. Clutter Scoring: Weight assignment

Weighting is applied to the results of state determination to calculate the level of room clutter. A weight of 1 is assigned to correctly stored objects and 30 to un-stored objects. However, un-stored objects on the floor are assigned a weight of 200. The weighting is done intuitively. A large weight on objects on the floor indicates a high possibility of room clutter.

Fig. 5 shows the results of storage determination and weighting. Green indicates stored, and red indicates not stored.

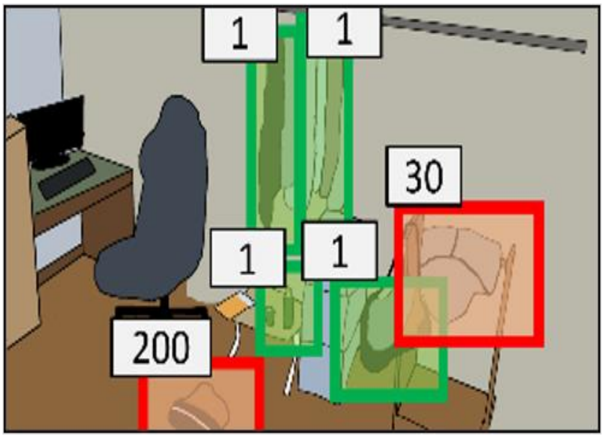


Fig. 5. Weight assignment

2.10. Clutter Scoring: Calculating the degree of room clutter

Finally, the degree of randomness calculated ranged from 0 to 100 points. The tidier the room, the higher the score will be. High clutter rooms are scored close to 0 points. The following two equations are used for scoring. $(h(x,y))$ is the score based on the weights and $f(x,y)$ assigned the final room score.

$$h(x,y) = 0.1 * \frac{\sum_{k=0}^{x} Area_k * Weight_k}{\sum_{k=0}^{x+y} Area_k * Weight_k} \quad (1)$$

$$f(x,y) = 100 * \mathrm{h(x,y)} + \frac{100}{1+e^{-0.1(x-10)}} \quad (2)$$

where x: number of cluttered objects, y: number of organized objects, Area: area of object, Weight: weight of object

2.11. Clutter Scoring: Visualization of results

After scoring, the storage results and scores are written into the input image and output as a result image. Fig. 6 shows the resulting image after writing the scores.

Fig. 6. Scoring results

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the file "MSW_USltr_format".

3. Results and Discussion

3.1. Experimental environment

31 room images were used (five different rooms) each 5184 x 3888 pixels. For object detection, improved YOLOv5 is trained using self-acquired data and detected. During learning, 900 images for training and 107 images for verification were used. Nine persons with no prior knowledge of the room where the photos were taken performed the scoring. Each person was asked to score the room tidiness assigning 0 for dirty and 100 for clean respectively.

3.2. Comparison with test subject results

Scoring was performed 2 times per room by nine subjects and compared with the system's scoring results. Table 1 show the results of comparing the scoring results.

Table 1. Scoring results

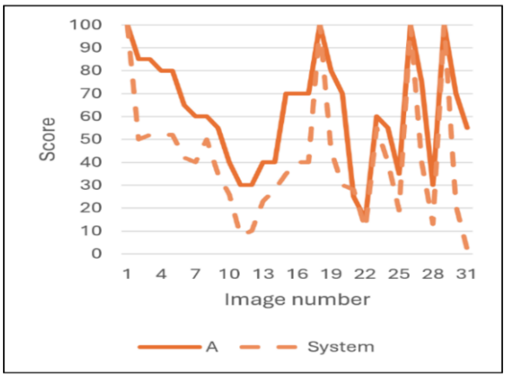
Room No	System	A	B	C	D	E	F	G	H	I
1	100	100	95	85	95	100	100	100	85	96
2	50	85	65	50	80	90	90	80	70	90
3	52	85	55	40	80	90	90	80	70	92
4	52	80	55	40	80	80	80	78	68	90
5	52	80	55	35	70	75	85	78	68	89
6	42	65	20	10	75	70	70	73	65	70
7	40	60	15	5	68	65	60	75	63	68
8	50	60	25	30	70	70	67	75	65	70
9	35	55	35	40	70	65	63	73	61	60
10	26	40	10	0	60	55	65	70	50	20
11	8	30	30	10	40	45	30	65	40	15
12	10	30	50	30	38	45	28	55	35	15
13	23	40	90	55	58	55	40	70	40	38
14	28	40	15	0	60	60	45	70	40	25
15	35	70	10	5	65	60	50	75	45	40
16	40	70	95	60	65	65	58	78	45	40
17	40	70	35	45	65	65	55	75	52	35
18	100	100	70	70	95	100	100	100	82	100
19	45	80	55	40	70	80	80	78	70	80
20	30	70	70	45	50	70	72	73	65	76
21	28	25	55	30	30	30	12	50	30	15
22	11	15	25	5	20	20	10	43	23	10
23	55	60	25	25	50	70	25	73	55	78
24	40	55	35	45	40	50	10	72	50	70
25	19	35	50	25	15	30	0	68	47	40
26	100	100	70	65	85	70	100	100	75	90
27	40	75	25	35	50	50	66	70	69	70
28	13	30	100	95	15	30	30	65	50	50
29	100	100	65	55	100	100	100	100	90	100
30	21	70	65	40	50	50	70	70	50	86
31	2	55	70	45	40	30	55	63	40	68

