International Journal of Innovative Computing, Information and Control
Volume **20**, Number **5**, October **2024**

ICIC International ©2024 ISSN 1349-4198
pp. 1319–1330

AN IMPROVED ENSEMBLE LEARNING-BASED APPROACH FOR RETAIL PRODUCT RECOGNITION

Huei-Yung Lin[1], Po-Yu Hsieh[2], Sen-Yih Chou[2] and Chin-Chen Chang[3,*]

[1]Department of Computer Science and Information Engineering
National Taipei University of Technology
No. 1, Section 3, Zhongxiao East Road, Taipei 10608, Taiwan
lin@ntut.edu.tw

[2]Center for Measurement Standards
Industrial Technology Research Institute
No. 195, Section 4, Chung Hsing Road, Chutung, Hsinchu 310401, Taiwan
{ poyu429; senyih }@itri.org.tw

[3]Department of Computer Science and Information Engineering
National United University
No. 2, Lienda Road, Miaoli 360302, Taiwan
*Corresponding author: ccchang@nuu.edu.tw

Received December 2023; revised April 2024

Abstract. *Due to the recent trend of unmanned economy, retail stores have gradually reduced their service and cashier manpower. The retail product recognition becomes one of key issues for unmanned shopping. Although the success of deep neural networks has made object recognition feasible in a variety of applications, it still struggles to perform well on a large number of classes of retail products. In this paper, we propose an improved ensemble learning-based approach for retail product recognition. In the proposed approach, the object classification network is first improved through feature extraction and block attention. An ensemble model is then built by integrating multiple network models and using loss selection as model weights. In the experiments, the feasibility of our ensemble learning-based approach has been evaluated through many product items. The results demonstrate the effectiveness of the proposed approach compared with previous retail product recognition methods.*
Keywords: Unmanned store, Ensemble learning, Retail product recognition, Convolutional neural network

1. **Introduction.** As declining birth rates and high labor costs become a growing problem around the world, traditional retailers are forced to reduce the manpower and cost requirements of store management. Some common methods include using Internet of Things (IoT) monitoring systems to ensure product health and safety [1] and deploying surveillance cameras for security purposes [26]. However, most current methods do not completely eliminate the need for on-site manpower. Therefore, the concept of unmanned stores in the retail industry has often been proposed and studied [24].

Due to the recent success of using deep neural networks for object recognition, it is possible to detect and recognize product items using image-based methods. Convolutional neural networks (CNN) have been widely used for object classification. Existing literature reports some promising results for different applications [14,19]. A major problem with these techniques for recognizing retail products is the excessive number of classes. There can be hundreds of different retail products in the same class on the racks which makes

DOI: 10.24507/ijicic.20.05.1319

recognition a challenging task. Furthermore, similar designs and patterns used for specific types of retail products further complicate feature extraction and network training [20].

In order to solve the problem of significant performance degradation when the number of classes increases, it is necessary to establish a more complex classification network architecture. In addition, a sufficient number of samples are needed to extract representative features for network training. However, neither is a trivial task and usually involves improving upon existing methods. An important strategy is to use the concept of ensemble learning to improve the recognition performance by combining different classification algorithms [7,11,17,22]. In most supervised learning techniques, data classification is achieved by training a classifier using ground-truth annotations. However, each learning method may have different classification effects on different datasets. Therefore, through the concept of ensemble learning, multiple learning algorithms can be used to obtain better prediction results than a single learning algorithm.

In this paper, we propose an ensemble learning-based approach for retail product recognition. We first improve the object classification network based on feature extraction and block attention. We then build an integrated model by integrating multiple network models and using loss selection as model weights. Our approach demonstrates the feasibility of recognizing up to fifty retail products. The contributions of the proposed approach are summarized below.

- We propose an improved ensemble learning-based approach for retail production recognition.
- We modify the classification network using new feature extraction and use attention modules to improve the basic recognition performance.
- We build an ensemble model by combining the network models with a weighted loss.

The rest of this paper is organized as follows. Section 2 reviews the related works. Section 3 describes the proposed approach. Section 4 presents the results and discussion. Finally, Section 5 presents the conclusions and directions for future research.

2. Related Works.

According to the combination strategy of ensemble learning, current methods can be divided into three categories. The first method was boosting [3], which combined multiple weak classifiers into a strong classifier. It required the learning error rate of the weak classifier to update the weights of the training samples. The purpose was to increase the weight of misclassified data by the old classifiers and then train the new classifiers. This will provide the new classifiers with the features to learn misclassified samples, improve the classification effect, and obtain an ensemble model with high recognition accuracy. The second method was bagging [3]. It randomly selected n samples from the training data each time and put them back. Due to random sampling, subsets of samples collected multiple times were often different from each other and from the original training data set. Repeating this process m times can produce m weak classifiers. Then, averaging, voting or other learning methods were used for ensemble learning. The third and most widely used neural network ensemble learning technique was stacking [15]. In this method, the original data set was used to train a first weak classifier for prediction, and then the labeled data set of the second weak classifier was used to produce an ensemble model.

