

jowua paper

10881-29890-1-RV

 icst

Document Details

Submission ID

trn:oid::3618:117588619

Submission Date

20 Oct 2025, 13:39 GMT+7

Download Date

20 Oct 2025, 13:46 GMT+7

File Name

10881-29890-1-RV.docx

File Size

9.9 MB

10 Pages

5,175 Words

30,206 Characters

16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

- 74** Not Cited or Quoted 14%
Matches with neither in-text citation nor quotation marks
- 9** Missing Quotations 2%
Matches that are still very similar to source material
- 0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation
- 0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 7% Internet sources
- 9% Publications
- 10% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 74 Not Cited or Quoted 14%**
Matches with neither in-text citation nor quotation marks
- 9 Missing Quotations 2%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 7% Internet sources
- 9% Publications
- 10% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers		
	Universitas Prima Indonesia on 2025-06-14		1%
2	Student papers		
	Universitas Khairun on 2024-04-18		1%
3	Internet		
	arxiv.org		1%
4	Internet		
	www.mdpi.com		<1%
5	Publication		
	"A Semi-Supervised Method for Steel Surface Defect Detection Based on Soft-teac...		<1%
6	Student papers		
	Adama Science and Technology University on 2023-06-20		<1%
7	Publication		
	"Medical Image Computing and Computer Assisted Intervention - MICCAI 2020", ...		<1%
8	Publication		
	"Computer Vision - ECCV 2022", Springer Science and Business Media LLC, 2022		<1%
9	Internet		
	jutif.if.unsoed.ac.id		<1%
10	Student papers		
	Liverpool John Moores University on 2024-03-30		<1%

11	Internet	proceedings.stis.ac.id	<1%
12	Student papers	Beirut Arab University on 2025-03-17	<1%
13	Publication	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dharendra Kumar Shukla. "Re..."	<1%
14	Publication	Yuqiu Li, Jian Xue, Ke Lu. "Semi-supervised Object Detection for Remote Sensing I..."	<1%
15	Publication	Martin Wutke, Clara Lensches, Ulrich Hartmann, Imke Traulsen. "Towards autom..."	<1%
16	Student papers	National College of Ireland on 2023-08-13	<1%
17	Publication	Ramiz Yilmazer, Derya Birant. "Shelf Auditing Based on Image Classification Usin..."	<1%
18	Student papers	University of Bolton on 2025-05-11	<1%
19	Internet	ejournal.uin-suska.ac.id	<1%
20	Internet	jurnal.iaii.or.id	<1%
21	Internet	mdpi-res.com	<1%
22	Student papers	Anadolu University on 2023-05-09	<1%
23	Publication	Yue He, QinHong Du, Huiyu Wu, Yukun Du, Juan Xu, Yongming Xi, Huan Yang. "M..."	<1%
24	Internet	jst-ud.vn	<1%

25	Internet	pmc.ncbi.nlm.nih.gov	<1%
26	Publication	Jiannan Zhu, Chen Fan, Minglei Yang, Feng Qian, Vladimir Mahalec. "A semi-super...	<1%
27	Student papers	University of Nottingham on 2024-09-06	<1%
28	Internet	franky07724-57962.medium.com	<1%
29	Publication	Amitesh Badkul, Li Xie, Shuo Zhang, Lei Xie. "eMOAIC: Multi-modal Out-of-distrib...	<1%
30	Student papers	Sriwijaya University on 2024-02-11	<1%
31	Student papers	University of Technology on 2022-03-27	<1%
32	Publication	Xiangxin Zhao, Xin Wang. "Unsupervised domain adaptation based fracture segm...	<1%
33	Internet	ar5iv.org	<1%
34	Internet	kudos.dfo.no	<1%
35	Publication	"Information and Communication Technology", Springer Science and Business M...	<1%
36	Publication	Bo Xiao, Yuxuan Zhang, Yuan Chen, Xianfei Yin. "A semi-supervised learning dete...	<1%
37	Student papers	Cork Institute of Technology on 2021-02-21	<1%
38	Student papers	National Taiwan University on 2023-07-24	<1%

39	Publication	Sijin Chen, Yingyun Yang, Yan Hua. "Semi-Supervised Active Learning for Object D...	<1%
40	Student papers	University of East London on 2025-05-08	<1%
41	Student papers	University of Warwick on 2022-10-03	<1%
42	Student papers	University of Warwick on 2025-08-26	<1%
43	Publication	Yu-Shu Ni, Wei-Lun Chen, Yi Liu, Ming-Hsuan Wu, Jiun-In Guo. "Optimizing Autom...	<1%
44	Internet	constellation.uqac.ca	<1%
45	Internet	journal.unnes.ac.id	<1%
46	Student papers	Deakin University on 2022-04-11	<1%
47	Publication	Ponduri Vasanthi, Laavanya Mohan. "Ensemble of ghost convolution block with n...	<1%
48	Student papers	University of Queensland on 2023-06-05	<1%
49	Student papers	University of Sydney on 2024-11-04	<1%
50	Publication	Cong Chen, Shouyang Dong, Ye Tian, Kunlin Cao, Li Liu, Yuanhao Guo. "Temporal ...	<1%
51	Publication	Guangchen Chen, Benjie Shi, Yinhui Zhang, Zifen He, Pengcheng Zhang. "CGSNet: ...	<1%
52	Publication	Yeqin Shao, Chang Lv, Ruowei Zhang, He Yin, Meiqin Che, Guoqing Yang, Quan Jia...	<1%

