

Article

A Parallel Open-World Object Detection Framework with Uncertainty Mitigation for Campus Monitoring

Jian Dong ^{1,2,†}, Zhange Zhang ^{1,†}, Siqi He ³, Yu Liang ⁴ , Yuqing Ma ^{1,5,*}, Jiaqi Yu ⁶, Ruiyan Zhang ⁶ and Binbin Li ²

¹ State Key Lab of Software Development Environment, Beihang University, Beijing 100191, China; dongjian@cesi.cn (J.D.); zhangezhang@buaa.edu.cn (Z.Z.)

² China Electronics Standardization Institute, Beijing 100007, China; libb@cesi.cn

³ School of Computer Science, Peking University, Beijing 100871, China; hesq47@foxmail.com

⁴ Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; yuliang@bjut.edu.cn

⁵ Institute of Artificial Intelligence, Beihang University, Beijing 100191, China

⁶ Beijing Institute of Control and Electronic Technology, Beijing 100038, China; yujq2004@163.com (J.Y.); zhangruiyan65@foxmail.com (R.Z.)

* Correspondence: mayuqing@buaa.edu.cn

† These authors contributed equally to this work.

Abstract: The recent advancements in artificial intelligence have brought about significant changes in education. In the context of intelligent campus development, target detection technology plays a pivotal role in applications such as campus environment monitoring and the facilitation of classroom behavior surveillance. However, traditional object detection methods face challenges in open and dynamic campus scenarios where unexpected objects and behaviors arise. Open-World Object Detection (OWOD) addresses this issue by enabling detectors to gradually learn and recognize unknown objects. Nevertheless, existing OWOD methods introduce two major uncertainties that limit the detection performance: the unknown discovery uncertainty from the manual generation of pseudo-labels for unknown objects and the known discrimination uncertainty from perturbations that unknown training introduces to the known class features. In this paper, we introduce a Parallel OWOD Framework with Uncertainty Mitigation to alleviate the unknown discovery uncertainty and the known discrimination uncertainty within the OWOD task. To address the unknown discovery uncertainty, we propose an objectness-driven discovery module to focus on capturing the generalized objectness shared among various known classes, driving the framework to discover more potential objects that are distinct from the background, including unknown objects. To mitigate the discrimination uncertainty, we decouple the learning processes for known and unknown classes through a parallel structure to reduce the mutual influence at the feature level and design a collaborative open-world classifier to achieve high-performance collaborative detection of both known and unknown classes. Our framework provides educators with a powerful tool for effective campus monitoring and classroom management. Experimental results on standard benchmarks demonstrate the framework's superior performance compared to state-of-the-art methods, showcasing its transformative potential in intelligent educational environments.

Keywords: intelligent campus monitoring; uncertainty mitigation; objectness-driven object discovery; parallel network; open-world object detection; artificial intelligence



Citation: Dong, J.; Zhang, Z.; He, S.; Liang, Y.; Ma, Y.; Yu, J.; Zhang, R.; Li, B. A Parallel Open-World Object Detection Framework with Uncertainty Mitigation for Campus Monitoring. *Appl. Sci.* **2023**, *13*, 12806. <https://doi.org/10.3390/app132312806>

Academic Editor: Sungho Kim

Received: 25 October 2023

Revised: 17 November 2023

Accepted: 20 November 2023

Published: 29 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The continuous advancement of deep learning [1–4] technology has enabled the widespread application of artificial intelligence (AI) in multiple domains [5–7], with the field of education as a notably impacted area. Object detection [8–12] stands out as one of the key technologies driving this transformation in education, which plays a significant role in shaping intelligent campus and can be harnessed for various purposes, including

campus monitoring and class behavior detection. Zhou et al. [8] design an object detection algorithm to detect the raising hand action in real classroom scenarios. Zhao et al. [9] propose CNPH-Net with a multiscale feature extraction module to enhance the capability of small object detection in classroom scenarios. Xu et al. [13] applied object detection and gaze-tracking technology to analyze the students' attention in the classroom. These methods are all based on traditional closed-set object detection techniques, meaning that the models can only detect objects that are already present in the training set. However, due to the openness and diversity of objects in real campus and classroom scenarios, instances of objects and behaviors not included in the training datasets may arise. Traditional object detection methods are ill-equipped to detect such unexpected objects, making it challenging to meet the demands of real-world campus monitoring.