In addition to the above approaches, ensemble learning also had various implementation strategies. Lee et al. [8] proposed an approach of CNN models for object detection. This approach not only considered the quality of the ensemble model based on the overall mean average precision (mAP), but also selected different CNN models based on the class average precision (AP). The advantages of the models were complementary to each other and AP was used as a weight to determine confidence. Then, it was followed by the voting scheme to improve the performance of the ensemble model. Alam et al. [2] presented a dynamic ensemble learning algorithm to design and train the ensemble of neural networks. Traditional ensemble designing was still a manual process. This approach can automatically create ensemble architectures that maintained the accuracy and diversity of the neural networks, as well as the minimum number of parameters specified by the designer. Their method can achieve good generalization ability. Zhu et al. [29] proposed an approach to improve the classification on imbalanced data sets through a geometric structure ensemble learning framework. First, they generated a hypersphere through Euclidean metric to divide and eliminate redundant majority samples. Second, this method learned a basic classifier to enclose a small number of samples. Their approach can achieve higher efficiency during the training process. Finally, the remaining samples were used to train the next sample until the entire training process was completed.

Santra et al. [16] proposed an annotation-free machine vision system for detecting products on racks. Their system consisted of three modules: exemplar-driven region proposal, classification, and non-maximal suppression of region proposals. First, they estimated the scale of the rack images relative to product template images. Their system then used the estimated scale to generate potential object regions. Finally, a CNN was used to classify potential object regions. Their approach outperformed the previous methods improving detection accuracy. George and Floerkemeier [5] presented an approach for per-exemplar multi-label image classification of products in retail store images. Their method used discriminative random forests, deformable dense pixel matching, and genetic algorithm optimization to achieve high efficiency. Furthermore, they performed product image search using tagging tools for multi-label retail product image classification. Their approach can achieve good results for accuracy and efficiency. Wang et al. [21] proposed a fine-grained classification algorithm for retail product recognition based on self-attention destruction and construction learning. They used self-attention techniques to destroy and construct image information in an end-to-end manner. Therefore, they can compute accurate fine-grained classification predictions and large informative regions during inference. Their approach can achieve a higher accuracy than the previous methods in retail product recognition.

3. **Proposed Approach.** We first improve the RefineDet network [27] by modifying feature extraction modules and attention modules. The structure in the original network architecture is still remained. We then add the attention mechanisms to the RefineDet network structure for optimization. Finally, we build an ensemble learning model to achieve better performance.

3.1. **Improved RefineDet.** In the original network structure of RefineDet, VGG16 is used for feature extraction [18] and the single-shot detector (SSD) is used as the backbone architecture [10]. It combines region proposal network (RPN) and feature pyramid network (FPN) to ensure recognition performance under single small target detection [9]. The RefineDet network architecture consists of three main components, the anchor refinement module (ARM), the transport connection block (TCB), and the object detection module (ODM). We make modifications on both feature extraction and attention modules of RefineDet to optimize overall performance.

First, the feature extraction module is an important part for RefineDet to recognize images. Three different networks, ResNeSt [28], RegNet [13], and RexNet [6], are integrated into the feature extraction module of RefineDet. The improved network structure is shown in Figure 1, which still retains the original idea of using ARM, TCB, and ODM modules. The orange shaded areas are the changed areas and VGG16 in the original

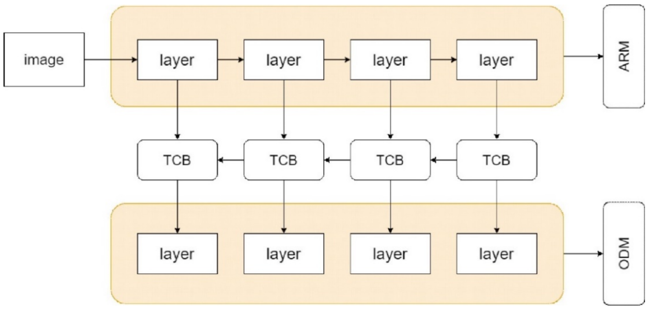

FIGURE 1. The improved network architecture

RefineDet is replaced with ResNeSt, RegNet, and RexNet. Next, the attention modules are incorporated into RefineDet to improve feature extraction of convolutional layers.