In the following sections, the results of each subject results are discussed in detail because the subjectivity of the subject produced varying results.

3.3. Subject A

Subject A's score were similar to the system's scoring results. In particular, for room No. 18, the scoring results of the system and subject were the same. The reason could be there were almost no objects on the floor. There are shelves, a TV, a desk, a chair, documents, bottles, etc. in the target images. However, they are all stored inside the shelves and are not scattered on the floor. As a result, the system's scoring process did not result in a significant point deduction, and a high score was output as a clean condition. Similarly, Subject A judged the floor to be clean based on the condition of the floor and the state of storage of objects and gave it a high score. The scoring results for A and the system are shown in Fig. 7.

(a) Room image 18



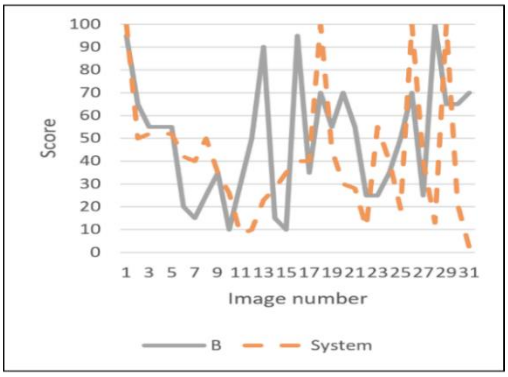
(b) Scoring Result

Fig. 7. A scoring results

3.4. Subject B

Subject B judged some images to be significantly more clean than the system. In room 13 there was a desk, a chair, a clothes hanger, an empty plastic bottle, and trash tissue on the desk. The large score difference between the subject and the system can be due to 2 reasons. First, the number of objects and the false detection of tissue waste. There are 12 objects on the desk, but they do not cover the top board. Therefore, it is thought that Subject B evaluated the target image as clean because the top board was exposed in some places and objects were not scattered on the surrounding floor. Second, because the system evaluates the number of objects in the scoring process, the points lost were large. Additionally, points were deducted for incorrectly detecting a piece of tissue paper large enough to cover a desk. The scoring results for B and the system are shown in Fig. 8.