To address the challenges of detection in open environments, such as campuses and classrooms, a new object detection task, Open-World Object Detection (OWOD) [14], is proposed to focus on the detection of objects in real-world open environments. This problem requires detection models not only to detect known objects but also to discover and incrementally learn about previously unseen unknown objects during the training phase. In the classroom and similar education scenarios, the OWOD models are required to discover and locate unknown targets and behaviors, and subsequently engage in incremental learning, which helps to facilitate intelligent monitoring of the teaching process.

Previous OWOD works mainly realized unknown class detection by designing complex unknown proposal generation mechanisms. The ORE method [14] incorporates an unknown object-aware Region Proposal Network (RPN), selecting top-k region proposals sorted by the objectness of the RPN from background as unknown objects, thereby endowing the model with the ability to detect unknown objects. Zhao et al. [15] further introduces an auxiliary proposal advisor that combines an unknown-aware RPN with the selective search method to generate more accurate pseudo-labels for unknown categories. Building on the recent success of closed-set detector D-DETR [16], OW-DETR [17] proposes an unknown pseudo-labeling method that selects query boxes with high attention scores which not match with the ground truth of known class.

However, in these OWOD works, two major uncertainties that limit the detection performance are commonly observed: the unknown discovery uncertainty and the known discrimination uncertainty. The unknown discovery uncertainty indicates the uncertain bias from the manual generation of pseudo-labels for unknown objects. In the previous OWOD works, the unknown regions were largely derived from high-score region proposals generated by RPN, which may potentially cover some background areas that introduce great uncertain bias. The known discrimination uncertainty arises from the uncertain perturbations that the learning of unknown classes introduces to the features of known class discrimination. The differentiation among known classes primarily relies on discriminative features, whereas the learning of unknown classes guides the model to extract more generic features among objects, which are only distinguished from the background. Therefore, while granting the model the ability to detect unknown classes, previous methods often resulted in an uncertain perturbation in the discrimination performance of known classes. These two kinds of uncertainty have impeded further improvements in detection performance in open-world scenarios, making it challenging to apply them to real-world environments like campus.

In this paper, we introduce a Parallel OWOD Framework to mitigate the unknown discovery uncertainty and the known discrimination uncertainty within the OWOD task. Specifically, the framework employs a parallel architecture to decouple the process of learning unknown classes from the detection of known classes, thereby preventing the influence of uncertain features learned through unknown guidance on the known classes. Through the parallel architecture, our framework successfully mitigates the known discrimination uncertainty during open-world training. To alleviate the unknown discovery uncertainty, we have designed and added an objectness-driven discovery module to replace the uncertain unknown object selection, driving the model to focus on the objectness of

various objects derived from the deterministic supervision of known classes. Therefore, the Objectness-Driven Discovery Module can learn objectness scores from stable supervision of known classes. For all objects, including both known and unknown classes, their objectness scores are expected to be significantly higher than those of the background regions. Consequently, the objectness-driven discovery module is capable of identifying potential objects with objectness more certainly from the background, including unknown objects. Furthermore, the output from the objectness-driven discovery module and a close-world known object detector is subsequently reconciled via a collaborative open-world classifier, which eliminates redundant unknown detection instances and facilitates collaborative learning of known and unknown classes. As shown in Figure 1, this framework may provide a powerful tool for educators and administrators to effectively monitor campus and manage classroom behaviors.