3.2. **Attention mechanism.** As shown in Figure 2, three attention modules are used: convolutional block attention module (CBAM) [23], TCBv2 [25], and enhanced map block (EMB) [4]. The CBAM is mainly divided into channel and spatial attention modules. The model uses max pooling to obtain feature differences in images and average pooling to extract common features for global feature learning. TCBv2 joins the ideas of PANet [12] and includes FPN with the bottom-up path. This is then integrated into the TCB module in RefineDet. EMB [4] contains attention streams and feature maps. It provides a feature-enhancing approach that improves detection capabilities without reducing processing speed.

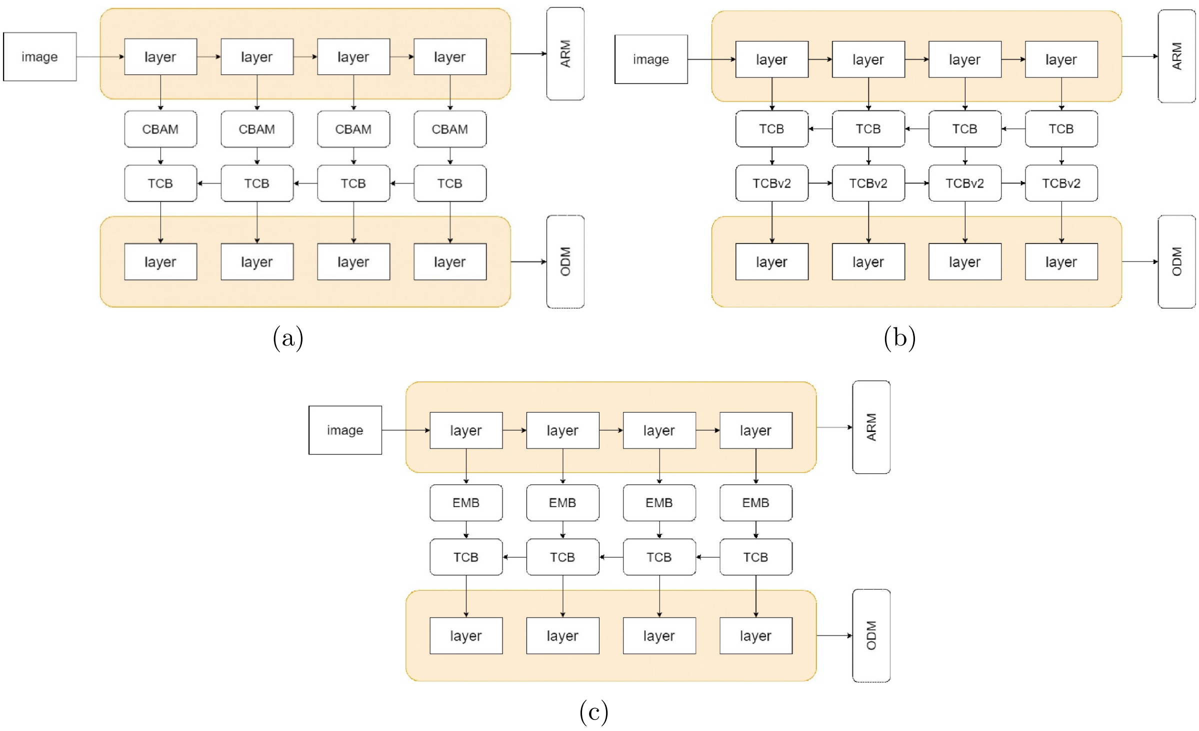

FIGURE 2. Three attention modules, CBAM, TCBv2, and EMB, are adopted to optimize RefineDet: (a) Optimized RefineDet architecture with CBAM; (b) optimized RefineDet architecture with TCBv2; (c) optimized RefineDet architecture with EMB

3.3. **Ensemble learning model.** In order to build our ensemble learning model and improve the recognition performance, we replaced VGG16 with ResNeSt, RegNet, and RexNet in RefineDet for feature extraction. The three models are trained separately, and their respective mAP is used as the basic recognition rate for integration and improvement. Our ensemble learning approach is based on the training losses of individual models. The feature extraction layers of different classification networks use the same training data. This process is performed in parallel, and the weighted sum of the training losses is used for performance evaluation of the ensemble model. We first set the weights based on the mAP of the base classification model. Although this configuration can provide good recognition rates, if the network training is updated with changed training data, the weights need to be adjusted accordingly.

To solve this problem, an automatic weight adjustment technique for our ensemble model is proposed. It consists of loss selection and joint loss calculation. Through our loss integration method of ensemble learning, we can achieve better performance than the original individual basic recognition models.

