Artificial Intelligence Evolution
http://ojs.wiserpub.com/index.php/AIE/

Research Article

A Framework for Open World Object Detection

Khadija Shaheen[1*], Muhammad Abdullah Hanif[2], Osman Hasan[1], Muhammad Shafique[2]

[1]School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan
[2]New York University Abu Dhabi, United Arab Emirates
E-mail: shaheenkhadija162@gmail.com

Received: 17 May 2023; **Revised:** 31 July 2023; **Accepted:** 15 August 2023

Abstract: Open World Object Detection (OWOD) is a computer vision task that focuses on real-world scenarios where object detection algorithms need to not only detect known and labeled objects but also handle novel and unknown objects that were not seen during training. This distinguishes OWOD from traditional object detection benchmarks, where the scope is limited to detecting only known object classes. The main challenge in OWOD lies in detecting and classifying unknown objects, which were not part of the training data. In standard object detection, objects not overlapping with labeled objects are automatically classified as background. However, these approaches are not suitable for OWOD, as unknown objects may be wrongly predicted as background due to the lack of specific supervision for distinguishing unknown objects from the background. The paper proposes a novel framework for Open World Object Detection called Open World Object Detection based on Non-Parametric classification (OWOD-NP). This method aims to address the challenges of identifying unknown objects and extending the knowledge base by incrementally introducing new object categories. OWOD-NP incorporates a non-parametric learning approach based on mean prototypes and rejection criteria into a standard detector model. The non-parametric learning model allows the system to detect whether the perceived region contains an unknown object and perform incremental learning in an end-to-end manner. The extensive experiments conducted on the benchmark dataset of Pascal Visual Object Classes (VOC) validate the effectiveness of OWOD-NP. Compared to the standard faster RCNN model, OWOD-NP achieves approximately 14% higher mean Average Precision (mAP) in class incremental scenarios. This improvement showcases the capability of OWOD-NP to handle open-world object detection tasks more efficiently. By combining non-parametric learning with object detection, OWOD-NP provides a promising solution for open-world scenarios, where the environment is dynamic and new objects may appear over time. The ability to detect and classify both known and unknown objects makes OWOD-NP a valuable approach for real-world applications in robotics, autonomous systems, and other computer vision tasks. It allows for continuous adaptation and learning, enabling the system to extend its knowledge and cope with ever-changing environments effectively.

***Keywords*:** open world object detector, continual learning, open-set learning, non-parametric learning

1. Introduction

In recent years, object detection has been increasingly used in many practical applications, such as autonomous driving, video surveillance, and robotics. Several approaches, particularly based on Convolution Neural Networks (CNN),

DOI: https://doi.org/10.37256/aie.4220233058

have shown great performance improvements on some public datasets, e.g., Pascal VOC [1-4]. Despite their success, these methods have restricted their understanding to the limited number of categories present in the training data, based on the closed world assumption, (i.e., the number of categories is assumed to be fixed and known as a priori). While, this assumption is not true for real-world applications that contain an infinite range of visual input conditions (e.g., lightning, pose, and environment) and concepts; it is practically impossible to capture all visual information about the real world in a single dataset. From these perspectives, it is important to make detection methods robust against unknown objects, thus allowing them to act in an open-world setting [5].

To exemplify an open-ended learning framework, consider a robotic detection framework in Figure 1, that is initially learned to identify a limited number of categories. Given an unknown concept (for example, a cat) within an image, the robot must be able to detect it as an unknown object and learn it in subsequent learning steps. To achieve this objective, a robot vision system must possess two essential capabilities: (1) *open-set learning* (detecting both seen and unseen objects) [6] and, (2) *incremental learning* (extending its base knowledge with new classes without forgetting the previously learned ones) [7].

Unknown object detection is a fundamental aspect of OWOD, as it involves identifying objects that were not seen during training and do not have specific labels in the dataset. The difficulty in unknown object detection arises from the lack of explicit supervision. Unlike known objects, which have labeled examples in the training data, unknown objects do not have such annotations. As a result, during the training of object detection models, object proposals that include unknown objects would be erroneously penalized as background. This misclassification leads to low recall rates for detecting unknown objects. In an attempt to address this challenge, most OWOD methods have employed various heuristics to differentiate between unknown objects.