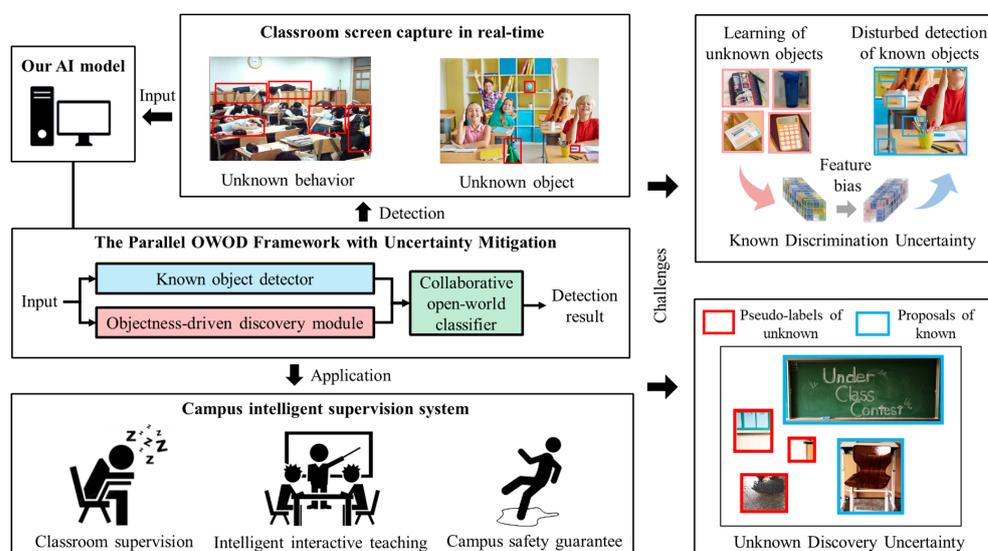


Figure 1. Introduction of our parallel OWO framework. Our framework will provide a powerful tool for educators and administrators to effectively campus monitoring. The framework effectually mitigates the unknown discovery uncertainty and the known discrimination uncertainty in the OWO task.

We summarize our contributions as follows:

- (1) We propose a Parallel OWO Framework comprised of two different detectors to mitigate the unknown discovery uncertainty and the known discrimination uncertainty within OWO, which can be used for effective management and monitoring of campus to ensure a conducive and productive learning atmosphere.
- (2) An objectness-driven discovery module is trained to guide the model to focus on the objectness of various objects derived from deterministic supervision of known classes, which could help alleviate the unknown discovery uncertainty.
- (3) A known object detector and a collaborative open-world classifier are designed to accomplish the collaborative learning of known and unknown classes, helping mitigate the known discrimination uncertainty.
- (4) Experiments on common benchmarks demonstrate that our framework shows superior results compared to state-of-the-art methods.

The upcoming sections of this article will delve into our parallel open-world object detection framework designed for campus monitoring, with a specific focus on uncertainty mitigation. Section 2 offers a review of existing methods for object detection in campus scenarios, open-set object detection, and open-world object detection. Following that, Section 3 details our proposed parallel open-world object detection framework, including its components and training process. In Section 4, we present the experimental evaluation of the proposed framework on standard benchmarks, as well as an ablation study, a parameter

analysis, and a visualization of the results. Finally, Section 5 summarizes our findings, outlines limitations, and explores potential avenues for future research.

2. Related Work

The continual progress in deep learning technologies has extensively applied AI across diverse domains, such as autonomous driving [18–20], medicine [21–23], and intelligent education. In the field of intelligent education, artificial intelligence plays a pivotal role in predicting students' performance [24,25], comprehending how students learn [26], detecting their behaviors [27], and fulfilling a series of other educational requirements. Notably, the development of object detection technology has elevated intelligent campus monitoring to a notable research and application domain within the field of intelligent education. In order to achieve intelligent campus monitoring, previous efforts were primarily based on the application of traditional closed-set object detection techniques in educational settings such as classrooms. On the other hand, to handle the challenge in the real-world object detection that an object detector may encounter several unknown objects which not appear in the training process, previous approaches have explored open-set and open-world settings. In this section, we discuss related work on object detection in campus scenarios, as well as work related to open-set and open-world target detection.

2.1. Object Detection in Campus Scenarios

With the advancement of deep learning, object classification and detection techniques have found application in various domains [28,29]. Existing object detection approaches in campus scenarios often involve the direct application of traditional closed-set detection models to the classroom environment. Ref. [9] proposes a single-stage object detector called CBPH-Net and designs a feature extraction module to capture more channel information and relevant features to enhance the multiscale recognition capability in classrooms. Ref. [8] proposes an automatic hand-raiser recognition algorithm to show who raises their hands in real classroom scenarios, which is of great importance for further analyzing the learning states of individuals. Ref. [13] proposes a fusion model based on gaze tracking and object detection to intelligently analyze the students' attention in the classroom from the first-person perspective and promote teachers' precise teaching and students' personalized learning. Although these methods successfully apply object detection to campus scenarios, they lack the capability to detect and learn from unknown targets, rendering them inadequate for meeting the detection requirements within the open environments of campus and classrooms.

