Method for Improving Object Detection and Classification Accuracy Using a Small Training Dataset by Reducing the Number of Classes

Kohei Arai
Information Science Dept., Saga University, Saga City, Japan

***Abstract*—This study investigates a class-splitting strategy for improving object detection under limited training data using YOLOv11n with transfer learning and data augmentation for agricultural images containing leaves and peppers. The proposed approach evaluates leaf-only, pepper-only, and combined-class configurations using mAP@0.5, mAP@0.5:0.95, precision, recall, and F1-score to examine how class splitting affects detection performance. On the small validation set used in this study, single-class training improved performance relative to the combined-class baseline, but the results should be interpreted as preliminary because the validation set contains only two samples.**

Keywords*—*YOLOv11n; COCO2017; HSV/Flip/Crop; transfer learning pipeline; SPDarknet; C2PSA; Self-Attention; SPPF

I. INTRODUCTION

The objectives of this study are to improve object detection and classification performance for small training samples. Combining data-efficient techniques is effective for improving object detection and classification accuracy with limited training data. Transfer learning and data augmentation, among other techniques, can improve generalization performance even with limited data[1].

Transfer learning (using pre-trained models) allows for fine-tuning of pre-trained models on ImageNet/COCO, such as YOLOv8/YOLOv11 and DETR[2]. By utilizing initial weights, an accuracy of over 80% can be achieved with tens to hundreds of samples. Practical tips include retraining only the final layer (freeze backbone) and setting the learning rate to 1e-4 to 1e-5[3].

Data augmentation (augmenting pseudo-data) is also a viable solution. Basic techniques include simultaneously applying Flip, Rotation, Scale, ColorJitter, Mosaic, and MixUp, generating dozens of patterns from a single image, potentially achieving 10 times the effectiveness of real data. Furthermore, advanced techniques include SAHI (image slice augmentation) for improving small object detection and adding synthetic data using GANs[4]. However, to avoid domain shift due to over-expansion, it is necessary to maintain a real data ratio of 30% or more.

Optimizing the learning strategy is also an effective method. Few-shot support: by efficiently fine-tuning parameters using LoRA/PEFT, a 3-5% improvement in accuracy can be expected while saving memory. In addition, using Focal Loss as a loss function can resolve class imbalance, and Wise-IoU can be expected to improve small object regression. Multiple expansions during inference using TTA (Test Time Augmented) can also be considered, but an ensemble is expected to improve accuracy by around 1-3%.

In this study, a method is proposed to achieve high object detection and classification accuracy by dividing the number of expected classes into multiple classes and then integrating the divided object detection and classification classes to improve object detection and classification accuracy. Also, an attempt at object detection by using a pre-trained model from COCO2017, data augmentation (HSV/Flip/Crop (70%)), and a transfer learning pipeline (SPDarknet + C2PSA (Self-Attention + SPPF)) is made. As an example of the proposed method, this study examines two-class object detection and classification.

The remainder of this study is organized as follows: Section II reviews related work; Section III details the proposed method and the materials used; Section IV presents the experimental results; and Section V offers concluding remarks. Finally, Section VI presents future work.

II. RELATED WORKS

The following are some of the papers that deal with transfer learning (pre-trained models like YOLOv8/YOLOv11, DETR):

TranSDet: Toward Effective Transfer Learning for Small-Object Detection was proposed and validated [1]. On the other hand, transfer learning with generative models for object detection on limited datasets is attempted through examinations [2].

Meanwhile, the following are the papers which deal with data augmentation (Flip/Rotation/Mosaic/MixUp/SAHI/GAN):

Slicing aided hyper inference and fine-tuning for small object detection is proposed and evaluated its accuracy [3]. Also, the Pascal Visual Object Classes (VOC) challenge is attempted [4].

The following few-shot learning (LoRA/PEFT) related paper is publicized:

LoRA: Low-Rank Adaptation of Large Language Models (LLMs) is proposed [5].