53

Publication

Yunqing Jiang, Sunyuan Qiang, Wuchen Li, Yanyan Liang. "LLM-DiffAug: Enhancin... <1%

JIKO (Jurnal Informatika dan Komputer)
 Vol. x, No. x, April 2025, pp. x-x
 DOI: 10.33387/jiko

Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
 p-ISSN: 2614-8897
 e-ISSN: 2656-1948

Implementation of Semi-Supervised Learning with YOLOv11 for On-Shelf Availability Detection of Retail

Author^{1*}, Author² (10pt, bold, and center)

¹ Department, Faculty, University, City, Postal code, Country (10pt)
² Department, Faculty, University, City, Postal code, Country (10 pt)
 Email: *¹author.one @xmail.ac.id, ²author2@xmail.ac.id (10pt)

(Received: dd mmm yyyy, Revised: dd mmm yyyy, Accepted: dd mmm yyyy)

Abstract (10pt)

On-Shelf Availability (OSA) is a critical aspect of retail operations that affects customer satisfaction and potential sales. Computer vision-based systems have emerged as a promising solution to monitor product availability on store shelves. However, their implementation faces the challenge of limited labeled data, which requires time-consuming manual annotation with precise bounding boxes. This study proposes a semi-supervised learning approach based on pseudo-labeling using the YOLOv11n architecture to address the scarcity of labeled data. We utilized a dataset of 918 retail product images with 174 classes, divided into four proportions of labeled data (20%, 40%, 60%, and 80%). The research stages included training a teacher model, generating pseudo-labels with a confidence threshold of 0.5, and training a student model using a combination of labeled and pseudo-labeled data. Experimental results show that this approach effectively improves detection performance. With 60% labeled data, the model achieved an mAP50 of 0.931 and an mAP50-95 of 0.864, along with high-quality pseudo-labels (F1-Score 0.727; IoU 0.819). This significant improvement indicates that pseudo-labels can enrich data variation without introducing excessive noise. The study demonstrates that semi-supervised learning can reduce dependence on large labeled datasets while offering a practical and efficient solution for OSA detection systems in retail environments.

Keywords: On-shelf Availability, Semi-Supervised Learning, Pseudo-labeling, Yolov11, Object Detection

This is an open access article under the [CC BY](#) license.



*Corresponding Author: Author1

1. INTRODUCTION

On-Shelf Availability (OSA) is a critical aspect of retail store operations, as it directly impacts customer satisfaction and sales potential. In this context, computer vision-based systems, particularly object detection from shelf images, have emerged as promising solutions for monitoring product availability, identifying shelf gaps, and supervising product placement on store shelves [1][2]. However, the implementation of such systems faces the challenge of requiring large amounts of accurately annotated training data. Unlike image classification, object detection not only focuses on recognizing object categories but also predicts the location of each object through the use of bounding boxes [3]. Therefore, it requires precise annotations in the form of bounding boxes for each object, which must be performed manually and is time-consuming [4]. This challenge

becomes increasingly complex in retail store scenarios, which typically involve packaging variations, visual similarities between products, and diverse lighting conditions and camera angles [5]. Furthermore, the requirement to include detailed product attributes such as brand, flavor, size, and type further increases the cost and complexity of the labeling process [6].

To address the challenges of limited labeled data and high annotation costs in object detection tasks, semi-supervised learning (SSL) approaches have gained significant attention in recent years [5]. SSL offers an efficient solution by leveraging a combination of a small portion of labeled data for model training [7]. This strategy enables improved model efficiency without heavy dependence on extensive data labeling processes [8][9][10]. One widely used SSL technique is pseudo-labeling, where an initial model trained in a supervised manner is used

to generate pseudo-labels on unlabeled data [11]. This technique is particularly relevant in retail contexts, which typically involve images with densely packed objects and high visual similarity between products. In such scenarios, semi-supervised object detection (SSOD) methods become an appropriate approach, as they can maximize the utilization of unlabeled data to enhance object detection performance [12].

A previous study on OSA [13] developed and evaluated several YOLO (You Only Look Once) model variants, namely YOLOv5n, YOLOv6-nano, YOLOv7-tiny, and YOLOv8n, for detecting product availability on retail store shelves (on-shelf availability/OSA). This study employed a fully-supervised learning approach with a total of 7,697 images and 125 product classes. Experimental results demonstrated that the YOLOv8n model delivered the best performance with an mAP50 score of 0.933 and an inference time of 13.4 ms, making it suitable for real-time implementation on resource-constrained devices. However, the approach used was entirely dependent on labeled data, without utilizing unlabeled data or semi-supervised strategies, rendering it less efficient for real-world scenarios with limited labels.