(a) Room image 13

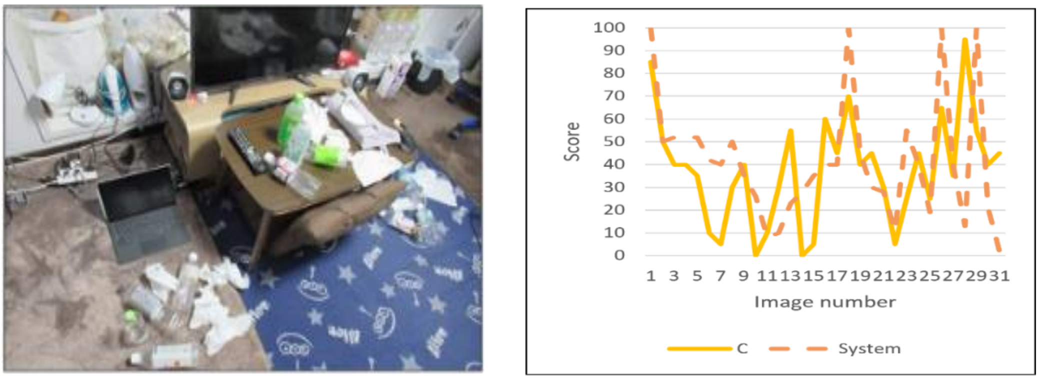


(b) Scoring Result

Fig. 8. B scoring results

3.5. Subject C

Subject C tended to rate the room as dirtier compared to the system. However, some rooms are rated cleaner than the system. In room 28, there were plastic bottles and tissue papers scattered around the desk. However, behind the chair and on the carpet on the left side of the image, there is almost no garbage. Therefore, maybe Subject C thought there was little trash and rated it as clean. On the other hand, the system's scoring process gave a low score because it gave a large weight to objects on the floor. The scoring results for C and the system are shown in Fig. 9.

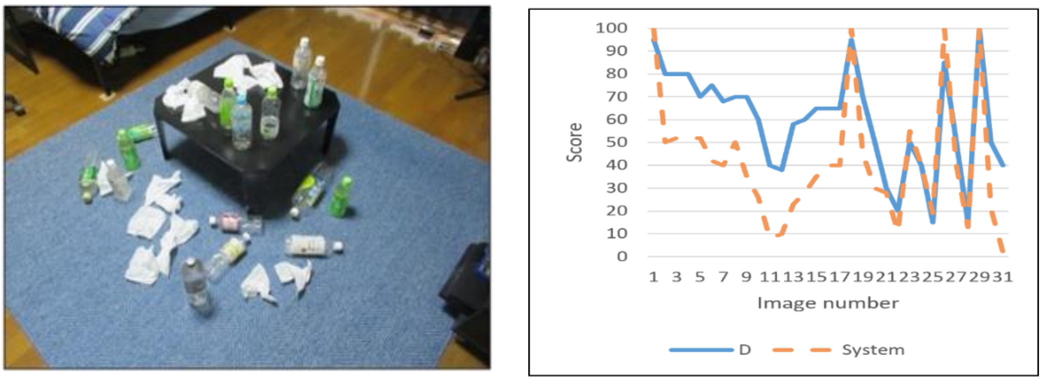


(a) Room image 28 (b) Scoring Result

Fig. 9. C scoring results

3.6. Subject D

Subject D's score are similar to the system's scoring results. The transition generally resembles subject A's but subjects A and D rated room image No. 31 cleaner than the system. In the target image, trash is concentrated in the center of the room. On the other hand, there are almost no objects at the periphery of the image, and the floor surface is clean. Therefore, the subject evaluated the room as clean because few objects were on the floor. The scoring results for subject D and the system are shown in Fig. 10.



(a) Room image 31 (b) Scoring Result

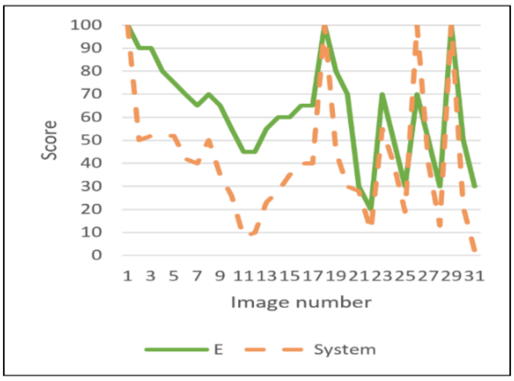
Fig. 10. D Scoring

3.7. Subject E

Subject E tended to rate the room image as cleaner than the system, but rated room image No. 26 as dirtier than the system. This is because the paper bag and hair dryer at the periphery of the image affected the subject. Subject E likely judged that the paper bag and hair dryer were left on the floor and gave a low score. On the other hand, paper bags and hair dryers were not subject to scoring by the system, so high scores were output.

The scoring results for E and the system are shown in Fig. 11.