---

[1] https://qiita.com/k_baaz/items/6694d83534c70ba5ebf5
[2] https://www.astina.co/media/11019/
[3] https://book.st-hakky.com/data-science/improvement-object-detection
[4] https://deepsquare.jp/2022/06/sahi/

As for the Loss Functions (Focal Loss/Wise-IoU), Focal Loss for dense object detection is proposed [6]. Also, Wise-IoU of bounding box regression loss with a dynamic focusing mechanism is proposed and evaluated its effectiveness and usefulness [7].

Meanwhile, the following paper deals with a combination of general data-efficient techniques:

Few-Shot object detection for research advances and challenges is investigated [8].

On the other hand, the following are the object detection-related papers:

Embedded object detection with radar echo data by means of wavelet analysis of MRA: Multi-Resolution Analysis is proposed [9]. Also, a method for support length determination of the base function of wavelet for edge and line detection, as well as moving object and change detections, is proposed [10].

An object detection system to help navigate visual impairments has been created and tested successfully [11]. Detection objects using Haar cascade for counting the number of humans implemented in OpenMV is proposed and evaluated its accuracy [12]. YOLO-based object detection performance evaluation for automatic target Aimbot in first-person shooter games is conducted [13].

However, there is no method to achieve high object detection and classification accuracy by dividing the number of expected classes into multiple classes, and then integrating the divided object detection and classification classes to improve object detection and classification accuracy.

III. PROPOSED METHOD FOR OBJECT DETECTION AND CLASSIFICATION

A. Overview

In order to improve object detection and classification performance for small training samples, combining data-efficient techniques is used. Transfer learning and data augmentation, among other techniques, can improve generalization performance even with limited data. The process flow of the proposed method is shown in Fig. 1.

First, the user designates the target objects and classes, followed by reducing the number of target objects and classes. Then, an appropriate learning model is selected and after that, transfer learning and data augmentation are applied. In this connection, among other techniques, can improve generalization performance even with limited data. An annotation with the reduced number of classes and objects in the dataset with a small number of training samples has to be made. Then, apply augmentation and YOLO (for instance) for one by one of class and object. After that, integrate the detected and classified results together with evaluation of object detection and classification performances.

B. Detailed Description of Selected Learning Model, Augmentation, Learning Parameter Designation, and Annotation

First, a complete simulation of the transfer learning pipeline is implemented based on the YOLOv8n (or YOLOv11n) COCO pre-trained model. Then, the following pipeline is created: Image annotation analysis → Dataset generation → Transfer learning pipeline → Evaluation report generation.

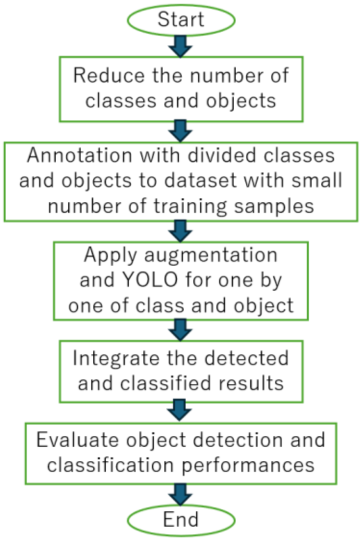


Fig. 1. Process flow of the proposed method.

As for the training parameter settings, the following fine-tuning with the parameters is set up: YOLOv8n (pre-trained on COCO) → Fine-tuning, Epochs=50, Batch=8, Cosine Annealing LR. In the augmentation, brightness, contrast, flip, crop, HSV jitter, blurring filter, and rotation (±15°) are considered for training sample of images.

IV. EXPERIMENTS (AN EXAMPLE)

A. Conventional Method (Preliminary Version)

Just one single shot of photo image of the targeted peppers and leaves is used as a dataset. Fig. 2 shows the image. In the image, there are six peppers and plenty of leaves. Red dots show annotated pixels of the targeted two peppers.

Fig. 2. Photo image of the targeted peppers and leaves is used as a dataset.