Research by Chauhan [12] applied a noisy student training approach for dense object detection in retail shelf images using the GFL (Generalized Focal Loss) model with a ResNet-50 backbone. They trained the model on 9,000 labeled images (PD9K) and progressively added 10K, 20K, up to 200K unlabeled images from the Retail Video Frames (RVF) dataset. Evaluation was conducted on the SKU110K dataset with high product density. The results showed that model accuracy improved from an mAP of 0.389 (baseline) to 0.403 after adding 200K pseudo-labels. They also proposed a pseudo-label filtering technique based on confidence thresholding and nms-inter to reduce false positives. Nevertheless, this study did not explore the use of YOLO architectures.

Study [14] compared three deep learning approaches—RetinaNet, YOLOv3, and YOLOv4—for monitoring on-shelf availability (OSA) in retail shelf images. The dataset used was WebMarket, consisting of 3,153 high-resolution shelf images captured from 18 different store shelves using three digital cameras. Of the entire dataset, 300 images were manually labeled using LabelImg and classified into five classes: three product categories (Beverage, Breakfast, Food) and two shelf conditions (Empty Shelf, Almost Empty Shelf). Experiments were conducted on varying proportions of labeled data (20%–80%) using a semi-supervised learning (SSL) approach with pseudo-labeling methods. The YOLOv4 model with CSPDarkNet53 backbone achieved the best performance in the full-supervised scenario, with an mAP of 0.9187, F1-score of 0.91, and recall of 0.96. In the SSL scenario, the SOSA approach still demonstrated competitive results, particularly when using 60–80% labeled data. However, accuracy gradually declined when the

proportion of labeled data was reduced to 40% and 20%, with the "Almost Empty Shelf" class being the most affected.

Research conducted by Dipendra [15] focused on automating the on-shelf availability inspection process, which was previously performed manually. They used a custom dataset containing 1,000 product shelf images that underwent data cleaning processes to improve training data quality. Two main object detection architectures were employed: the EfficientDet family and YOLOv5. The best results for the EfficientDet family were obtained from the EfficientDet-D1 model, with 57.0% precision, 63.8% recall, and 60.2% F1-score. Meanwhile, the YOLOv5n6 model demonstrated superior performance with 76.3% precision, 63.8% recall, and 71.3% F1-score, making it a more effective choice for retail product detection.

Research [16] developed an object detection model capable of running in real-time on mobile devices to monitor on-shelf availability (OSA) of retail products, specifically powdered milk. The dataset consisted of 4,637 images comprising 106 product classes, collected using smartphone cameras. The study employed YOLOv4-tiny due to its advantages in small model size, inference speed, and computational efficiency. The model was trained through six stages: data collection, preprocessing, annotation, training, evaluation, and testing. Experimental results showed that the model achieved an mAP of 92.14% and an inference time of 600–700 ms on Android devices. This demonstrates that YOLOv4-tiny is capable of detecting retail products in real-time on mobile devices, although there remains room for development, such as improving image resolution and exploring lighter model architectures.

Unlike previous studies, this research integrates a pseudo-labeling-based semi-supervised learning approach with the latest YOLOv11 architecture. The selection of YOLOv11 in this study is based on its advantages as the newest version of the YOLO architecture, offering significant improvements in accuracy, processing speed, and parameter efficiency. YOLOv11 demonstrates strong adaptability for various Computer Vision tasks such as object detection and instance segmentation, making it highly suitable for retail product detection. Furthermore, its efficiency makes it ideal for deployment across different platforms, including edge computing and cloud-based systems [17]. Although semi-supervised and pseudo-labeling approaches have shown great potential in enhancing object detection efficiency within the retail sector, most previous studies remain limited to earlier YOLO architectures or other detection models.

Therefore, this research aims to apply a semi-supervised pseudo-labeling approach to YOLOv11 and evaluate its performance across various proportions of labeled data (20%, 40%, 60%, and 80%). This study also compares pseudo-labeling

results with ground truth to assess the quality of generated pseudo-labels. It is expected that this approach will contribute to the development of more efficient, adaptive, and practically deployable OSA detection systems for real-world applications.

2. RESEARCH METHOD

This research was conducted through five stages to implement a pseudo-labeling-based semi-supervised learning approach on the YOLOv11 architecture for retail product object detection tasks. These stages include data collection and preparation, teacher model training, pseudo-labeling, student model training, and evaluation of model performance and pseudo-label quality. The research workflow is presented in Figure 1.

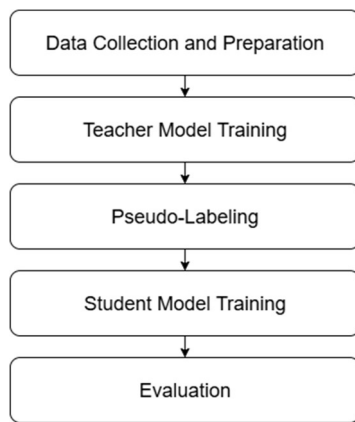


Figure 1. Research Methodology

2.1 Data Collection and Preparation

The initial stage of this research involves the collection and preparation of the dataset to be used for training and testing the object detection model. The dataset employed is a custom-built dataset collected manually from retail environments such as grocery store shelves and minimarkets, with variations in viewing angles, lighting conditions, and product arrangements to represent real-world conditions as shown in Figure 2.