(a) Room image 26



(b) Scoring Result

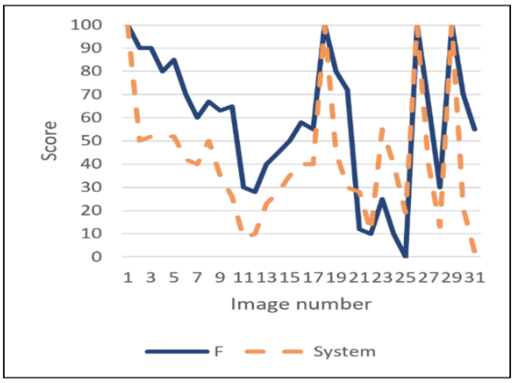
Fig. 11. E Scoring

3.8. Subject F

Subject F rated room image 23 as cleaner than the system, but rated room image 23 as dirtier. This is because the pods and cases on the desk in the center of the image affected the subjects. No objects are scattered around the desk, and much of the floor is exposed. However, because tissue waste and pods were clustered on the desktop, it is likely that Subject F judged the room to be dirty and gave a low score. On the other hand, since cases are not subject to scoring by the system, high scores were output.

The scoring results for F and the system are shown in Fig. 12.

(a) Room image 23



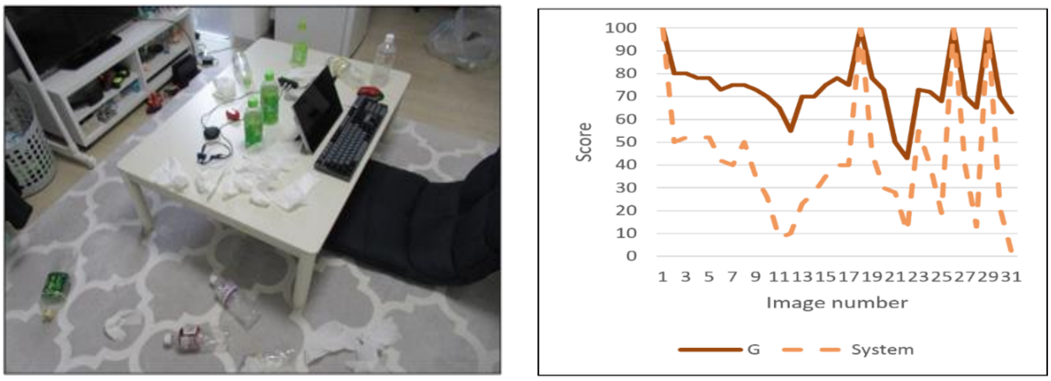
(b) Scoring Result

Fig. 12. F Scoring

3.9. Subject G

Subject G rated the system as being cleaner overall. Room 12 is an example of a room image with particularly different scores. The possible reasons for the difference in scores are the object being partly on the floor, and stored inside the shelf. Plastic bottles and tissue trash are scattered on the desktop, but on the floor, there is only trash in the front area. Therefore, it is thought that Subject G evaluated the area as relatively clean because there were almost no objects except the desk, and other objects were similarly stored on the shelves. On the other hand, the system outputs a low score due to the number of objects and false detection of tissue waste. Due to the above factors, the score for room image No. 12 was significantly different.

The scoring results of G and the system are shown in Fig. 13.



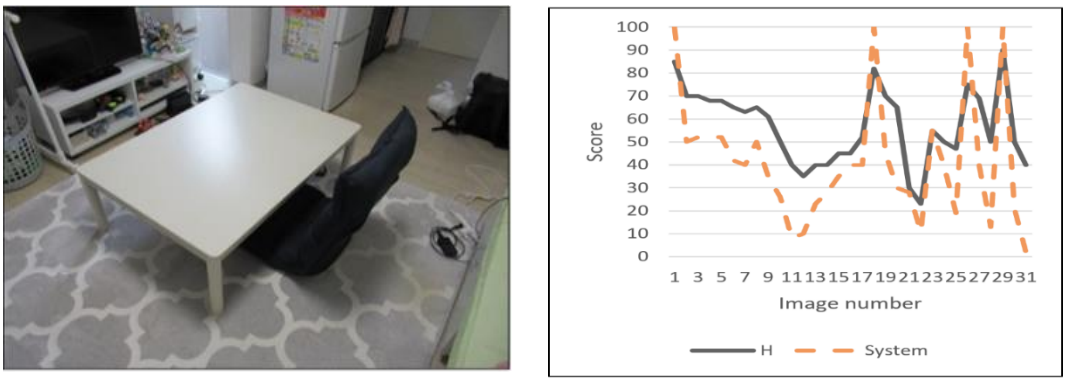
(a) Room image 23 (b) Scoring Result

Fig. 13. G Scoring

3.10. Subject H

There is a tendency for room images up to number 17 to be rated as cleaner than the system. However, the first image is rated as dirtier than the system. The cause is the backpack and plastic bag in the upper right corner of the image and the game equipment on the floor near the TV stand. In the target image, most objects are stored on shelves, but only game equipment, backpacks, and plastic bags are left on the floor. Therefore, subject H's score may have been slightly lower due to these factors.