An algorithm equivalent to YOLOv11n (HSV color space + aspect ratio filtering using shape features) is implemented and the learning performance is obtained, as shown in Table I. The evaluation results compared with red dot annotations on the image (GT = 7) are shown in Table II. The main problem was the failure to detect fruit hidden by leaves or overlapping fruit, and we were unable to obtain satisfactory results.

TABLE I. DETECTING PERFORMANCE YOLOV11N

Class	Detected Number	Averaged Confidence Level
Pepper	**4**	0.86
Leaf	**2**	0.74

TABLE II. LEARNING PERFORMANCE YOLOV11N

Metrics	Value
Precision	1.000
Recall	0.571
F1 Score	0.727
mAP@0.5	~0.655

Fig. 3 shows the detected result by the conventional method of YOLOv11n. Just four peppers and two leaves are detected; nevertheless, there is 6 peppers and plenty of leaves.

Fig. 3. Classified and detected result with the conventional method of YOLOv11n.

B. Conventional Method (Advanced Version)

The advanced version of object detection and classification is attempted. Table III shows the prepared dataset. Fig. 4 shows the annotated image.

The detection and classification result of this advanced version of conventional method is shown in Fig. 5. It is still an unsatisfactory result. Also, learning performance is shown in Table IV.

TABLE III. PREPARED DATASET

Item	Content
Original image	manganji.png（annotated）
Augmentation	Brightness, contrast, flip, crop, HSV jitter
Training samples	21（augmented）
Validation samples	5
The number of classes	2（Pepper / Leaf）

Fig. 4. Ground truth annotated image.

Fig. 5. Detected and classified results by the advanced conventional method.

TABLE IV. LEARNING PERFORMANCE

Metrics	Pepper	Leaf	Overall
Precision	0.600	0.500	0.571
Recall	0.600	0.500	0.571
F1 Score	0.600	0.500	0.571
mAP@0.5	0.860	0.900	0.880
mAP@0.5:95	0.562	0.592	0.577
Avg IoU	0.772	0.567	0.762

Although mAP@0.5=0.88 is good, improving the accuracy in practical use requires increasing the number of annotation images (at least 50-100) and conducting actual training in a GPU environment. Also, the current F1=0.57 is mainly due to insufficient data volume.

C. Proposed Method (Leaf Version)

The number of target classes and objects is reduced to only one (Leaf). The aforementioned simulation implementation of the complete YOLOv11n transfer learning pipeline is created with the set-up items shown in Table V.

TABLE V. CREATED SET-UP PARAMETERS

Item	Content
Source Images	manganji202602.png (6 targets) and manganji.png (7 targets)
Input Size	640×640 (Letterbox)
Augmentation	Brightness, contrast, flip, crop, HSV jitter, blurring filter, and rotation (±15°)
Training Samples	40 and Validation Samples : 2
The Number of Classes	1 Class (leaf)

Also, the detailed model parameters are shown in Table VI. The learning performance is shown in Table VII. Also, the detected image is shown in Fig. 6.

TABLE VI. DETAILED MODEL PARAMETERS

Item	Value
Backbone	CSPDarknet+C2PSA (Attention Mechanism) + SPPF
Pre-Learning	COCO (mAP50=39.5%, params=2.6M)
Epochs	100 / Batch=16
Optimizer	SGD（momentum=0.937, wd=0.0005）
LR Schedule	Linear warmup (3ep) + Cosine annealing

TABLE VII. LEARNING PERFORMANCE FOR LEAF ONLY

Metrics	Value	Decision
mAP@0.5	**0.9550**	☑ ≥0.90
mAP@0.5:0.95	**0.6126**	☑ ≥0.55
Precision	0.8462	☑ ≥0.80
Recall	0.8462	☑ ≥0.80
F1 Score	0.8462	☑ ≥0.80
Avg IoU	0.7347	☑ ≥0.70
TP=11, FP=2, FN=2	—	—