2.2. Open-Set Object Detection

In the open-set setting, the incomplete information extracted from the training set causes the model couldn't classify the unknown categories that are not encountered during the training phase. Based on several assumptions, a number of previous works [30–35] attempted to step into open-set tasks. A metric learning framework [36] and a model with OpenMax classifier [37] have been proposed to address open-set classification. Some works also exploited self-supervised learning [38] and unsupervised learning [39] to handle open-set classification tasks. Open-set object detection protocol was first proposed in [40], followed by works [41–43] that improved detection performance by measuring object uncertainties. OpenDet [44] designed a manual unknown-discover strategy based on feature density to enhance the identification of unknown proposals.

2.3. Open-World Object Detection

Compared with open-set tasks, open-world tasks are more demanding since these tasks require not only identifying unknown classes, but also learning incrementally based on newly obtained category data. Open-world tasks go beyond open-set by requiring incremental learning based on newly obtained category data. Bendale et al. [45] proposed the first open-world classification model, and Xu et al. [46] introduced a meta-learning

method for identifying unknown classes. Previous works [36,47,48] explored more complex scenarios, e.g., long-tail distribution [49], few-shot learning [50] and zero-shot learning [51] on the open-world classification, respectively.

ORE, the first OWOD method proposed by Joseph et al. [14] designed a specific RPN, selecting top-k region proposals sorted by the objectness of the RPN from the background to discover unknown objects. SA proposed by Yang et al. [52] defined semantic centroids in feature space to exploit embedded semantic topology. Wu et al. [53] designed a two-stage detector distinguishing diverse classes on the basis of clustering and similarity to handle the raised Unknown-Classified OWOD problem. Zhao et al. [15] proposed a model coupling an auxiliary proposal advisor and a class-specific expelling classifier to improve the detection performance of unknown classes. Inspired by the powerful presentation ability of transformer [54], Gupta et al. [17] first extended transformer to OWOD problem and proposed a transformer-based open-world detector OW-DETR, which combines an attention-driven pseudo-labeling scheme for selecting unknown query boxes and an objectness branch to effectively separate foreground objects and background.

Previous methods used complex strategies to generate pseudo-labels for unknown classes but introduced excessive uncertainties, harming the learning of unknown objects and influencing the known classification as well. In contrast, our method only explores the unknown information from the known instances through a reasonable decoupling framework, which improves the detection performance of the unknown objects while maintaining that of the known objects.

3. Method

In this section, we introduce our parallel OWOD framework which is designed to mitigate the detection uncertainties of known and unknown objects in open environments such as campuses and classrooms, with the primary objective of facilitating campus monitoring. We begin by analyzing the task of open-world object detection in Section 3.1. Subsequently, we provide a comprehensive overview of the proposed framework's structure in Section 3.2. Finally, Section 3.3 outlines the training process of our framework.

3.1. Problem Analysis

In the setting of open-world object detection, the whole training process is separated into T tasks. At any task $t \in \{1, \dots, T\}$, we consider the set of known classes as \mathcal{K}^t and the set of unknown classes as \mathcal{U}^t , where $\mathcal{K}^t \cap \mathcal{U}^t = \emptyset$. During the training phase at task t , the annotations of known classes \mathcal{K}^t are accessible to ensure the model has the capacity to classify the known instances into correct classes accurately, i.e., an instance $\mathbf{x}_i^{(\mathcal{K}^t)}$ of class $c \in \mathcal{K}^t$ is annotated as $\mathbf{y}_i^{(\mathcal{K}^t)} = [l_i^{(\mathcal{K}^t)}, b_i^{(\mathcal{K}^t)}]$, where $l_i^{(\mathcal{K}^t)} = c \in \mathcal{K}^t$ denotes corresponding label and $b_i^{(\mathcal{K}^t)} = [x, y, w, h]$ denotes the coordinates of corresponding bounding box. In contrast, the unknown classes \mathcal{U}^t appear during the training phase without annotations but are required to be classified as an *unknown* denoted by label $l_i^{(\mathcal{K}^t)} = c \in \mathcal{K}^t$ when inferring.