For loss selection, as shown in Figure 3, the loss of each model can be calculated by selecting one from n batches of training at each epoch. The obtained model loss is used to adjust the contribution ratio among the underlying network structures. In general, the higher loss corresponds to the poorer recognition result and a lower model weight is assigned accordingly. For the joint loss computation, the ensemble loss is calculated by the weighted sum of model training losses.

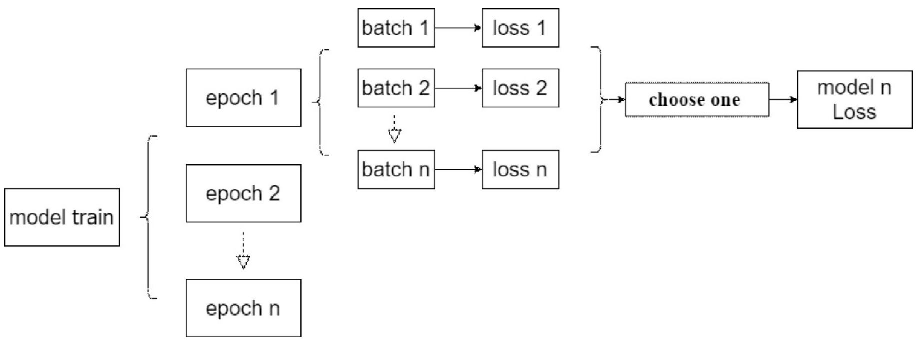

FIGURE 3. The loss for each model derived by selecting one from n batches of training at each epoch

The weight ratio of each epoch during model training is derived based on the reciprocal of the model loss in the previous epoch. The better-performing model has a relatively good loss contribution value. For a more general formulation with n base network models, the ensemble loss L^i is given by

$$L^i = \sum_{j=1}^{n} W_j^i \cdot L_j^i, \tag{1}$$

where

$$W_j^i = \frac{1}{L_j^{i-1} \sum_{j=1}^{n} \left(1/L_j^{i-1}\right)}. \tag{2}$$

Here L_j^i is the training loss of the network model j at the i-th epoch; W_j^i is the weight ratio of the network model j at the i-th epoch. For the initialization at the first training epoch, equal weights are assigned to compute the ensemble loss. Finally, the ensemble loss is used for backpropagation at each epoch. Through our loss integration technique, we can achieve better performance than the original individual basic recognition models.

4. **Results.** We tested the proposed ensemble learning-based approach and network models for retail product recognition with real scene images. Product images were captured by a camera (GoPro HERO 7) with a resolution of 1920×1440, mounted on the top of the fixed bracket and viewing downward. Experimental product items, including various snacks, were placed in different orientations and lighting for image acquisition. The calculations for retail product recognition run on a PC with an Intel Core i5 8600 CPU and an NVIDIA Geforce GTX1080 GPU.

We first used 10 product items to verify the reliability of the proposed approach. Each product item had 200 images for training and 50 images for testing. Figure 4 showed several sample images acquired in our experiments and used as the training data. As shown in Figure 5, the captured testing images were slightly different from the training samples.

FIGURE 4. Several images acquired in our experiments and used as the training data

FIGURE 5. The testing images captured slightly different from the training samples

We replaced the feature extraction layer of RefineDet with ResNeSt50, RegNet, and RexNet. Table 1 showed the mAP results of comparing the original RefineDet and modified RefineDet network models. The mAP increased by approximately 30% for all three conditions. Figure 6 showed the results using the modified RefineDet network models of three feature extraction layers ResNeSt50, RegNet, and RexNet. Although some snacks were blocked by hands, the three modified RefineDet network models can still recognize the product class and location. Compared with the original RefineDet network model using VGG16, the recognition effect of the proposed approach was more stable.

Table 1. The mAP comparison for different modified RefineDet network structures

Model	RefineDet	w/ResNeSt50	w/RegNet	w/RexNet
pocky	86%	79%	85%	74%
cocopuff	85%	81%	80%	92%
doritos	73%	75%	58%	79%
noodles	64%	92%	91%	94%
greengood	51%	87%	96%	88%
kolanut	43%	84%	77%	84%
oysterchip	35%	81%	82%	86%
oricracker	34%	77%	78%	86%
oreo	18%	77%	88%	91%
cookie	18%	71%	64%	94%
mAP	50.7%	80.4%	79.9%	86.8%