The scoring results for H and the system are shown in Fig. 14.



(a) Room image 17 (b) Scoring Result

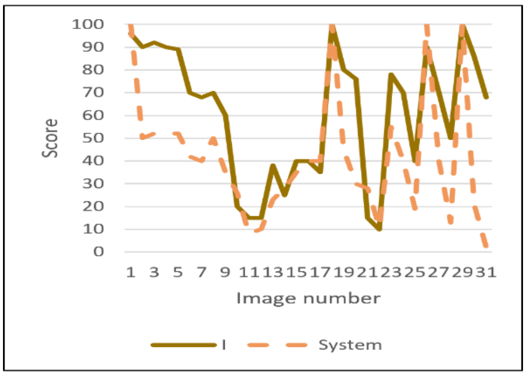
Fig. 14. H Scoring

3.11. Subject I

Subject I scored similarly to the system's scoring results. Among them, number 20 is cited as a room image with a large score difference. A possible reason for the large score difference is the number of objects on the floor. Most of the floor is exposed due to the no objects on the floor. Therefore, it is thought that Subject I evaluated that the room was clean. A higher score than the system was assigned due to the above factors.

The scoring results for I and the system are shown in Fig. 15.

(a) Room image 20



(b) Scoring Result

Fig. 15. I Scoring

3.12. Measurement Results

Table 2 shows the degree of similarity between each subject in the experiment and the system's scoring results. Furthermore, the similarity is the value calculated using equation (3) rounded to the second decimal place. The similarity by subject was over 70% except for subject G, and notably over 80% for subject H. Therefore, relatively good results have been obtained.

Table 2. Degree of similarity

SubjectScore	Degree of similarity(%)
A	79.1
B	73.9
C	79.4
D	79.7
E	77.2
F	77.1
G	67.5
H	80.4
I	78.0

$$\text{Degree of similarity}[\%] = 100 - \frac{\sum_{i=1}^{n=31}|sys_{score}*human_{score}|}{31} \tag{3}$$

where:

sys¬score: systemScore

humanscore: SubjectScore

4. Conclusion

In this work, leveraging AI, we proposed a system that can be used to determine the degree of clutter in a room to help caregivers determine when a visit to an independently living elderly person's home is necessary. Using the improved YOLOv5, object detection was performed on 31 room images. An overhead camera detects items like clothing, laptops, garbage, etc, determines their position, and judges whether items are collectively stored. This information is then used to determine the level of clutter in a room. A comparison was made with the scoring results performed by 9 people. The effectiveness of the proposed method was confirmed using the experimental results. In a nutshell, despite the subjectivity of the 9 human evaluators, the proposed system produced promising generalization. Future challenges include bringing the system's scoring results closer to human results and collecting more scoring data from more rooms. Moreover, it is necessary to find a new method to calculate the room clutter since the method relies on some intuitively decided parameters.

References

[1] H. A. Abbass, "Social Integration of Artificial Intelligence: Functions, Automation Allocation Logic and Human-Autonomy Trust," *Cognit. Comput.*, vol. 11, no. 2, pp. 159–171, Apr. 2019, doi: 10.1007/s12559-018-9619-0.

[2] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[3] P. Dhamija and S. Bag, "Role of artificial intelligence in operations environment: a review and bibliometric analysis," *TQM J.*, vol. 32, no. 4, pp. 869–896, Jul. 2020, doi: 10.1108/TQM-10-2019-0243.

[4] M. Zhao, J. Gao, M. Li, and K. Wang, "Relationship Between Loneliness and Frailty Among Older Adults in Nursing Homes: The Mediating Role of Activity Engagement," *J. Am. Med. Dir. Assoc.*, vol. 20, no. 6, pp. 759–764, Jun. 2019, doi: 10.1016/j.jamda.2018.11.007.

[5] S. Padhan, A. Mohapatra, S. K. Ramasamy, and S. Agrawal, "Artificial Intelligence (AI) and Robotics in Elderly Healthcare: Enabling Independence and Quality of Life," *Cureus*, vol. 15, no. 8, Aug. 2023, doi: 10.7759/cureus.42905.