Figure 2. Example of data

The dataset consists of 918 images, divided into 724 images for the training set, 92 images for the

validation set, and 92 images for the test set. These images contain various retail products with a total of 174 classes and are labeled in YOLO format (.txt) as follows: class, x center, y center, width, height, as shown in Figure 3, using a labeling tool [18].

```

    9 0.105957 0.195964 0.135742 0.313802
    11 0.249512 0.197266 0.157227 0.308594
    12 0.398926 0.197266 0.153320 0.316406
    12 0.548340 0.194661 0.153320 0.308594
    2 0.699219 0.198568 0.148438 0.313802
    2 0.844238 0.202474 0.141602 0.319010
    2 0.734375 0.582031 0.134766 0.265625
    12 0.713379 0.860677 0.131836 0.236979
    10 0.583984 0.855469 0.126953 0.229167
    10 0.454590 0.852865 0.129883 0.229167
    10 0.329590 0.852865 0.129883 0.221354
    10 0.205078 0.847656 0.130859 0.221354
    
```

Figure 3. Data label format

2.2 Teacher Model Training

The initial supervised training stage serves as the foundation for the pseudo-labeling process. In this stage, manually annotated data with several proportions (20%, 40%, 60%, and 80%) are used to train a teacher model, which is later utilized to generate pseudo-labels for unlabeled data. The training employs the YOLOv11n model pretrained on the COCO dataset as initialization, enabling the application of transfer learning to accelerate convergence and enhance detection performance. The model is trained to recognize 174 retail product classes through a supervised learning process using manual bounding box annotations. The selection of the nano version (YOLOv11n) is based on the objective of evaluating the model's performance starting from the smallest variant of YOLOv11. With 181 layers, 2,678,022 parameters, 2,678,006 gradients, and a computational requirement of 6.9 GFLOPs, YOLOv11n offers an optimal balance between model complexity and detection efficiency, making it an ideal choice for the initial experiments in this study [19]. As an additional variation of testing, a Teacher Model training (Pseudo-Label Data), also referred to in this study as the pseudo model, was conducted. This model is the result of supervised training that was retrained using only pseudo-labeled data.

2.3 Pseudo-labeling

After the supervised training stage is completed, the trained teacher model is used to perform inference on unlabeled data to generate pseudo-labels for the next training phase. This process begins with object detection on unlabeled images, where the model predicts object locations and classes based on knowledge learned from previously labeled data. Next, a confidence filtering step is applied with a confidence threshold of 0.5, ensuring that only high-confidence predictions are retained to maintain the quality of the generated pseudo-labels. The resulting predictions are then converted into YOLO annotation format (.txt) to

maintain consistency with the structure of the labeled dataset. These pseudo-labels are subsequently used directly alongside the labeled data in the next semi-supervised training stage. This approach reflects the core principle of semi-supervised learning, which leverages the teacher model's capability to effectively expand the training dataset, enabling the model to learn from a combination of labeled and unlabeled data to enhance overall detection performance [20].

2.4 Student Model Training

After pseudo-labels are generated on unlabeled data, the subsequent stage involves retraining the model by combining manually labeled data and pseudo-labeled data. This process is referred to as semi-supervised training, which aims to optimally utilize the entire training dataset, including both human-annotated and teacher model-generated labels. In this research, three student training experiments were conducted with the following variations:

- Pretrained Model YOLOv11n: Training using the base pretrained YOLOv11n model
- Teacher Model (Labeled Data): Model resulting from supervised training on manually labeled data
- Pseudo Model: Model resulting from supervised training using only pseudo-labeled data.

All three experiments were executed across each proportion of labeled data (20%, 40%, 60%, and 80%), enabling performance analysis of the model with respect to variations in the amount of labeled data within the semi-supervised learning framework

2.5 Evaluation

The final stage of this research involves evaluating the performance of the student model resulting from semi-supervised training. The evaluation aims to measure the model's capability to accurately detect objects on test data and assess the extent to which pseudo-labels contribute to improved detection performance. The evaluation process is conducted using 92 manually annotated test images, allowing for objective comparison of test results. Additionally, an assessment of the quality of pseudo-labels generated by the teacher model is performed. The evaluation encompasses:

- Visual evaluation of pseudo-labels: Comparing pseudo-labels with ground truth directly on images to examine the alignment of bounding boxes and class assignments.
- Metric-based evaluation: Calculating metric scores using ground truth as the reference standard, enabling determination of how closely pseudo-labels approximate the quality of manual labels.

The evaluation metrics employed include Precision, Recall, mAP@0.5, mAP@0.5:0.95, F1-Score, IoU, and Inference time. By utilizing a combination of model performance evaluation and

pseudo-label quality assessment, this research is expected to provide a comprehensive understanding of the effectiveness of the semi-supervised approach compared to supervised training, while simultaneously assessing the actual contribution of pseudo-labels to improved object detection performance.

3. RESULT AND DISCUSSION

All experiments were executed on the Kaggle Notebook platform utilizing NVIDIA T4 GPU (2x accelerators) as the primary processing device. The programming environment comprised several main libraries, including Ultralytics version 8.3.189 for YOLOv11 model management, Torch version 2.8.0, and Torchvision version 0.19.0. Model training was conducted with the following parameter configurations: 100 epochs, image size of 640x640 pixels, batch size of 16, and AdamW as the optimizer. The obtained results encompass model performance across various training scenarios, along with comparative performance analysis between models and data proportions used. The discussion focuses on interpreting evaluation results to assess the effectiveness of the semi-supervised learning approach in enhancing object detection performance for retail products.