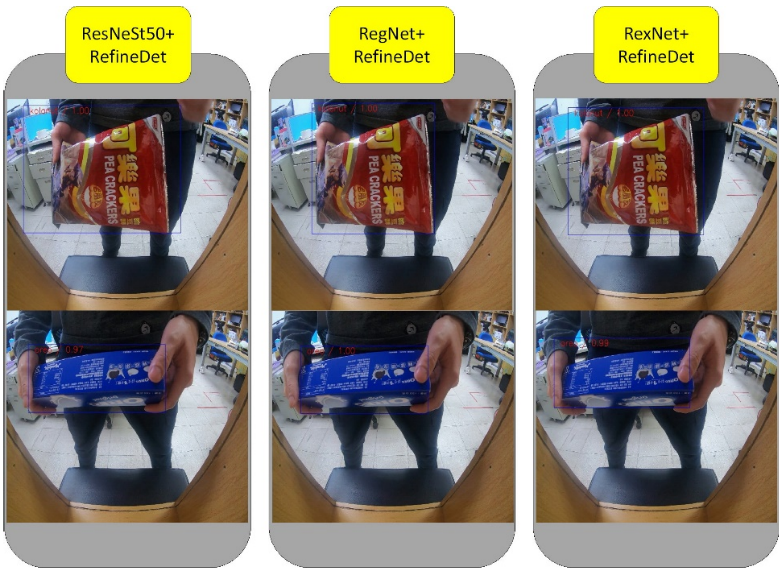

Figure 6. The results using the modified RefineDet network models of three feature extraction layers ResNeSt50, RegNet, and RexNet

We then collected a total of 4000 images of 20 retail products for evaluation. The number of testing images was 1000. In order to further improve recognition performance on a variety of product items, we used three ensemble models, namely, ResNeSt50+RexNet, ResNeSt50+RegNet, and RexNet+RegNet. We take the ensemble model ResNeSt50+RexNet (Res50+Rex for short) with the best mAP as our ensemble model. We compared our ensemble model with YOLOv5m and YOLOv5x [7]. The mAP of our ensemble model was 88.42%. The mAPs of YOLOv5m and YOLOv5x were 84.99% and 88.33%, respectively. Hence, our ensemble model performed better than both YOLOv5m and YOLOv5x. Figure 7 depicted the comparison of the proposed approach with YOLOv5m and YOLOv5x. Our ensemble model can recognize both "seaweedcookie" and "cocostick". For YOLOv5m, "cocostick" was recognized as "oricracker". For YOLOv5x, "seaweedcookie" cannot be correctly recognized at a certain perspective view.

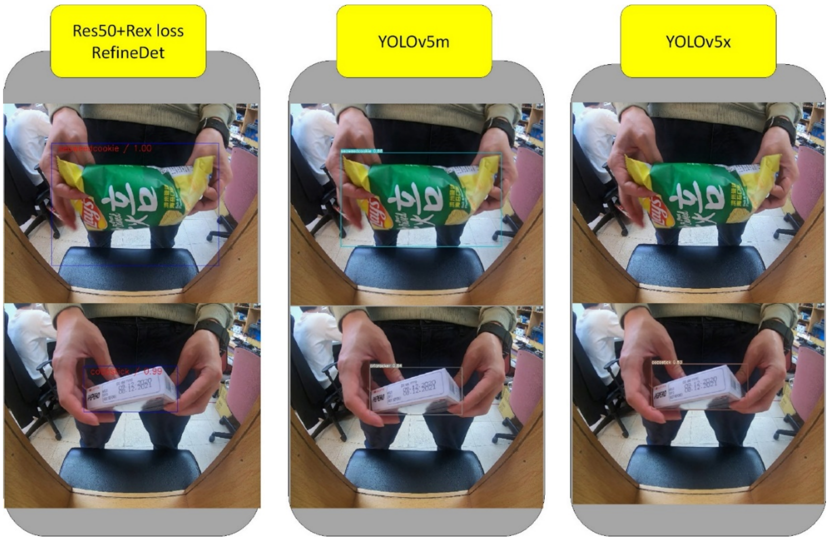

FIGURE 7. Comparison of Res50+Rex loss RefineDet, YOLOv5m, and YOLOv5x. The retail products on top and bottom were "seaweedcookie" and "cocostick", respectively.