[6] G. Cheung, V. Wright-St Clair, E. Chacko, and Y. Barak, "Financial difficulty and biopsychosocial predictors of loneliness: A cross-sectional study of community dwelling older adults," *Arch. Gerontol. Geriatr.*, vol. 85, p. 103935, Nov. 2019, doi: 10.1016/j.archger.2019.103935.

[7] A. Frolli, M. C. Ricci, A. Cavallaro, S. Rizzo, and F. Di Carmine, "Virtual Reality Improves Learning In Children With Adhd," In *Edulearn21 Proceedings*, Jul. 2021, vol. 1, pp. 9229–9236, doi: 10.21125/edulearn.2021.1859.

[8] A. Frolli *et al.*, "Children on the Autism Spectrum and the Use of Virtual Reality for Supporting Social Skills," *Children*, vol. 9, no. 2, p. 181, Feb. 2022, doi: 10.3390/children9020181.

[9] M. Di Gregorio, M. Romano, M. Sebillo, and G. Vitiello, "Dyslexeasy-App to Improve Readability through the Extracted Summary for Dyslexic Users," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, Jan. 2022, pp. 1–6, doi: 10.1109/CCNC49033.2022.9700618.

[10] "Current status of the nursing," *Ministry of Health, Labor and Welfare*, p. 24, 2018.[Online]. Available at: https://www.mhlw.go.jp/content/12602000/000489026.pdf.

[11] D. Chishima, M. Nihei, And M. Kamata, "2G5-3 The relationship between Mess and Stress," *Japanese J. Ergon.*, vol. 54, no. Supplement, pp. 2G5-3-2G5-3, Jun. 2018, doi: 10.5100/jje.54.2G5-3.

[12] I. Logothetis, P. Rani, S. Sivasothy, R. Vasa, and K. Mouzakis, "Smart Home Goal Feature Model -- A guide to support Smart Homes for Ageing in Place," Nov. 2023. [Online]. Available at: https://arxiv.org/abs/2311.09248v1.

[13] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9905 LNCS, Springer Verlag, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[14] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 2980–2988, Dec. 2017, doi: 10.1109/ICCV.2017.322.

[15] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” in *Computer Vision and Pattern Recognition*, Apr. 2018, pp. 1–6. [Online]. Available at: https://arxiv.org/abs/1804.02767v1.

[16] R. H. Krishnan, B. A. Naik, G. G. Patil, P. Pal, and S. K. Singh, “AI Based Autonomous Room Cleaning Bot,” in *2022 International Conference on Futuristic Technologies (INCOFT)*, Nov. 2022, pp. 1–4, doi: 10.1109/INCOFT55651.2022.10094492.

[17] V. Suryani, K. Agustriana, A. Rakhmatsyah, and R. R. Pahlevi, “Room cleaning robot movement using A* algorithm and imperfect maze,” *J. INFOTEL*, vol. 15, no. 1, pp. 75–81, Mar. 2023, doi: 10.20895/infotel.v15i1.901.

[18] Y. Liu, Z. Shao, Y. Teng, and N. Hoffmann, “NAM: Normalization-based Attention Module,” *arxiv Artif. Intell.*, no. NeurIPS 2021, pp. 1–5, 2021, [Online]. Available at: http://arxiv.org/abs/2111.12419.

[19] M.-H. Guo *et al.*, “Attention mechanisms in computer vision: A survey,” *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, Sep. 2022, doi: 10.1007/s41095-022-0271-y.

[20] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[21] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, vol. 07-09-Janu, pp. 1027–1035, 2007, [Online]. Available at: https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf.

[22] K. Krishna and M. Narasimha Murty, “Genetic K-means algorithm,” *IEEE Trans. Syst. Man Cybern. Part B*, vol. 29, no. 3, pp. 433–439, Jun. 1999, doi: 10.1109/3477.764879.

[23] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, “Focal and efficient IOU loss for accurate bounding box regression,” *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022, doi: 10.1016/j.neucom.2022.07.042.

[24] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, pp. 12993–13000, Apr. 2020, doi: 10.1609/aaai.v34i07.6999.

[25] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, “UnitBox,” in *Proceedings of the 24th ACM international conference on Multimedia*, Oct. 2016, pp. 516–520, doi: 10.1145/2964284.2967274.