For incremental learning, part of the unknown classes $\bar{\mathcal{U}}^t \in \mathcal{U}^t$ are labeled and update the known classes as $\mathcal{K}^{t+1} = \mathcal{K}^t \cup \bar{\mathcal{U}}^t$ and unknown classes as $\mathcal{U}^{t+1} = \mathcal{U}^t \setminus \bar{\mathcal{U}}^t$ at task $t + 1$. The model adaptively updates itself with new knowledge and identifies current known classes \mathcal{K}^{t+1} accurately while classifying current unknown classes \mathcal{U}^{t+1} . This cycle continues over the life of the object detector.

3.2. Parallel OWOD Framework with Uncertainty Mitigation

Our parallel OWOD framework is based on the state-of-the-art object detection model D-DETR [16]. As illustrated in Figure 2, our framework comprises an objectness-driven discovery module, a known object detector, and a collaborative open-world classifier. The framework utilizes a parallel architecture to decouple the learning processes to address the challenges of uncertainties within OWOD. The objectness-driven discovery module converts all known objects into the same class and guides the model to focus on the object-

ness derived from the supervision of different objects, thereby alleviating the unknown discovery uncertainty. The objectness-driven discovery module effectively aids in the discovery of more potential objects distinct from the background including unknown objects. The known object detector retains the fundamental approach of D-DETR to ensure precise detection performance for known classes. Detection of known and unknown classes is carried out separately by different sub-modules within the parallel architecture, which ensures that the learning of known and unknown classes does not interfere with each other, effectively alleviating the known discrimination uncertainty.

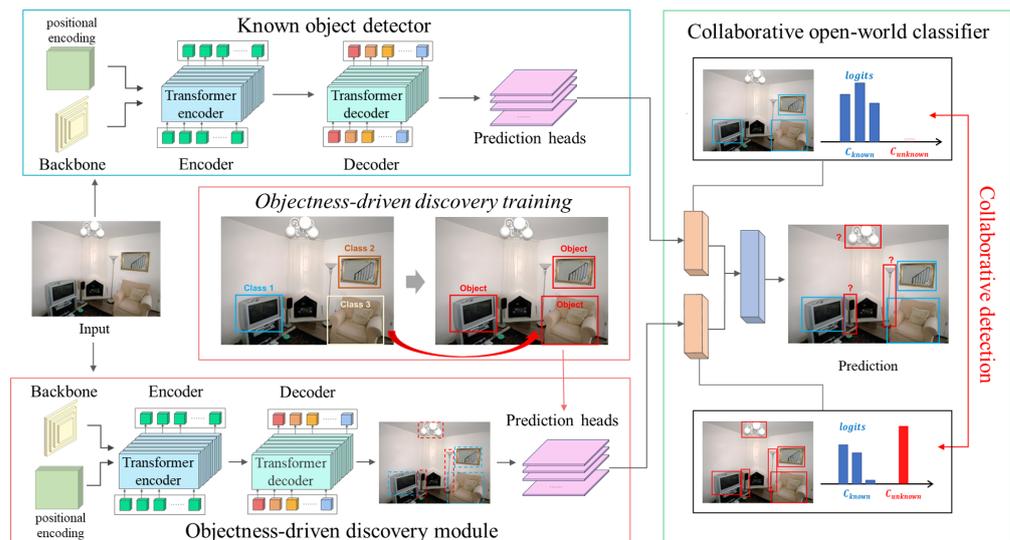


Figure 2. Architecture of our parallel OWO framework. The parallel OWO framework comprises an objectness-driven discovery module, a known object detector, and a collaborative open-world classifier.

During the inference phase, the image is separately processed by both the objectness-driven discovery module and the known object detector. The outputs are then integrated through the collaborative open-world classifier, facilitating cooperative learning for both known and unknown classes. The framework ultimately leads to strong detection performance for both known and unknown classes.

3.2.1. Objectness-Driven Discovery Module

To address the challenge posed by the unknown discovery uncertainty introduced from the strategy of unknown pseudo-labeling, we introduced an objectness-driven discovery method to learn the unknown objects certainly from the deterministic supervision of known classes.