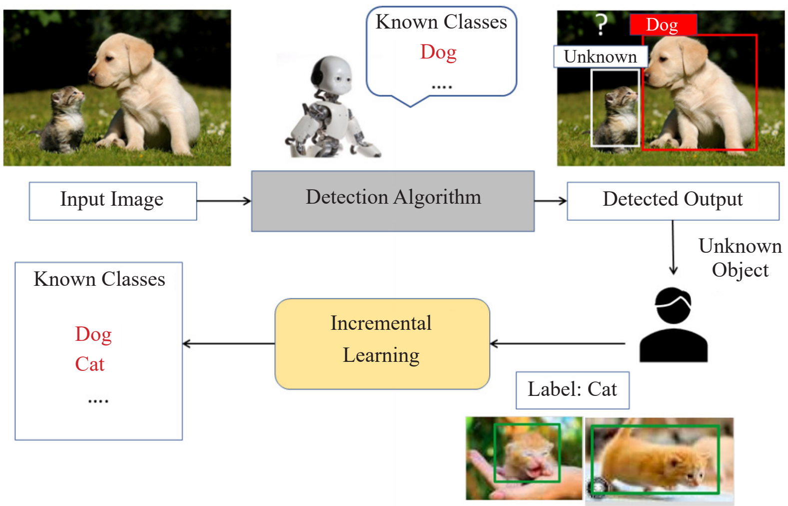

Figure 1. In an open world setting, a robotic vision system must be able to correctly detect the known object (dog) as well as the unseen object (cat). If the novel object is detected, its label and relative training images are provided through an external human source

Finally, the new training data is used to update the existing model through incremental learning and background during training. For instance, existing OWOD techniques [8, 9] use a pseudo-labeling scheme, where image patches with high backbone feature activation are considered to likely contain unknown objects. These patches are then used to create pseudo-labels, which are used for supervising the object detection model during training. While these heuristics have shown some improvement, the field of OWOD still has much room for improvement to achieve its goal of effectively detecting unknown objects in real-world scenarios. More research is needed to develop robust and generalizable methods that can accurately and reliably identify unknown objects without relying heavily on heuristics or pseudo-labeling schemes. In contrast to the previous methods that reason about known and unknown objects separately using labels and pseudo-labels, our approach takes a more direct and unified approach. We aim to learn a model with a non-parametric classifier head that is capable of generalizing to any object, whether it is known or unknown. The idea is that all objects, regardless of their category, possess general features that can distinguish them from the background. By learning these general features, the model can improve both the detection of unknown and known objects. In essence, our approach focuses on developing a more comprehensive understanding of the underlying features that define objects, without relying on explicit knowledge of specific categories. This non-parametric classifier head allows the model to detect and differentiate between objects and the background without explicitly labeling or pseudo-labeling unknown objects.

In this paper, we introduce a novel Open World Detection system known as OWOD-NP. In OWOD-NP, the classifier head of Faster R-CNN is replaced with a non-parametric classification architecture, which consists of the online estimated mean prototypes representing each known class and class-specific rejection thresholds that allow the detection of novel objects in a life-long fashion. We show the effectiveness of our approach on the class incremental object detection protocols and demonstrate that the inclusion of the non-parametric classifier in the object detector model significantly improves the performance in real-world scenarios. The main contributions of the paper are as follows:

• The problem setting of open world object detection is formalized, which explains the actual real-world scenarios.

• A novel open world detection (OWOD-NP) framework is developed by incorporating the non-parametric classifier in the Faster R-CNN model, which allows the detection of unknown objects and performs incremental learning in an end-to-end fashion.

• Extensive experiments are conducted on the PASCAL VOC to show the effectiveness of our method in the class incremental setting.

2. Related work

The Open World Object Detection (OWOD) task, first introduced by [8], has quickly gained significant attention in the research community [9-15] due to its potential real-world applications. In their work, [8] proposed an approach called ORE, an approach of open-set recognition of objects using energy-based models. ORE incorporates several key components including feature-space contrastive clustering, a Region Proposal Network (RPN)-based unknown detector, and an Energy-Based Unknown Identifier (EBUI). To identify unknown instances, ORE employed auto-labeling unknown supervised learning. Similar to ORE, existing OWOD technique models like [15] and [9] use the same spirit. Likewise, [13] proposed Unknown-Classified Open World Object Detection (UC-OWOD) that can identify both known and unknown objects by using an unknown label-aware proposal and an unknown discriminative classification head. Further, [14] made efforts to enhance ORE by minimizing the overlap between the feature-space distributions of known and unknown classes. To achieve this, they set the number of feature clusters to be equal to the number of classes. By doing so, they were able to reduce the confusion between known and unknown objects. On the other hand, [12] also extended the ORE approach by introducing a second objectness detection head, inspired by the work [16]. This additional localization-based objectness detection head aimed to improve the recall of unknown objects. Their results demonstrated the utility of integrating the objectness detection component, which contributes to better detection and classification of unknown objects in the open-world setting. [9] introduced the Open-World Detection Transformer (OW-DETR)-based method, which adopted the deformable DETR model for open-world object detection and utilizes a pseudo-labeling scheme to supervise the detection of unknown objects, where unmatched object proposals with high backbone activation are selected as potential unknown objects.