Table 1. Teacher Model Using Manual Labels

Mode	Precision	Recall	mAP50	mAP50-95	Inference Time
20%	0.464	0.689	0.621	0.565	3.1ms
40%	0.734	0.825	0.852	0.787	15.0ms
60%	0.801	0.886	0.898	0.830	2.6ms
80%	0.870	0.915	0.953	0.885	4.6ms

Table 1 shows the results of training the teacher model using manually labeled data. It can be seen that increasing the proportion of labeled data has a positive impact on improving model performance. With 20% labeled data, the mAP50-95 value only reaches 0.565, whereas with 80% labeled data it increases significantly to 0.885. This confirms that the amount of labeled annotation data greatly affects the quality of model detection.

Table 2. Pseudo-Label Quality Metrics

Model	Jumlah True Positives	Jumlah True Negatives	Jumlah False Positives	Jumlah False Negatives	F1-Score	IoU
20%	587	2374	1013	2751	0.3675	0.575
40%	440	2904	892	1080	0.6942	0.842
60%	294	2050	539	399	0.7279	0.819
80%	247	1055	314	115	0.6160	0.680

Table 2 presents the evaluation of pseudo-label quality generated by the teacher model by comparing it against ground truth across various proportions of

labeled data. Overall, there is an observable improvement in pseudo-label quality up to the 60% proportion, but it decreases at 80%.

With 20% labeled data, the number of TPs at 2,374 and FN reaching 2,751 indicates that many objects failed to be detected, resulting in a low F1-Score of 0.367 and IoU of only 0.575. When the proportion of labeled data increases to 40%, there is a significant improvement in pseudo-label quality. The F1-Score rises to 0.694 with IoU reaching 0.842. The increased number of TPs and decreased numbers of FP and FN indicate that the model is increasingly capable of recognizing and labeling objects properly.

At 60% labeled data, the pseudo-labels achieve the highest performance with an F1-Score of 0.727 and IoU of 0.819. However, at 80% labeled data, the performance actually decreases slightly with an F1-Score of 0.616 and IoU of 0.680. This decline may be caused by the model having learned from sufficient manual data, such that there are objects that were not labeled in the ground truth labels but the model assigned labels to them, resulting in false positives being counted, even though they are visually correct.

Table 3. Teacher Model Using Pseudo Labels

Model	Precision	Recall	mAP50	mAP50-95	Inference Time
20%	0.439	0.452	0.480	0.440	3.2ms
40%	0.734	0.731	0.799	0.738	2.9ms
60%	0.564	0.652	0.682	0.627	4.5ms
80%	0.366	0.467	0.391	0.355	3.0ms

Table 3 shows the results of training the teacher model using pseudo-label data without involving manually labeled data. The pseudo-label teacher model demonstrates the best performance at a 40% data proportion with an mAP50 value of 0.799. However, increasing the pseudo-label proportion up to 80% actually decreases performance significantly. This is caused by the increasing inaccuracy of pseudo-labels generated by the teacher model, thus causing noise in the training process.

Table 4. Training Student Model 20%

Model	Precision	Recall	mAP50	mAP50-95	Inference Time
Pre Trained Teacher Model	0.582	0.767	0.759	0.702	3.4ms
Pseudo Model	0.581	0.755	0.752	0.695	3.1ms
Pseudo Model	0.595	0.736	0.737	0.68	3.1ms

Table 4 displays the training results of the student model at a 20% labeled data proportion. Overall, the performance of all three models is still relatively low due to the limited amount of labeled data used in the training process. The student model trained using the pre-trained model obtained an mAP50 value of 0.759 and mAP50-95 of 0.702, slightly higher compared to the student model trained using the teacher model. Additionally, the student model trained using the

pseudo model achieved the lowest values, which is attributed to the poor quality of pseudo-labels generated at the 20% data proportion. The relatively low overall performance across all models indicates that at this proportion, the models still struggle to detect objects stably due to insufficient data variation.

Table 5. Training Student Model 40%

Model	Precision	Recall	mAP50	mAP50-95	Inference Time
Pre Trained Teacher Model	0.798	0.817	0.863	0.801	3.5ms
Pseudo Model	0.760	0.843	0.866	0.806	7.7ms
Pseudo Model	0.778	0.862	0.886	0.826	3.2ms

Table 5 shows the training results of the student model with a 40% labeled data proportion. It can be seen that all models experienced significant performance improvements compared to the 20% experiment. This improvement demonstrates that increasing the proportion of labeled data has a positive impact on the model's ability to perform object detection.

In this experiment, the student model trained using the pseudo model shows the best performance, with an mAP50 value of 0.886 and mAP50-95 of 0.826. The recall value of 0.862 also indicates that the model is capable of detecting most objects present in the test images. This performance is slightly higher compared to the other two models. This indicates that pseudo-labels can contribute to improving the model's generalization capability.