Since both unknown and known classes inherently share the common characteristic of being “generalized objects”, we can learn objectness from known classes to guide the learning of unknown classes. We simply convert all known object classes to a generalized class and only distinguish them from background regions without objects, extracting generalized features from all known classes while attenuating discriminative features to enhance the model’s potential for object discovery. Through this approach, the detector can not only detect known objects but also discover other unknown objects with universities that are similar to known ones. Compared to previous methods for unknown object discovery, our objectness-driven discovery module extracts stable objectness from the supervised information of known classes, rather than manually selecting potential regions for unknown objects from the background. Therefore, our framework effectively reduces the uncertainties and enhances the detection performance of unknown objects.

Specifically, during the training phase of the detector, the instances of all known classes except the background region are converted to a universal class “generalized object”. Given

the i -th instance of known classes $\mathbf{x}_i^{(\mathcal{K}^t)}$ with its annotations $\mathbf{y}_i^{(\mathcal{K}^t)}$, the supervision label used for classification regularization could be denoted as $\tilde{\mathbf{y}}_i$:

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i^{(\mathcal{G}^t)} = [l^{(\mathcal{G}^t)}, b_i^{(\mathcal{K}^t)}] \quad (1)$$

where $\mathbf{y}_i^{(\mathcal{G}^t)}$ denotes the generalized annotation of the instance after convert, and $l^{(\mathcal{G}^t)} = o$ denotes the universal label "generalized object" of every known instance. The classification regularization for the objectness-driven discovery module can be presented as follows:

$$\begin{aligned} \tilde{\ell}_{\text{universal}}^{\text{cls}} &= - \sum_i \tilde{\mathbf{y}}_i \log \mathbf{x}_i^{(\mathcal{K}^t)} \\ &= - \sum_i \mathbf{y}_i^{(\mathcal{G}^t)} \log \mathbf{x}_i^{(\mathcal{K}^t)} \end{aligned} \quad (2)$$

where $\tilde{\ell}_{\text{universal}}^{\text{cls}}$ is the loss of classification regularization calculated by the universal label for the objectness-driven discovery module.

Through the objectness-driven discovery module, our framework is capable of identifying generalized objects possessing the attribute of "being an object" in the input image, including numerous unknown objects that have never been encountered in the training phase.

3.2.2. Known Object Detector

To mitigate the known discrimination uncertainty, a known object detector built upon the D-DETR [16] is employed to operate in parallel with the objectness-driven discovery module to accurately detect known classes. In order to preserve the precise detection performance of known category objects in the framework. The Known Object Detector is designed to learn discriminative features that are as relevant as possible to the categories, enabling it to distinguish between different known categories effectively.

Specifically, during the training phase of the known object detector, we treat the known proposals as their origin label. The supervision label used for classification regularization of known proposal i could be denoted as \mathbf{y}_i :

$$\mathbf{y}_i = \mathbf{y}_i^{(\mathcal{K}^t)} \quad (3)$$

The classification regularization for the known object detector can be presented as follows:

$$\begin{aligned} \ell_{\text{known}}^{\text{cls}} &= - \sum_i \mathbf{y}_i \log \mathbf{x}_i^{(\mathcal{K}^t)} \\ &= - \sum_i \mathbf{y}_i^{(\mathcal{K}^t)} \log \mathbf{x}_i^{(\mathcal{K}^t)} \end{aligned} \quad (4)$$

where $\ell_{\text{known}}^{\text{cls}}$ is the loss of classification regularization for the known object detector, and $\mathbf{y}_i^{(\mathcal{K}^t)}$ denotes original label of proposal i .

At each task, the known object detector is independently trained. During testing, the bounding boxes and logits produced by the Known Object Detector are input into the Collaborative open-world Classifier to facilitate the fusion of results with the detection of unknown categories.

3.2.3. Collaborative Open-World Classifier

To accomplish the collaborative learning of known classes and unknown classes, we introduce a collaborative open-world classifier to integrate the output of the two detectors of our framework and eliminate the impact of learning unknown classes on the detection performance of known classes.