The recent advancements in Open World Object Detection (OWOD) have highlighted the importance of integrating objectness estimation and pseudo-labeling to achieve robust performance. However, prior approaches have typically treated objectness estimation and class prediction as separate components, and pseudo-labeling may require sampling unknown instances during training, which limits their effectiveness in diverse and dynamic real-world scenarios where prior knowledge of unknown objects is unavailable. In contrast, our method takes a more direct and robust approach to open-world object detection by formulating the problem without relying on any prior assumptions about unknown objects. This means that our approach is designed to handle diverse and dynamic environments without prior knowledge of the types of unknown objects that may appear. Moreover, our method introduces a direct integration of class prediction for identifying both known and unknown objects. This integration improves the detection of unknown objects, a critical challenge in open-world scenarios where new objects may emerge over time.

3. Proposed methodology

In this section, we discuss the proposed methodology for open world object detection. First, we formalize the problem of open world object detection and then discuss the proposed method that incorporates non-parametric learning into the Faster R-CNN model.

3.1 *Problem definition*

The main objectives of open world object detection include (i) identifying all object and non-object regions (background region), (ii) recognizing all the known object categories, (iii) identifying the unknown objects, and (iv) adding the new classes incrementally. Let the input space of the system be denoted by X and the output space as O. Under the closed world assumption, the input space is composed of a set of possible regions within an image and the output space contains the labels of known object classes and the background class (i.e., $O = K + b$ where $K = 1, 2, 3, ..., C$). During incremental learning, the output space evolves, and after the t^{th} incremental step, the output space becomes $O_t = K_t + b$. Under open world circumstances, the output should have a special category $U = \{C + 1, ...\}$ for unknown objects, extending output space as $O = K + b + U$. Hence, our primary goal is to develop a model that can classify various regions of interest from an input image as the known object label ($K + b$) or unknown object label (U). A brief overview of the problem set is illustrated in Figure 2.

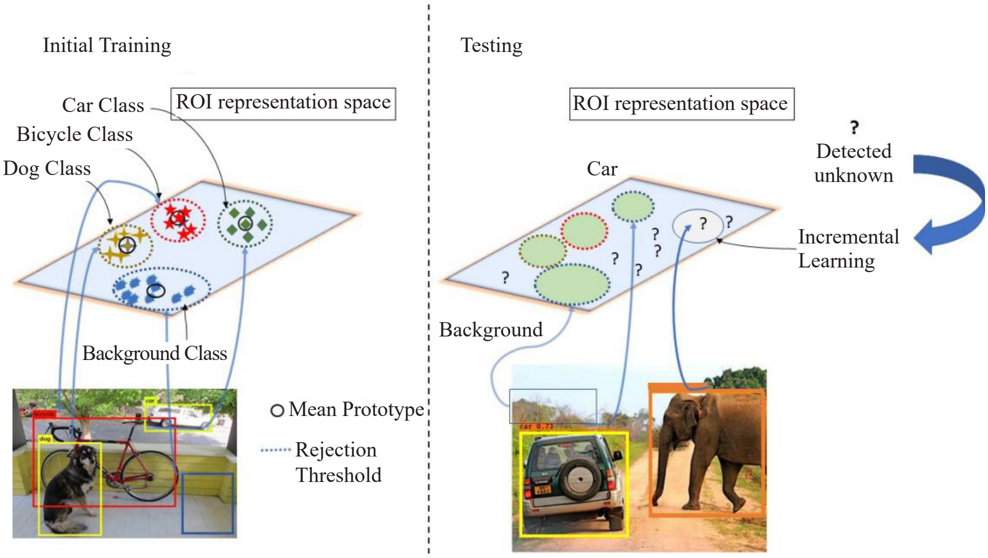

Figure 2. Overview of open world object detection problem setting: During the initial training phase, the model learned the mean prototypes and rejection thresholds that define the clusters of known classes (e.g., Car, Bicycle, and dog) and background class within Region of Interest (ROI) representation space. During test time, the samples that fall outside the threshold of known class clusters (green circle) are detected as unknown (denoted as ?). The incremental learning is performed to add a new unknown class to the existing representation space