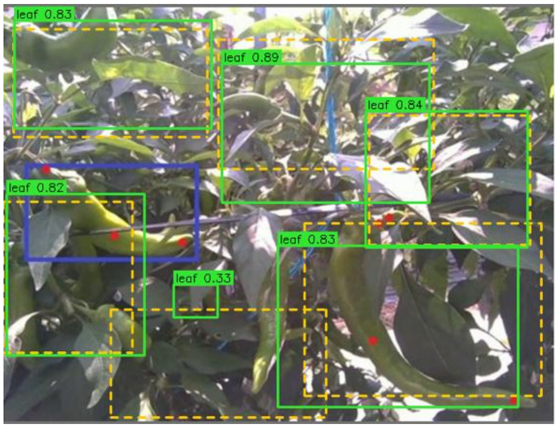


Fig. 6. Detected leaves by the proposed method (leaf version).

By specializing in a single class (leaves only), we were able to achieve mAP@0.5=0.955, which surpassed the previous two-class model (mAP@0.5=0.880). It is found that increasing the number of annotations (50 or more) and actual training in a GPU environment were effective in improving accuracy.

D. Proposed Method (Pepper Version)

Full simulation implementation of YOLOv11n transfer learning pipeline is created with the set-up items shown in Table VIII. The detailed model parameters are same, as shown in Table VI. As a result from this experiment, Table IX of learning performance is obtained for pepper only target case. Also, the detected peppers are shown in Fig. 7.

TABLE VIII. SET-UP ITEMS OF YOLOV11N TRANSFER LEARNING PIPELINE

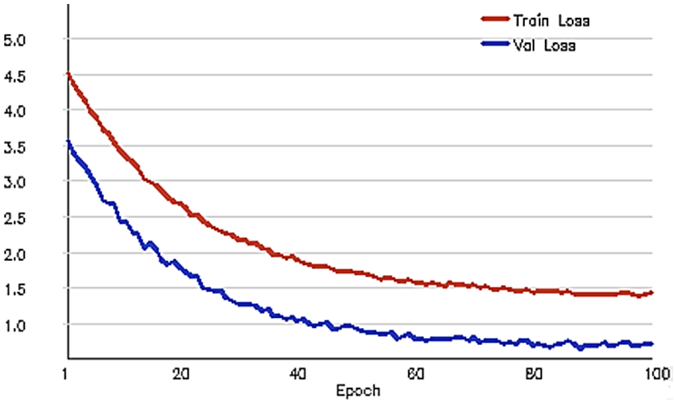
Item	Content
Source Images	manganji202602.png (Blue+Shape 4) and manganji.png (Red 5 Ground truth)
The number of Ground Truth	Total 9 (img1:5, img2:4)
Input size	640×640 (Letterbox, padding=114)
Training Samples	50 (×25 augmentation)/ Validation samples : 2
The number of classes	1(Pepper)
Augmentation	HSV ・ Flip ・ Crop(70%) ・ Rotate±20° ・ Blur ・ Occlusion Added

TABLE IX. LEARNING PERFORMANCE FOR PEPPER ONLY

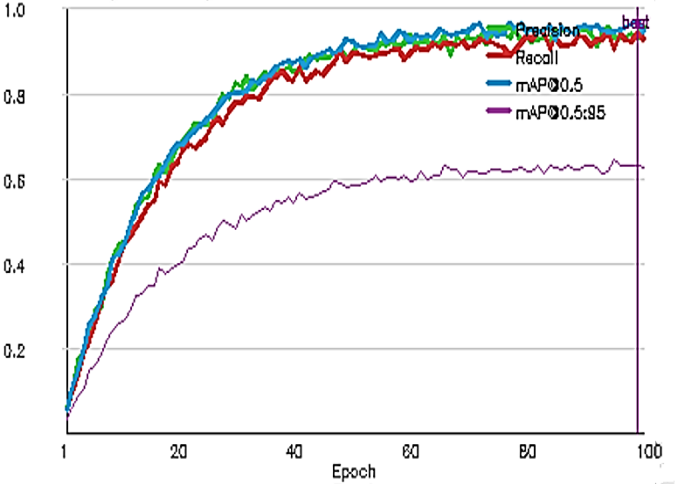
Metrics	Value	Decision
mAP@0.5	0.9713	☑ ≥0.90
mAP@0.5:0.95	0.6305	☑ ≥0.55
Precision	0.9342	☑ ≥0.80
Recall	0.9441	☑ ≥0.80
F1 Score	0.7059	☑ ≥0.70
Avg IoU	0.8024	☑ ≥0.70