Table 6. Training Student Model 60%

Model	Precision	Recall	mAP50	mAP50-95	Inference Time
Pre Trained Teacher Model	0.808	0.881	0.907	0.843	3.0ms
Pseudo Model	0.863	0.878	0.918	0.854	2.8ms
Pseudo Model	0.868	0.892	0.931	0.864	3.5ms

Table 6 presents the training results of the student model with a 60% labeled data proportion. In this experiment, all models show performance improvements compared to the 40% proportion. The mAP50 and mAP50-95 values across all models are above 0.84, which indicates a fairly mature object detection capability. Among the three models tested, the pseudo model again demonstrates the best results with an mAP50 value of 0.931 and mAP50-95 of 0.864. This model also has the highest precision at 0.868 and recall of 0.892. This performance improvement shows that at the 60% labeled data proportion, pseudo-labels still provide a positive contribution to the model's learning process.

Table 7. Training Student Model 80%

Model	Precision	Recall	mAP50	mAP50-95	Inference Time
Pre Trained Teacher Model	0.903	0.898	0.945	0.877	3.3ms
Pseudo Model	0.915	0.929	0.954	0.890	3.0ms
Pseudo Model	0.856	0.918	0.941	0.871	2.7ms

Table 7 shows the training results of the student model with an 80% labeled data proportion. At this stage, all models achieve high performance with mAP50-95 values above 0.87, indicating that the models are capable of performing object detection with a very good level of accuracy. The teacher model obtains the best results with an mAP50 value of 0.954 and mAP50-95 of 0.890, as well as the highest precision at 0.915. Meanwhile, the pseudo model still shows competitive performance with an mAP50-95 of 0.871 and recall of 0.918. This demonstrates that the larger the proportion of manually labeled data, the better the model's ability to recognize objects with precision and consistency.

In the visual evaluation, 3 samples per experiment were taken to examine the pseudo-label results compared to the ground truth labels. In the 20% experiment, it can be seen in image (a) that the pseudo-labels are able to label some objects, but not all of them match the ground truth labels. For example, for Lactogrow objects, several boxes can be labeled but some are not detected. However, for Dancow boxes, the model can generate pseudo-labels for every object. Even for Batita objects, the model can label them, which were not labeled in the ground truth, although the results are visually correct, thus increasing the false positive count.

In image (b), the model cannot label most objects, which may be caused by the model only learning from relatively limited data. In image (c), the model can generate pseudo-labels quite well, which is also because the objects in that image are not too varied, making it easier for the model to provide pseudo-labels compared to image (b).



Figure 4. Visual pseudo-label 20%, ground truth : green (left), pseudo-label : red (right)



Figure 5. Visual pseudo-label 40% ground truth : green (left), pseudo-label : red (right)

In the 40% experiment, there is a visible improvement from the previous one. In image (a), the model can now label almost all Lactogrow objects. For Batita objects, it can only provide 1 label out of 4 objects, while for Dancow objects it can provide 3 out of 4 objects. In image (b), the model performs very well and can label all objects. Then in image (c), the ground truth only provides one label, whereas the pseudo-labels successfully add more labels. Although

this may count as false positives when compared to the ground truth labels, it can add object variation.

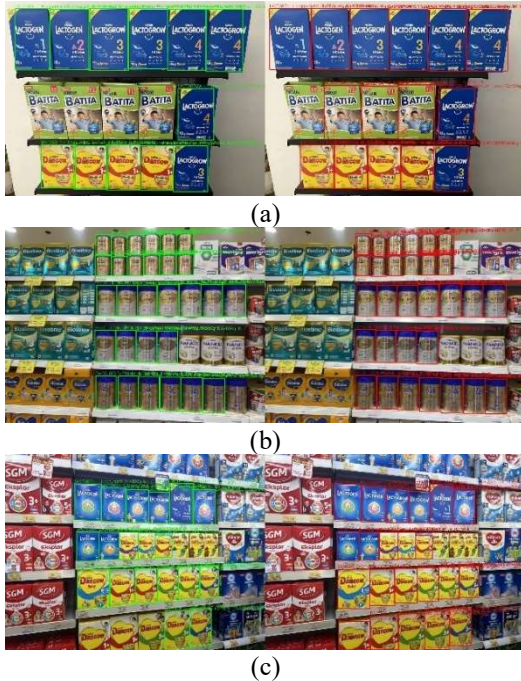


Figure 6 Visual pseudo-label 60%, ground truth : green (left), pseudo-label : red (right)

In the 60% experiment, it can be seen in image (a) that the model can provide labels for all objects that were labeled in the ground truth, although Batita objects still count as false positives. However, this can add variation to the dataset objects. In image (b), the model can label objects well, consistent with what was manually labeled, indicating an improvement in pseudo-label quality as the proportion of labeled dataset increases. In image (c), the model can label objects but not all of them. This may be caused by the image having a fairly high level of density and variation, so the model still experiences difficulty in maintaining labeling consistency.