3.2 *OWOD-NP: Open world object detection using non-parametric learning*

Open World Object Detection is capable of solving real-world challenges by detecting unknown objects and performing incremental learning on the detected novel objects without forgetting previously learned categories. We developed a new algorithm, OWOD-NP that addresses the difficult problem setting of open world detection.

The key component of our design is non-parametric learning that creates an efficient clustering of each class sample. The clustering at the intermediate latent space distinguishes the feature representations of known and unknown classes, allowing us to identify the unknown object as a novel category. In addition, it also facilitates incremental learning and significantly reduces forgetting by eliminating feature representation overlaps between the new and previously learned classes. To get optimal clusters, we employ the mean prototype and rejection threshold-based classifier head that creates a clear distinction between features related to known, unknown, and background regions.

The high-level architecture of OWOD-NP is shown in Figure 3. In OWOD-NP, the feature vector from the Region of Interest (RoI) of Faster R-CNN is provided to the non-parametric classifier head. Additionally, an embedded module is used to reduce the dimension of the pooled ROI feature vector, making the end-to-end learning process more stable [17]. In the following subsection, each of the components of OWOD-NP is briefly explained.

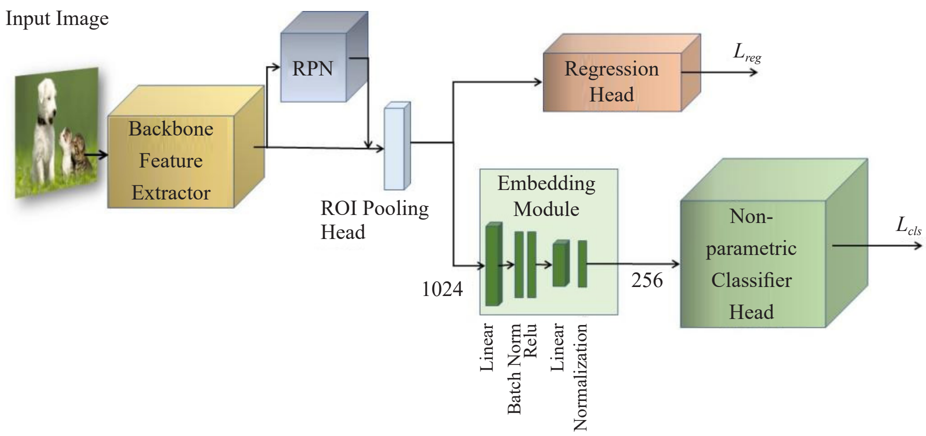

Figure 3. The proposed deep architecture of OWOD-NP that combines the Faster R-CNN model with a non-parametric classifier through an embedding module

3.2.1 *Non-parametric learning algorithm*

Existing approaches to the open world problems mostly use non-parametric learning techniques on top of learned metric spaces, which enforce features of the same class close together while dissimilar class features are pushed apart. The most common technique uses mean prototypes and rejection criteria, with each class represented by a mean feature vector, i.e., centroid [18]. It assigns a class label to the test sample based on the minimum distance between its representative features vector and the centroids. Given an m-dimensional feature vector $f(x)$ extracted from the intermediate ROI layer, the class label y of an input image x is computed by following distance criteria:

$$y = argmin_{\mathbf{c \epsilon Ct}} \| f(x) - \mu_c \| \tag{1}$$

When new samples arrive, the entire architecture of OWOD-NP is trained in an end-to-end manner, causing the feature representations $f(x)$ to be changed. During each iteration, class mean vectors are updated using an approximate strategy to track these changes. We compute the mean vectors for a mini-batch of samples $B = (x_1, k_1), \ldots, (x_b, k_b)$ by:

$$\mu_c^{t+1} = \frac{n_c \mu_c^t + n_{c,B} \mu_{c,B}^t}{n_c + n_{c,B}} \tag{2}$$

where n_c is the number of data points from class c that the network has seen upto the current training step t, $n_{c,B}$ is the number of data points from class c in the current batch, and $\mu_{B,c}$ is the current mini-batch mean vector belongs to class c features.