Next, we used a total of 6000 images of 30 product items for evaluation. The testing images contained 500 images for each class. For our ensemble model Res50+Rex, the mAP was 77.72%. The mAPs of YOLOv5m and YOLOv5x were 70.91% and 74.67%, respectively. Our ensemble model still performed better than both YOLOv5m and YOLOv5x for more product items. As shown in Figure 8, our ensemble model can correctly recognize "coconutscookiee" and "cocostick". For YOLOv5m, "cocostick" was recognized as

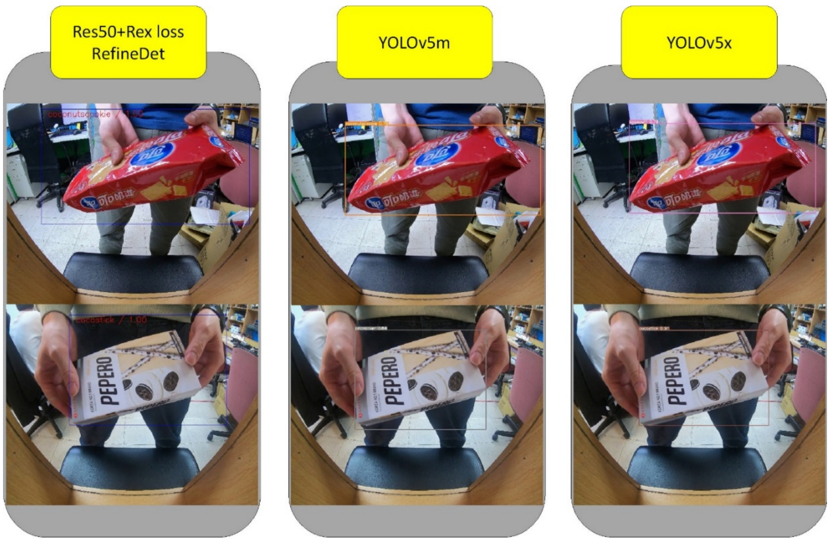

FIGURE 8. Comparison of Res50+Rex loss RefineDet, YOLOv5m, and YOLOv5x. The retail products on top and bottom were "coconutscookiee" and "cocostick", respectively.

"oricracker" because these two items were white and square in appearance. For YOLOv5x, "seaweedcookie" was recognized as "cocopuff" since these two items were red in appearance.

Finally, we used 50 product items for model training and testing with a total of 10,000 images. Each product item had 150 images for training and 50 images for testing. However, the differences of the recognition rates of our ensemble model Res50+Rex for different items were quite large indicating the limited capabilities of the model generalization. To further increase the stability of recognition on a variety of product items, we modified the base network structures. We replaced the initial feature extraction module of ResNeSt50 with ResNeSt101 (Res101 for short) and RexNet was changed to the width of 3 (Rex3 for short). Furthermore, the number of convolutional layers and kernel size were increased. For our ensemble model Res101+Rex3, the mAP was 72.94%. Moreover, the recognition results of Res101+Rex3 were stable for all product classes. For adding attention modules, we modified Res101+Rex3 to present different ensemble models. Res101+Rex3 CBAM was constructed by using pooling to enhance feature extraction. Res101+Rex3 TCBv2 was built by enhancing the data conversion of ARM and ODM and fusing the features from shadow and deep layers. Res101+Rex3 EMB was built by adding EMB modules to optimize ARM-derived features. The mAPs of Res101+Rex3 CBAM, Res101+Rex3 TCBv2, and Res101+Rex3 EMB were 71.49%, 78.24%, and 79.53%, respectively. The mAPs of YOLOv5m and YOLOv5x were 67.96% and 71.93%, respectively. The results illustrated that our ensemble models provided better mAPs than YOLOv5m and YOLOv5x. Figure 9 showed that our ensemble model Res101+Rex3 EMB with the best mAP can recognize challenging cases "cocostick" and "smallstick". For YOLOv5m and YOLOv5x, they were recognized as "oricracker" and "greengood", respectively.

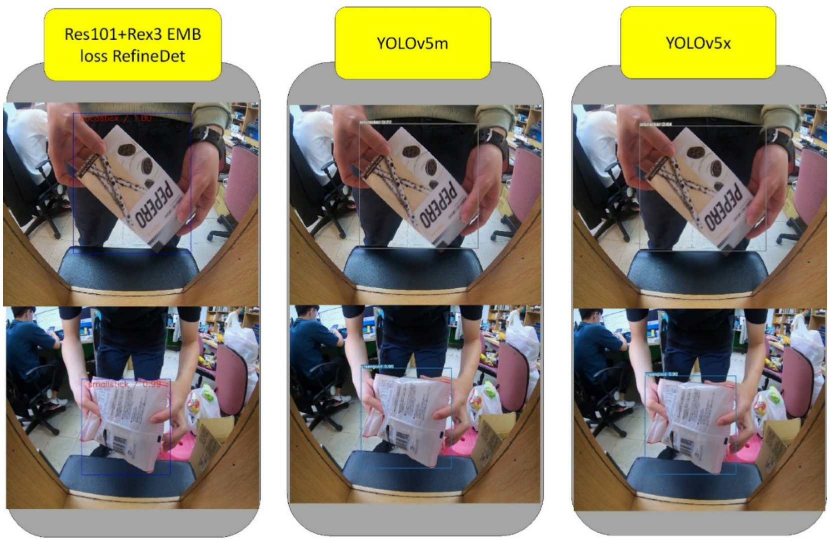

Figure 9. Comparison of Res101+Rex3 EMB loss RefineDet, YOLOv5m, and YOLOv5x. The retail products on top and bottom were "cocostick" and "smallstick", respectively.