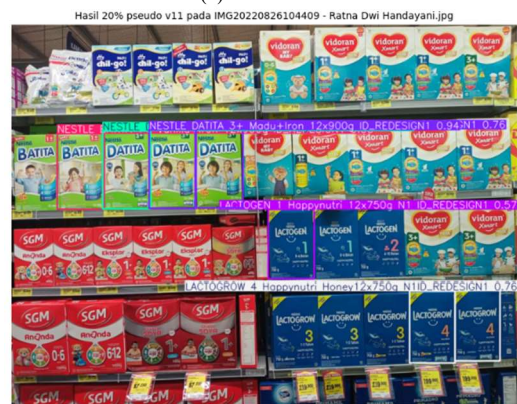


Figure 7 Visual pseudo-label 80%, ground truth : green (left), pseudo-label : red (right)

In the 80% experiment, it can be seen in image (a) that all objects labeled in the ground truth are successfully replicated by the pseudo-labels. Additionally, the Batita milk class is also successfully labeled in the pseudo-labels, although not all are detected. Next, in image (b), the model can provide accurate annotations for all objects, which is possible because the objects in the image are very clear and the model has learned from sufficient data. In image (c), the model is able to label all objects that were labeled in the ground truth, but there are additional objects that were not labeled in the ground truth. This condition again creates false positive cases, caused by the limitations of manual annotation. Overall, the model shows improvement in detection quality along with the increase in dataset size.



(a) Teacher Model



(b) Pseudo Model



(c) Pretrained Model
Figure 8. Testing Student Model 20%



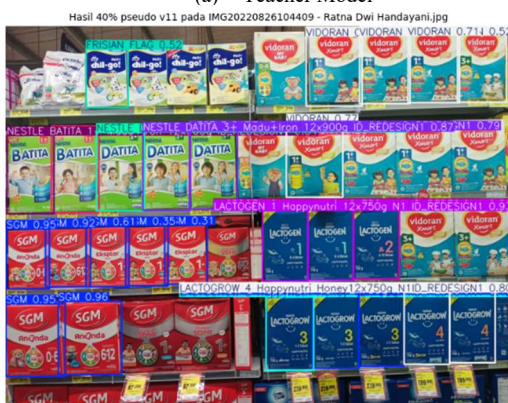
(c) Pretrained Model
Figure 9. Testing Student Model 40%

The testing results of the student model in the 20% experiment still fail to perform detection well, especially for the Vidoran, SGM, and Morinaga classes, whereas for the Lactogrow, Batita, and Datita classes, detection is possible although not all objects can be detected yet. This may be caused by limited object samples, and this is also consistent with the metric results where recall tends to be low, indicating that many objects are not successfully detected

The testing results in the 40% experiment show performance improvement along with the addition of labeled data, such as in the Vidoran and SGM classes which are starting to be detected, although there are still some that cannot be detected yet. Detection in the Batita, Datita, and Lactogrow classes is already quite good. However, the model still cannot detect the Morinaga class. Overall, it produces a trend of performance improvement across all models.



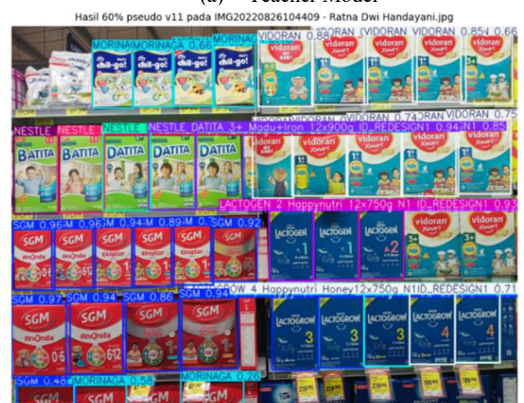
(a) Teacher Model



(b) Pseudo Model



(a) Teacher Model



(b) Pseudo Model



(c) Pretrained Model
Figure 10. Testing Student Model 60%



(c) Pretrained Model
Figure 11. Testing Student Model 80%

There is a performance improvement in the 60% experiment. The model is now able to detect almost all classes, especially the student model trained using the model that has learned from pseudo data. Out of all objects, only two objects are not detected. Although there are two false negatives, it shows the best performance among the other models, which is also consistent with the metric results shown.

Then in the 80% experiment, there is not much visible improvement compared to the 60% experiment visually, as it still fails to detect all objects, especially in the Morinaga class, which may be caused by limited class samples. Although it performs quite well with other classes, the best performance is delivered by the student model trained using the pre-trained YOLOv1 l model that has not yet learned from previous data.

Overall, the experimental results show that the semi-supervised learning approach utilizing pseudo-labels is capable of improving object detection performance in the retail product domain, especially at medium labeled data proportions such as 40% and 60%. The pseudo-labels generated by the teacher model are proven to provide a positive contribution to the student model's learning with significant increases in mAP and recall values. However, at proportions of labeled data that are too small or too large, model performance tends to decline due to limited information or the emergence of noise in the pseudo-labels. Thus, the balance between the amount of labeled data and pseudo-labels becomes an important factor in achieving optimal performance in the application of semi-supervised object detection.



(a) Teacher Model



(b) Pseudo Model

4. CONCLUSION

This study successfully implemented a semi-supervised learning approach based on pseudo-labeling using the YOLOv11n architecture for On-Shelf Availability (OSA) detection of retail products. The experimental results demonstrate that this approach effectively improves detection performance even with limited labeled data. Although the highest overall performance was achieved with 80% labeled data, the 60% proportion was already able to reach an mAP50 of 0.931 and an mAP50-95 of 0.864, indicating a significant improvement with higher labeling efficiency. The 60% proportion also produced high-quality pseudo-labels (F1-Score 0.727; IoU 0.819), enriching data variation without introducing excessive noise. Therefore, this approach has proven effective in reducing dependence on large labeled datasets while offering a practical and efficient

solution for OSA detection systems in retail environments, as well as opening opportunities for further research on optimizing pseudo-labeling strategies and exploring the latest YOLO architectures.