For the open-set scenarios, the class-specific rejection criteria are employed such that each class c exhibits a threshold or maximal distance Δ_c that is used to decide if the sample belongs to class c. The classification prediction thus becomes:

$$o(x) = \begin{cases} unk & \text{if } f(\phi(x), \mu_c) > \Delta_c \quad \forall_c \epsilon K_t \\ argmin_c f(\phi(x), \mu_c) & else \end{cases} \tag{3}$$

with $f(a,b) = \frac{1}{\sigma^2} \| a - b \|^2$. These thresholds are determined explicitly for each class by minimizing the following loss function:

$$L_\Delta(x,c) = \sum_{k \epsilon K_t} max\left(0, r\frac{1}{\sigma^2} \| \phi(x) - \mu_k \| - \Delta_c\right) \tag{4}$$

where r = -1 when $c = k$ and r = 1 otherwise.

To train the network, the following clustering loss is minimized. Formally, the loss term is given as follows with image x and its class label c.

$$l(x,c) = -log \frac{s_c(x)}{\sum_{k \epsilon K_t} s_k(t)} \tag{5}$$

where $s_c(x)$ denotes class-specific scores, which is defined as:

$$s_c(x) = \frac{e^{-\frac{1}{T}\|\phi(x)-\mu_c\|^2}}{\sum_{x \epsilon C_t} e^{-\frac{1}{T}\|\phi(x)-\mu_k\|^2}} \tag{6}$$

where T represents the temperature coefficient regulating the classifier's behavior. It is the variance of activation volume produced in feature space.

Overall, the training process involves two steps: first, the feature extractor on the training set minimizes the clustering loss function, and second, the rejection criteria are learned on the samples that were left out of the training set.

3.2.2 *Incremental learning*

After the unknown objects have been identified, an open world detector must be able to learn new classes when some of the unknown classes of interest are provided with labeled examples. Interestingly, retraining from scratch is not possible, all training data related to previous tasks are not available. Only practicing with new class instances will result in a catastrophic forgetting of previous classes. To alleviate the forgetting, we employ rehearsal and knowledge distillation methods.

• The rehearsal method is based on memory that keeps the most hard samples of each task. The number of bounding boxes determines the hard samples, the images with a large number of boxes are the most difficult to recognize. The rehearsal memory augments the new training set with previous task samples, allowing the network to retain the mean estimates of previous classes. The size of memory is kept constant and pruned after each incremental step to prevent unlimited expansion. Furthermore, a batch sampler is also used, which ensures that a specific proportion of the batch is made up of samples collected from the memory, regardless of the size of the memory. This approach prevents the incremental learning process from being biased toward novel categories [19].

• The Knowledge Distillation (KD) is based on parameter regularization which uses the network from the previous task and avoids deviating the important features for recognizing old classes. Typically, KD employs a teacher-student framework, in which the previous class knowledge is stored in the form of a frozen model that serves as the teacher, and a new learning model, which serves as the student, is used to create an adapted network for new classes. To maintain performance on older classes, the learning of the adapted network is restricted by introducing a distillation loss that minimizes the discrepancy between the response of two networks over the new training samples [20]. We apply KD feature-changing loss at Regional Proposal Network (RPN) and ROI pooling layer of faster RCNN, which considerably reduces forgetting [21].

4. Experiments and results

4.1 *Dataset and evaluation*

• **Dataset:** For the evaluation, we use the PASCAL VOC 2007 dataset [1], which is composed of 20 different classes, including, 10 K images and 25 K annotated objects. All 20 classes are alphabetically sorted and evenly divided into four different class groups to create a multi-class detection dataset. Each group contains images having at least one object category from that group's classes and bounding boxes of other class groups are excluded [21].

• Protocols for training and testing: For incremental learning under both closed and open world scenarios, the following training and testing protocols are used.

- Training: Using the example groups, the detection model should be trained in multiple incremental training sessions. The first training session uses traditional techniques to train an untrained model with the first example group. The pre-trained model from the previous training session is then used for subsequent sessions, and incremental learning is performed to train the model with data from the corresponding example group, allowing the model to recognize more object classes. Since the open world setting requires an unknown set of classes, we split class groups into two sections: two groups are referred to as known categories, and the other two are referred to as unknown classes. Only the known groups of classes are used for training purposes under both closed and open world scenarios.