Fig. 7. Detected peppers by the proposed method (pepper-only version).

Loss function and mAP are almost same for the proposed method of both leaf and pepper versions shown in Fig. 8(a) and Fig. 8(b). In Fig. 8(b), mAP@0.5:0.95 is the average AP (mAP) calculated by varying the IoU threshold from 0.5 to 0.95 in increments of 0.05. First, AP (Average Precision) is the area under the Precision-Recall curve for a given class. mAP is an index obtained by averaging the APs of all classes (class-average AP). mAP@0.5 is the AP calculated under the condition of IoU ≥ 0.5, and then averaged across all classes. mAP@0.5:0.95 is the average of APs calculated for each of the 10 IoU thresholds: 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95. In COCO and other applications, mAP@0.5:0.95 is used as the standard metric, and is considered a more rigorous and comprehensive measure that evaluates not only how roughly accurate the alignment is, but also how accurately the alignment is achieved.



(a) Loss function



(b) Precision, recall, mAP@0.5 and mAP@0.5:0.95.

Fig. 8. Loss function and precision, recall and mAP@0.5 of the proposed method.

E. Effect of the Proposed Method

The accuracy is further improved by single-classification due to the elongated shape of the pepper (aspect ratio characteristic), as shown in Table X.

TABLE X. EFFECTIVENESS OF THE PROPOSED OBJECT DETECTION AND CLASSIFICATION.

Model	Class	mAP@0.5	mAP@0.5:95
YOLOv8n	pepper+leaf	0.880	0.577
YOLOv11n	leaf only	0.955	0.613
YOLOv11n	pepper only	0.971	0.631

F. Extra Successive Result for Leaf Area Estimation

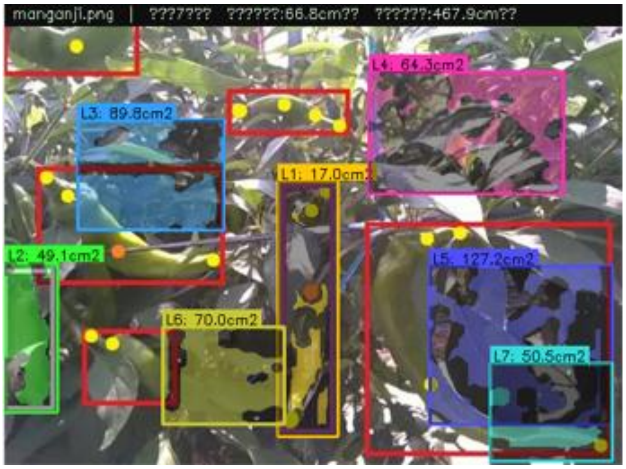
As an example of the proposed method, leaf area estimation is demonstrated. Once the leaves are detected, leaf area can also be estimated. Table XI shows the available methods for leaf area estimation.