5. **Conclusions.** We have proposed an ensemble learning-based approach for retail production recognition. The classification network, RefineDet, is first modified using new feature extraction and attention modules to improve the basic recognition accuracy. The ensemble models are then built by combining the network models with a weighted loss. The proposed ensemble learning-based technique is evaluated using 50 retail products.

Compared with previous networks, our approach demonstrates the feasibility of recognizing a large number of object classes. In future research, we will try to add more ensemble learning methods for analysis and increase the number of basic recognition models for integration through different techniques. In addition, we will add more classes to test whether the recognition performance is maintained for the proposed approach.

Acknowledgment. The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-194-004 and Center for Measurement Standards, Industrial Technology Research Institute is gratefully acknowledged. Moreover, this paper is a revised and expanded version of a paper entitled "Ensemble Learning for Retail Product Recognition with a Large Number of Classes", presented at 2022 IEEE International Conference on Systems, Man and Cybernetics (SMC 2022), Prague, Czech Republic, October 2022.

REFERENCES

[1] M. H. Ahmadzadegan, M. Mohammadzadeh, G. Eftekharnejad and H. Ghorbani, Intelligent monitoring systems for transportation of perishable products based Internet of Things (IoT) technology, *Proc. of 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*, Gwalior, India, 2020.
[2] K. M. R. Alam, N. Siddique and H. Adeli, A dynamic ensemble learning algorithm for neural networks, *Neural Computing and Applications*, vol.32, no.12, pp.8675-8690, 2020.
[3] B. Asadi and R. Hajj, Prediction of asphalt binder elastic recovery using tree-based ensemble bagging and boosting models, *Construction and Building Materials*, vol.410, Article no.134154, 2024.
[4] H. T. Choi, H. J. Lee, H. Kang, S. Yu and H. H. Park, SSD-EMB: An improved SSD using enhanced feature map block for object detection, *Sensors*, vol.21, no.8, Article no.2842, 2021.
[5] M. George and C. Floerkemeier, Recognizing products: A per-exemplar multi-label image classification approach, *Proc. of the 13th European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, 2014.
[6] D. Han, S. Yun, B. Heo and Y. Yoo, ReXNet: Diminishing representational bottleneck on convolutional neural network, *arXiv Preprint*, arXiv: 2007.00992, 2020.
[7] P. Y. Hsieh, H. Y. Lin and S. Y. Chou, Ensemble learning for retail product recognition with a large number of classes, *Proc. of 2022 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Prague, Czech Republic, 2022.
[8] J. Lee, S. K. Lee and S. I. Yang, An ensemble method of CNN models for object detection, *Proc. of 2018 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea, 2018.
[9] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan and S. Belongie, Feature pyramid networks for object detection, *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.
[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A. C. Berg, SSD: Single shot multibox detector, *Proc. of the 14th European Conference on Computer Vision (ECCV)*, Amsterdam, The Netherlands, 2016.
[11] H. Liu and S. M. Chen, Multi-level fusion of classifiers through fuzzy ensemble learning, *Proc. of 2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, 2018.
[12] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, Path aggregation network for instance segmentation, *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018.
[13] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He and P. Dollár, Designing network design spaces, *Proc. of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 2020.
[14] W. Rawat and Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, *Neural Computation*, vol.29, no.9, pp.2352-2449, 2017.
[15] O. Sagi and L. Rokach, Ensemble learning: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol.8, no.4, 2018.