Considering that the generalized objects logits predicted by the objectness-driven discovery module may overlap with the output of the known object detector with high probability, the collaborative open-world classifier first filters the generalized objects logits

by calculating an Intersection over Union (IoU) matrix $\mathcal{P} = \{p_{mn}\}$ between the generalized object boxes of the objectness-driven discovery module $\mathcal{B}^G = \{b_m^G\}$ and the known object boxes $\mathcal{B}^K = \{b_n^K\}$. The calculation of matrix \mathcal{P} could be formulated as follows:

$$p_{ij} = \text{IoU}(b_m^G, b_n^K) \quad (5)$$

where b_m^G denotes each generalized object box in \mathcal{B}^U and b_n^K denotes each known object box in \mathcal{B}^K . Subsequently, any generalized bounding box is labeled as a known class if it satisfies that its IOU with any predicted bounding box of a known class exceeds a threshold P_t . Conversely, the remaining generalized bounding boxes are categorized as unknown class targets. The filter process of the generalized object box could be presented as follows:

$$\mathcal{B}^U = \{b_m^G | \forall n, p_{mn} \leq P_t\} \quad (6)$$

The final output \mathcal{B}^{out} of the collaborative open-world classifier is formed by merging the filtered unknown class predicted boxes with the original known class predicted boxes:

$$\mathcal{B}^{out} = \mathcal{B}^U \cup \mathcal{B}^K \quad (7)$$

The objective of the collaborative open-world classifier is to achieve collaborative detection of both known and unknown class targets by extracting unknown class targets with the attribute of "object" from the generalized objects outputted by the objectness-driven discovery module, excluding objects belonging to the known class.

3.3. Training Process

In accordance with the OWOD setup, the training of our framework consists of two stages: the initial training stage of the task 1 and the incremental learning stage. During the initial training of the task 1, the known object detector and the objectness-driven discovery module are trained separately. The training loss for the known object detector ℓ_{known} is the same as that of D-DETR [16] and can be represented as:

$$\ell_{\text{known}} = \ell_{\text{known}}^{cls} + \ell_{\text{known}}^{reg} \quad (8)$$

where $\ell_{\text{known}}^{reg}$ represents the loss for bounding box regression of the known object detector. The objectness-driven discovery module is trained using labels that have undergone the objectness-driven discovery training, and its loss $\tilde{\ell}_{\text{universal}}$ can be represented as:

$$\tilde{\ell}_{\text{universal}} = \tilde{\ell}_{\text{universal}}^{cls} + \ell_{\text{universal}}^{reg} \quad (9)$$

where $\ell_{\text{universal}}^{reg}$ represents the loss for bounding box regression calculated by the universal label of the objectness-driven discovery module.

In this way, our objectness-driven discovery module learns to detect generalized objects regardless of their class. The loss curve during the initial training of the task 1 of the known object detector and the objectness-driven discovery module is shown in Figure 3a.

During the Incremental Learning stage, for training efficiency, only the known object detector is actively trained, and the parameters of the objectness-driven discovery module are frozen. The training process of the known object detector for incremental learning is similar to the task 1, and it employs the data replay method used in OW-DETR [17] to overcome catastrophic forgetting. This strategy allows the model to retain knowledge of previously learned categories while adapting to new data and categories introduced during the incremental learning phase. The loss curve during incremental learning from task 2 to task 4 of the known object detector is shown in Figure 3b.

After training the two detectors separately, we use the collaborative open classifier to perform post-processing to obtain the final predicted output of the model. The collaborative open-world classifier does not partake in training but rather solely contributes to the inference process of the model.