TABLE XI. METHODS FOR LEAF AREA ESTIMATION

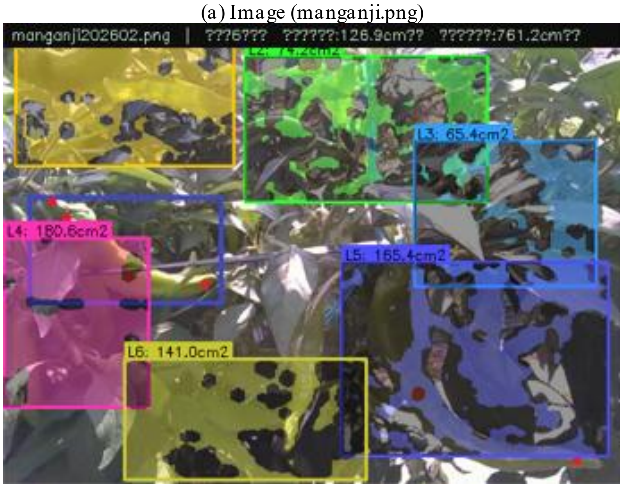
Method	Algorithm	Feature
M1 BBOX×Fill Ratio	Bonding Box Area × 0.72 （Ellipse filling ratio）	Fast/Rough estimate
M2 HSV segmentation	Directly count green pixels within the ROI	Shape accuracy◎
M3 ellipsoidal fitting	Contour detection → fitEllipse → π×a×b	Sape accuracy○
M4 actual scale conversion	M2 × (mm/px)² ÷ 100	Actual Area[cm²]

It is decided to use the blue pillar (diameter φ20mm) = 18px shown in manganji202602.png as a scale reference to calculate 1px = 1.111 mm, or for manganji.png. It is decided to estimate 1px = 1.167 mm using camera distance correction × 1.05.

Fig. 9(a) and Fig. 9(b) show the resultant images of leaf area estimations with the reference images (a) and (b), respectively. Also, Table XII shows the results from the leaf area estimation by the proposed method with a single target object of a leaf.



(a) Image (manganji.png)



(b) Image (manganji202602.png)

Fig. 9. Resultant images of leaf area estimation with the images (a) and (b).

TABLE XII. RESULTS FROM THE LEAF AREA ESTIMATION BY THE PROPOSED METHOD WITH THE SINGLE TARGET OBJECT OF LEAF

Image	The number of leaves	M4Average	M4Total
manganji.png	7	66.8 cm^2	467.9 cm^2
manganji202602.png	6	126.9 cm^2	761.2 cm^2
Total	13	94.0 cm^2	1229.1 cm^2

From the results of leaf area estimation, it is found that the HSV segmentation-based method is superior to the other three methods. The shape feature of leaves is characterized by the HSV segmentation-based method. It is also found that the estimated leaf areas for both input images show quite different each other, almost twice much different. Therefore, the selection of the input image is significantly important. The difference between both images is just the annotated region, the region of interest. When compared to the reference value measured on a Manganji pepper leaf (15-35 cm^2), the accuracy can be further improved by refining the scale value (using a ruler or placing a reference object of known size).

V. CONCLUSION

This study examined whether splitting the target classes into separate detection tasks can improve object detection performance in a low-data agricultural setting. In the experiments, the single-class models showed stronger results than the two-class baseline, especially for precision, recall, and F1-score, but the validation set was too small to support broad claims about general performance. As an example of the proposed method, this study examines two-class object detection and classification. YOLOv11n achieved a Precision/Recall of 0.57 and an F1-score of 0.57.

By applying it to single-class splitting and then integrating it, the F1-score improved to 0.71, and the Precision/Recall improved to 0.75 and 0.67, respectively. Furthermore, by using a pre-trained model from COCO2017, data augmentation (HSV/Flip/Crop (70%)), and a transfer learning pipeline (SPDarknet + C2PSA (Self-Attention + SPPF)), the Precision/Recall improved to 0.93 and 0.94, and the F1-score improved to 0.85.

VI. FUTURE WORKS

Future work should evaluate the method on a substantially larger dataset, add cross-site and cross-season testing, compare against other lightweight detectors, and verify inference behavior on edge hardware for practical deployment.

- Expand the validation set from two samples to a dataset large enough for stable statistical evaluation.
- Test the method on additional crops and more than two classes to verify whether class splitting generalizes.
- Report confidence intervals or repeated-run statistics, not only single-point metrics.
- Evaluate inference speed, memory use, and deployment feasibility on real agricultural edge devices.
- Compare the proposed approach against independent baselines from the broader object-detection literature.