REFERENCE

- [1] Y. Cai, L. Wen, L. Zhang, D. Du, and W. Wang, "Rethinking Object Detection in Retail Stores," *35th AAAI Conf. Artif. Intell. AAAI 2021*, vol. 2A, pp. 947–954, 2021, doi: 10.1609/aaai.v35i2.16178.
- [2] Y. Wei, S. Tran, S. Xu, B. Kang, and M. Springer, "Deep Learning for Retail Product Recognition: Challenges and Techniques," *Comput. Intell. Neurosci.*, vol. 2020, 2020, doi: 10.1155/2020/8875910.
- [3] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, no. d, pp. 39–64, 2020, doi: 10.1016/j.neucom.2020.01.085.
- [4] G. Li, X. Li, Y. Wang, Y. Wu, D. Liang, and S. Zhang, "PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13669 LNCS, pp. 457–472, 2022, doi: 10.1007/978-3-031-20077-9_27.
- [5] Y. Ouali, C. Hudelot, and M. Tami, "An Overview of Deep Semi-Supervised Learning," pp. 1–43, 2020, [Online]. Available: <http://arxiv.org/abs/2006.05278>
- [6] V. Guimarães, J. Nascimento, P. Viana, and P. Carvalho, "A Review of Recent Advances and Challenges in Grocery Label Detection and Recognition," *Appl. Sci.*, vol. 13, no. 5, 2023, doi: 10.3390/app13052871.
- [7] K. Sohn *et al.*, "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence," *IEEE Trans. Ind. Informatics*, vol. 37, no. 10, pp. 1575–1585, 2022, [Online]. Available: <https://doi.org/10.1016/j.isprsjprs.2020.01.013><https://doi.org/10.1016/j.isatra.2020.08.010><https://doi.org/10.1016/j.knosys.2023.110634><https://doi.org/10.1016/j.energy.2023.126726><https://doi.org/10.1016/j.est.2022.105074><https://doi.org/10.1>
- [8] A. Mey and M. Loog, "Improved Generalization in Semi-Supervised Learning: A Survey of Theoretical Results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4747–4767, 2023, doi: 10.1109/TPAMI.2022.3198175.
- [9] J. Smith, "Advances in Semi-Supervised Learning Techniques for Real-World Applications," vol. 6, no. 1, pp. 9–20, 2025.
- [10] T. Shehzadi, Ifza, D. Stricker, and M. Z. Afzal, "Semi-Supervised Object Detection: A Survey on Progress from CNN to Transformer," pp. 1–21, 2024, [Online]. Available: <http://arxiv.org/abs/2407.08460>
- [11] J. Qi, M. Nguyen, and W. Q. Yan, "CISO: Co-iteration semi-supervised learning for visual object detection," *Multimed. Tools Appl.*, vol. 83, no. 11, pp. 33941–33957, 2024, doi: 10.1007/s11042-023-16915-4.
- [12] J. Chauhan, S. Varadarajan, and M. M. Srivastava, "Semi-supervised Learning for Dense Object Detection in Retail Scenes," pp. 1–4, 2021, [Online]. Available: <http://arxiv.org/abs/2107.02114>
- [13] D. H. Fudholi, A. Kurniawardhani, G. I. Andaru, A. A. Alhanafi, and N. Najmudin, "YOLO-based Small-scaled Model for On-Shelf Availability in Retail," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 8, no. 2, pp. 265–271, 2024, doi: 10.29207/resti.v8i2.5600.
- [14] R. Yilmazer and D. Birant, "Shelf Auditing Based on Image Classification Using Semi-Supervised Deep Learning to Increase On-Shelf," *sensors Artic.*, vol. 21, no. 2, 2021.
- [15] D. Jha, A. Mahjoubfar, and A. Joshi, *Designing an Efficient End-to-end Machine Learning Pipeline for Real-time Empty-shelf Detection*, vol. 1, no. 1. Association for Computing Machinery, 2022. [Online]. Available: <http://arxiv.org/abs/2205.13060>
- [16] R. Digo Saputra and D. Hatta Fudholi, "Model Mobile untuk Deteksi Objek pada On-Shelf Availability Produk Retail," 2023.
- [17] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," vol. 2024, pp. 1–9, 2024, [Online]. Available: <http://arxiv.org/abs/2410.17725>
- [18] Tzutalin, "LabelImg." Accessed: Aug. 15, 2025. [Online]. Available: <https://github.com/tzutalin/labelImg>
- [19] G. Jocher and J. Qiu, "Ultralytics YOLO11." Accessed: Oct. 03, 2025. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [20] K. Sohn, Z. Zhang, C.-L. Li, H. Zhang, C.-Y. Lee, and T. Pfister, "A Simple Semi-Supervised Learning Framework for Object Detection," 2020, [Online]. Available: <http://arxiv.org/abs/2005.04757>