[16] B. Santra, A. K. Shaw and D. P. Mukherjee, An end-to-end annotation-free machine vision system for detection of products on the rack, *Machine Vision and Applications*, vol.32, Article no.56, 2021.
[17] S. Saravanan and G. M. K. Nawaz, Ensemble-based time series data clustering for high dimensional data, *International Journal of Innovative Computing, Information and Control*, vol.10, no.4, pp.1457-1470, 2014.
[18] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv Preprint*, arXiv: 1409.1556, 2014.
[19] P. Tang, X. Wang, B. Shi, X. Bai, W. Liu and Z. Tu, Deep FisherNet for image classification, *IEEE Transactions on Neural Networks and Learning Systems*, vol.30, no.7, pp.2244-2250, 2018.
[20] V. L. Tran, H. Y. Lin and H. W. Liu, MBNet: A multi-task deep neural network for semantic segmentation and lumbar vertebra inspection on X-ray images, *Proc. of the Asian Conference on Computer Vision*, Kyoto, 2020.
[21] W. Wang, Y. Cui, G. Li, C. Jiang and S. Deng, A self-attention-based destruction and construction learning fine-grained image classification method for retail product recognition, *Neural Computing and Applications*, vol.32, no.18, pp.14613-14622, 2020.
[22] B. Wattanacheep and O. Chitsobhuk, A CNN-based multi-model ensemble method for indoor and outdoor multi-view stereo reconstruction, *International Journal of Innovative Computing, Information and Control*, vol.18, no.5, pp.1571-1586, 2022.
[23] S. Woo, J. Park, J. Y. Lee and I. S. Kweon, CBAM: Convolutional block attention module, *Proc. of the 15th European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018.
[24] H. C. Wu, C. H. Ai and C. C. Cheng, Experiential quality, experiential psychological states and experiential outcomes in an unmanned convenience store, *Journal of Retailing and Consumer Services*, vol.51, pp.409-420, 2019.
[25] H. Xie and Z. Wu, A robust fabric defect detection method based on improved RefineDet, *Sensors*, vol.20, no.15, Article no.4260, 2020.
[26] Q. Xu, W. Zheng, X. Liu and P. Jing, Deep learning technique based surveillance video analysis for the store, *Applied Artificial Intelligence*, vol.34, no.14, pp.1055-1073, 2020.
[27] S. Zhang, L. Wen, X. Bian and Z. Lei, Single-shot refinement neural network for object detection, *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp.4203-4212, 2018.
[28] Z. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li and A. Smola, ResNeSt: Split-attention networks, *Proc. of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA, 2022.
[29] Z. Zhu, Z. Wang, D. Li, Y. Zhu and W. Du, Geometric structural ensemble learning for imbalanced problems, *IEEE Transactions on Cybernetics*, vol.50, no.4, pp.1617-1629, 2018.

Author Biography

Huei-Yung Lin is a professor at the Department of Computer Science and Information Engineering, National Taipei University of Technology. He also has a joint appointment with the Department of Electrical Engineering, National Chung Cheng University. Professor Lin received his Ph.D. degree in Electrical and Computer Engineering from State University of New York at Stony Brook, US. In 2002 he joined National Chung Cheng University, Taiwan, as an assistant professor and was promoted to a full professor in 2013. He served as the director of Research Liaison Division from 2009 to 2013, and the director of Academic Development Division from 2012 to 2014, both positions are with Office of Research and Development, National Chung Cheng University. He is an author of over 200 international conference and journal papers, and 5 book chapters. Professor Lin holds 11 US and 9 Taiwan invention patents. He also serves as the organizing committee member and program committee member of over 50 international conferences. He is recipients of Excellent Research Award from National Chung Cheng University, Outstanding Academic-Industry Cooperation Award from Taiwan Association of System and Science and Engineering (TASSE), and Outstanding Robotics Engineer Award from Robotics Society of Taiwan (RST). His research interests include computer vision, machine learning, robotics, and mechatronics. He is a fellow of IET and RST, and a senior member of IEEE and Optica.

Po-Yu Hsieh received his M.S. degree from the Institute of Electrical Engineering, National Chung Cheng University in Chiayi, Taiwan, in 2021. As a software engineer at the Industrial Technology Research Institute, he has dedicated the past two years to specializing in medical image processing and analysis. His expertise lies in the application of artificial intelligence, object recognition, as well as edge computing units and cloud software deployment. He continuously refines his skills in the professional field and keeps abreast of the latest technological advancements.

Sen-Yih Chou received his Ph.D. degree from the Institute of Electro-Optical Engineering, National Chiao Tung University in Hsinchu, Taiwan, in 2012. As a principal engineer at the Industrial Technology Research Institute, he has dedicated 15 years to mastering high-speed optical inspection, honing his skills in practical field applications. His core competencies encompass AI applications, sensing system integration, instrument system design, and fiber optics. He is committed to staying abreast of the latest advancements in technology, market dynamics, and intelligent system developments.

Chin-Chen Chang received his M.S. degree in Applied Mathematics from National Chiao Tung University in Hsinchu, Taiwan, in 1994, and his Ph.D. degree in Computer Science from National Chiao Tung University in Hsinchu, Taiwan, in 1998. At present, he is a professor in the Department of Computer Science and Information Engineering, National United University in Miaoli, Taiwan. He is a member of IICM and IPPR in Taiwan. His research interests include computer graphics, image processing, and virtual reality.