- Testing: The testing protocol is formed by combining the testing sections of all the observed groups after every training session, resulting in a hybrid testing dataset. The model is evaluated on the hybrid dataset to see how well it performs on both previous and new classes. This protocol is used for both scenarios. While the open world setting additionally uses unknown groups during testing.

• Metrics for Evaluation: Mean Average Precision (mAP) at 0.5 IoU is used to measure the performance of the model.

The Average Precision for class (c) is defined as

$$mAP = \frac{1}{|\,classes\,|} \sum_{c \in classes} \frac{TP_{(c)}}{TP_{(c)} + FP_{(c)}} \tag{7}$$

where

True Positive $TP_{(c)}$: A proposal was made for class (c) and there actually was an object of class (c) in the ground truth.

False Positive $FP_{(c)}$: A proposal was made for class (c) but there is no object of class (c) in the ground truth.

For open world evaluation, Wilderness Impact (WI) metric is employed that characterizes the behavior of detector for unknown objects.

$$Wilderness\ Impact\ (WI) = \frac{P_K}{P_{K \cup U}} - 1 \tag{8}$$

where P_K represents the model's precision for known classes, and $P_{K \cup U}$ represents the model's precision when evaluated on both known and unknown classes. Ideally, WI should be lower, as unknown objects should not affect the model's precision over known object categories.

4.2 *Experimental set-up*

For all experiments, Stochastic Gradient Descent (SGD) is used to train the network with a weight decay of 10^3 and a momentum of 0.9. All networks are trained for 55,000 iterations. For the first 50 k iterations, the learning rate is set to 0.002 and then reduced to 0.0002 for the last 5,000 iterations.

We choose Resnet-50 Feature Pyramid Network (FPN), pre-trained with the Common Objects in Context (COCO) train2017 dataset, as the backbone of the model. We utilize the Faster R-CNN model from Torchvision, which is a PyTorch library. The entire framework is trained and tested using the Google Colab repository, which provides convenient access to Graphics Processing Unit (GPU) resources for accelerated computations. The embedding module, used in our architecture, is composed of two fully connected layers: the first layer with a width of 1,024, followed by batch normalization and Relu activation, and a final layer of width 256 with linear activation, followed by L2 normalization. For the rehearsal strategy, we use a fixed size memory of 800 samples, and a batch sampler to construct each batch having 50% of the instances from the memory. The class-specific threshold is updated using 20% of memory samples that were never used during training.

For performance evaluation, we compare results against the state-of-the-art class-incremental object detector-Incremental Learning Object Detection (ILOD) and Class Incremental Faster RCNN with Non-Parameteric (CIFRCN-NP) [21].

4.3 *Results*

4.3.1 *Closed world detection*

For the closed world, we compare model performance with and without the rejection criteria. In OWOD-NP without the rejection option, the model is only evaluated on the known set of groups, removing the chance of classifying a known sample as unknown. This scenario evaluates the model's ability to correctly classify samples into the specified categories.

With rejection criteria, the model can classify the sample as one of the known or unknown categories. As samples from known groups may be misclassified as unknowns, this situation is more complicated than the previous one.

Table 1. Comparison test result of incremental learning under the closed world scenario without rejection criteria on the VOC 2007, when the groups of five classes are added sequentially. The rows show results of two known groups added during each incremental step

classes added	algorithms	A	B	all (mAP)
1-5	ILOD	63.9	-	63.9
	CIFRCN-NP	56.8	-	56.8
	OWOD-NP	58.28	-	58.28
6-10	ILOD	44.8	59.2	52.0
	CIFRCN-NP	40.4	68.8	54.6
	OWOD-NP	**50.55**	56.23	53.40

• **Without Rejection Criteria:** The results under the closed-world scenario without rejection criteria are compared in Table 1.

- **Overall performance:** After incremental training sessions, the learned model could face some difficulties in identifying objects from the oldest classes. The model focuses solely on new classes and suffers from a notable case of knowledge forgetting. ILOD has mitigated the forgetting phenomenon to some extent, and CIFRCN-NP has also shown less forgetting. While our method significantly reduces forgetting and archives 10 mAP gain.