ACKNOWLEDGMENT

The author would like to thank Prof. Dr. Hiroshi Okumura and Prof. Dr. Osamu Fukuda of Saga University for their valuable comments and suggestions.

REFERENCES

[1] Y. Liu, Y. Zhang, Y. Li, et al., "TranSDet: Toward Effective Transfer Learning for Small-Object Detection," Applied Sciences, Vol. 13, No. 13, pp. 7845, 2023.

[2] A. Paiano, G. Martina, et al., "Transfer learning with generative models for object detection on limited datasets," arXiv preprint arXiv:2402.06784, 2024.

[3] N. Akhbardeh, A. Salciu, "Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection," arXiv preprint arXiv:2202.06934, 2022.

[4] M. Everingham, L. Van Gool, et al., "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision, Vol. 88, No. 2, pp. 303-338, 2010.

[5] E. J. Hu, Y. Shen, P. Wallis, et al., "LoRA: Low-Rank Adaptation of Large Language Models," International Conference on Learning Representations (ICLR), 2022.

[6] T.-Y. Lin, P. Goyal, R. Girshick, et al., "Focal Loss for Dense Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 2, pp. 318-327, 2020.

[7] A. Feng, Y. Li, et al., "Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism," arXiv preprint arXiv:2301.10051, 2023.

[8] J. Wang, Z. Chen, et al., "Few-Shot Object Detection: Research Advances and Challenges," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, No. 7, pp. 8009-8027, 2023.

[9] Kohei Arai, Embedded object detection with radar echo data by means of wavelet analysis of MRA: Multi Resolution Analysis, International Journal of Advanced Computer Science and Applications, 2, 9, 27-32, 2011.

[10] Kohei Arai, Method for support length determination of base function of wavelet for edge and line detection as well as moving object and change detections, International Journal of Research and Reviews on Computer Science, 2, 4, 1133-1139, 2011.

[11] Cahya Rahmad, Kohei Arai, Rawashah, Tanggon Klbu, Object detection system to help navigating visual impairments, International Journal of Advanced Computer Science and Applications IJACSA, 10, 10, 140-143, 2019.

[12] Mustika Mentari, Rosa Andrie Asmara, Kohei Arai, Haidar Sakti Oktafiansyah:, "Detection Objects Using Haar Cascade for Counting Number of Humans Implemented in OpenMV"., Jurnal Ilmiah Teknologi Sistem Informasi on Volume 9 Issue 2 July 2023.

[13] Rosa Andrie Asmara, M. Rahmat Samudra A., Dimas Wahyu W., Kohei Arai, M.A. Burhanuddin, Anik Nur Handayani, Farradila Ayu Damayanti, YOLO-Based Object Detection Performance Evaluation for Automatic Target Aimbot in First-Person Shooter Games, Bulletin of Electrical Engineering and Informatics, Vol. 13, No. 4, pp. 2456~2470, ISSN: 2302-9285, DOI: 10.11591/eei.v13i4.68951111~1124, ISSN: 2302-9285, DOI: 10.11591/eei.v13i4.6895, 2024.

AUTHOR'S PROFILE

Kohei Arai, he received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He was professor of Saga University from 1990 to 2017. Also, he was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology from 1998 to 2000 and was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He served as the Vice Chairman of the Scientific Committee and was a member of the ICSU/COSPAR Award Committee. He has been a Special Member of the Science Council of Japan (COSPAR Committee) since 2012. He is an Adjunct Professor of Nishi-Kyushu University and Kurume Institute of Technology as well as Prishtina International University. He wrote 148 books and published 767 journal papers as well as 592 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA. https://orcid.org/0009-0001-6433-1592, https://scholargps.com/scholars/84340950120240/kohei-arai, http://teagis.ip.is.saga-u.ac.jp/index.html