- Performance on old classes: As Group A is evaluated in each session, so we compare the results for that particular group. Even when using only new classes of images during an incremental step, our OWOD-NP retains the majority of previously learned knowledge. At first, our OWOD-NP achieves 58.28% mAP on Group A, and after one increment process, it achieves 50.58% mAP, which is considerably better than other strategies, suggesting that the proposed OWOD-NP is capable of overcoming catastrophic forgetting to a large extent.

- Performance on new classes: For Group B, OWOD-NP attains mAPs of 56%, illustrating that the proposed framework does not prevent the model from learning new information.

• **With Rejection Criteria:** The comparison of the closed world with rejection criteria is given in Table 2. It is clearly observed that our approach based on the rejection threshold rejects only a few known samples. We compare the performance of OWOD-NP with that of OWOD-NP having no rejection option, and CIFRCN-NP. OWOD-NP without rejection is our upper bound in terms of performance in the closed world because it does not reject any instance of known classes as unknown. The purpose of this baseline is to show that the rejection criteria are always applicable.

Table 2. Comparison test result of incremental learning under the closed world scenario with rejection criteria on the VOC 2007, when the groups of five classes are added sequentially. The rows show results of two known groups added during each incremental step

classes added	algorithms	A	B	all (mAP)
1-5	CIFRCN-NP	56.8	-	56.8
	no rejection	58.28	-	58.28
	OWOD-NP	47.98	-	47.98
6-10	CIFRCN-NP	40.4	68.8	54.6
	no rejection	50.55	56.23	53.40
	OWOD-NP	**48.94**	46.93	47.94

4.3.2 *Open world detection*

Table 3 compares the results of OWOD-NP against Faster R-CNN using the proposed open world evaluation protocol. WI metrics are used to measure how the unknown instances are confused with known samples after each learning step. Overall, OWOD-NP has achieved fewer WI scores, indicating that it is less affected by unknown classes. OWOD-NP outperforms Faster R-CNN during incremental learning steps, demonstrating its effectiveness for the open-world scenario.

Table 3. Comparison test result of incremental learning under the open world scenario on the VOC 2007, when the groups of five classes are added sequentially. The rows show results when both known and unknown classes are used for evaluation

classes added	algorithm	A	B	all (mAP)	WI
1-5	Faster RCNN	49.32	-	49.32	0.2567
	OWOD-NP	29.05	-	29.05	0.65
6-10	Faster RCNN	0	50	25.44	0.2020
	OWOD-NP	**42.62**	36.62	**39.68**	**0.208**

5. Conclusion

The Open World Object Detection (OWOD) task is a complex and challenging objective that involves integrating various aspects of generalized open-set object detection and incremental learning. The detection and understanding of unknown objects play a crucial role in ensuring the robustness of OWOD methods. To address this challenge, we proposed a novel framework called Open World Object Detection with Non-Parametric Classification (OWOD-NP), which significantly improves the detection of unknown objects in the benchmark. Our method OWOD-NP integrates the non-parametric learning approach into the standard Faster R-CNN model, enabling the detection of unseen objects and supporting incremental learning in an end-to-end fashion. Through comprehensive ablations, we shed light on the underlying mechanisms of our approach and demonstrate the benefits of each component. Extensive experiments on popular benchmark datasets clearly show the effectiveness of the proposed method. However, there is still ample room for improvement, not only in the realm of unknown object detection but also in other aspects of the OWOD task. Future research efforts will be dedicated to addressing the challenges of open-world detection in single-stage object detectors such as You Only Look Once (YOLO) and Single Shot Detector (SSD), further advancing the field of OWOD.

Acknowledgements

This work was partially supported by the New York University Abu Dhabi (NYUAD) Center for Artificial Intelligence and Robotics (CAIR), funded by Tamkeen under the NYUAD Research Institute Award CG010.

Conflict of interest

The authors declare that there are no personal or organizational conflicts.

References

[1] Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*. 2010; 88(2): 303-338.

[2] Jaiswal S, Jaiswal T. Deep learning approaches for object detection. *Artificial Intelligence Evolution*. 2020; 1(2): 122-144.

[3] Jaiswal T, Pandey M, Tripathi P. Real time multiple-object detection based on enhanced ssd. In: *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*. Raipur, India: IEEE; 2022. p. 1-5. Available from: doi: 10.1109/ICPC2T53885.2022.9776899.

[4] Jaiswal T. Image captioning through cognitive iot and machine-learning approaches. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*. 2021; 12(9): 333-351.

[5] Bendale A, Boult T. Towards open world recognition. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE; 2015. p. 1893-1902. Available from: doi: 10.1109/CVPR.2015.7298799.

[6] Scheirer WJ, Rezende Rocha A, Sapkota A, Boult TE. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012; 35(7): 1757-1772.

[7] Castro FM, Marín-Jiménez MJ, Guil N, Schmid C, Alahari K. End-to-end incremental learning. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Germany: Springer; 2018. p. 233-248.

[8] Joseph KJ, Khan S, Khan FS, Balasubramanian VN. Towards open world object detection. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE; 2021. p. 5826-5836. Available from: doi: 10.1109/CVPR46437.2021.00577.

[9] Gupta A, Narayan S, Joseph KJ, Khan S, Khan FS, Shah M. Ow-detr: Open-world detection transformer. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE; 2022. p. 9225-9234. Available from: doi: 10.1109/CVPR52688.2022.00902.

[10] Ma Z, Yang Y, Wang G, Xu X, Shen HT, Zhang M. Rethinking open-world object detection in autonomous driving

scenarios. In: *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*. New York, NY, USA: Association for Computing Machinery; 2022. p. 1279-1288. Available from: doi: 10.1145/3503161.3548165.
[11] Singh DK, Rai SN, Joseph KJ, Saluja R, Balasubramanian, VN, Arora C, et al. Order: Open world object detection on road scenes. In: *Machine Learning for Autonomous Driving Workshop at the 35th Conference on Neural Information Processing Systems (NeurIPS 2021).* Sydney, Australia; 2021. p. 1-12.
[12] Wu Y, Zhao X, Ma Y, Wang D, Liu X. Two-branch objectness-centric open world detection. In: *Proceedings of the 3rd International Workshop on Human-Centric Multimedia Analysis (HCMA '22)*. New York, NY, USA: Association for Computing Machinery; 2022. p. 35-40.
[13] Wu Z, Lu Y, Chen X, Wu Z, Kang L, Yu J. Uc-owod: Unknown-classified open world object detection. In: *Computer Vision-ECCV 2022: 17th European Conference*. Tel Aviv, Israel: Springer; 2022. p. 193-210.
[14] Yu J, Ma LY, Li ZL, Peng Y, Xie SR. Open-world object detection via discriminative class prototype learning. In: *2022 IEEE International Conference on Image Processing (ICIP)*. Bordeaux, France: IEEE; 2022. p. 626-630. Available from: doi: 10.1109/ICIP46576.2022.9897461.
[15] Zhao X, Liu X, Shen Y, Ma Y, Qiao Y, Wang D. *Revisiting open world object detection*. arXiv. 2022. Available from: doi: 10.48550/arXiv.2201.00471.
[16] Kim D, Lin T-Y, Angelova A, Kweon IS, Kuo W. Learning open-world object proposals without learning to classify. *IEEE Robotics and Automation Letters*. 2022; 7(2): 5453-5460.
[17] Karlinsky L, Shtok J, Harary S, Schwartz E, Aides A, Feris R, et al. Repmet: Representative-based metric learning for classification and few-shot object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA: IEEE; 2019. p. 5197-5206.
[18] Fontanel D, Cermelli F, Mancini M, Buló SR, Ricci E, Caputo B. Boosting deep open world recognition by clustering. *IEEE Robotics and Automation Letters*. 2020; 5(4): 5985-5992. Available from: doi: 10.1109/LRA.2020.3010753.
[19] Liu XL, Yang H, Ravichandran A, Bhotika R, Soatto S. M*ulti-task incremental learning for object detection*. arXiv. 2020. Available from: doi: 10.48550/arXiv.2002.05347.
[20] Hinton G, Vinyals O, Dean J. *Distilling the knowledge in a neural network*. arXiv. 2015. Available from: doi: 10.48550/arXiv.1503.02531.
[21] Hao Y, Fu Y, Jiang Y-G, Tian Q. An end-to-end architecture for class-incremental object detection with knowledge distillation. In: *2019 IEEE International Conference on Multimedia and Expo (ICME*). Shanghai, China: IEEE; 2019. p. 1-6. Available from: doi: 10.1109/ICME.2019.00009.
[22] Shmelkov K, Schmid C, Alahari K. Incremental learning of object detectors without catastrophic forgetting. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy: IEEE; 2017. p. 3420-3429. Available from: doi: 10.1109/ICCV.2017.368.