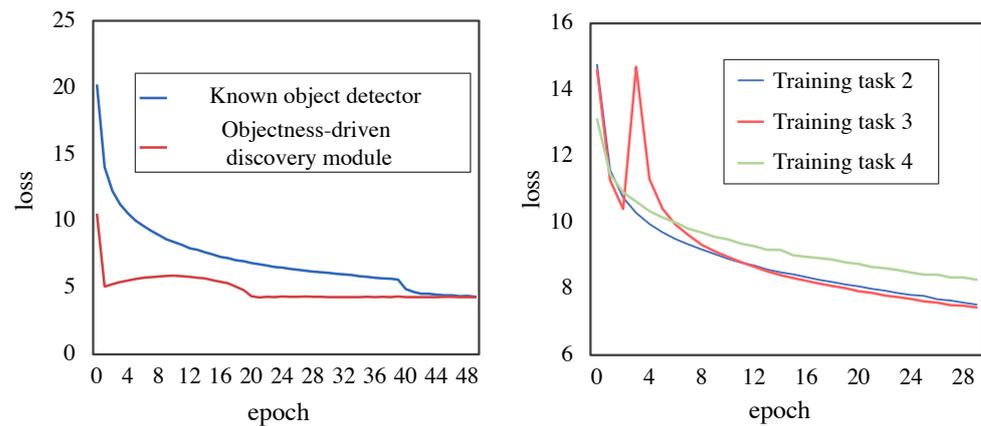


Figure 3. The loss curve during the training. (a) The loss curve of the known object detector and the objectness-driven discovery module during the initial training stage of the task 1. (b) The loss curve of the known object detector during the incremental learning stage of the task 1, task 2 and task 4.

4. Results

In this section, we perform comprehensive experiments and detailed analyses to demonstrate the effectiveness of the proposed method for open-world object detection.

4.1. Experiment Settings

Datasets. According to the typical setup of OWO, all classes from the training set are grouped in T incremental tasks. Following [14], we set T as 4, and adopt the Pascal-VOC [55] and MS-COCO [56] dataset. When learning the task 1, we treat classes and data from Pascal-VOC as the training set, and the remaining 60 classes of MS-COCO are considered unknown. For subsequent tasks, the class division strategy is exactly the same as that in ORE. For evaluation, we use the Pascal-VOC test set and MS-COCO validation set.

Implementation details. We implement our method based on closed world detection model D-DETR [16]. During the training of the known object detector and the objectness-driven discovery module, the SGD optimizer is used and the batch size is set to 8. The IOU threshold P_t in the collaborative open-world classifier is set to 0.4. During both the initial training and the incremental learning stage, the initial learning rate of the known object detector and the objectness-driven discovery module is set to 2×10^{-4} . We train the known object detector for 50 epochs in the initial training stage of the task 1, and train 50 epochs to learn new known classes during the incremental learning. Additionally, we trained for 50 epochs in each stage for data replay to address catastrophic forgetting, following a strategy similar to OW-DETR [17]. About the objectness-driven discovery module, we train it for 80 epochs only in the initial training stage.

Evaluation metrics. To conduct a fair and comprehensive evaluation of the proposed method, we first employ widely recognized metrics in object detection, such as mean Average Precision (mAP) and Recall, as evaluation benchmarks for known and unknown classes. Additionally, we utilize other common-used open-set evaluation criteria, such as Wilderness Impact (WI) [40], Unknown Detection Recall (UDR) and Unknown Detection Precision (UDP) [15].

4.2. State-of-the-Art Comparison

Table 1 shows the comparison of our method respectively with the state-of-the-art methods according to traditional detection metrics, such as mAP and Recall. Note that after completing incremental learning of the task 4, all our classes have become known, and there are no longer any unknown targets in the test set. As a result, metrics related to unknown targets, such as U-mAP, UDR, UDP WI, etc., are not included. From Table 1, our method outperforms other state-of-the-art OWO methods in most cases. Specifically, the unknown mAP of our method (4.84) is almost 7 times that for the model ORE (0.71)

with the better known mAP performance, while the unknown class recall of our method reaches 17.40 and is higher than OW-DETR by 9.75, which is the best method in the existing works. These phenomena clearly demonstrate the effectiveness of our parallel decoupling OWOD framework. As novel known classes are added in the subsequent tasks, the recognition ability of unknown classes decreases, which is consistent with the performance of ORE. However, our method still outperforms the comparison methods by a large margin in U-mAP and unknown Recall metrics. And our known mAP performance on the newly-annotated and previous known classes is maintained. That is to say, our model is capable of promoting the collaborative learning of both known and unknown classes with fewer uncertainties.

Table 1. State-of-the-art comparison for OWOD according to traditional detection metrics. “K-” indicates the known classes, and “U-” represents the unknown classes. Our model achieves superior performance in terms of traditional evaluation metrics in most cases. Note that U-mAP and U-Recall are not calculated because all classes are known in the task 4. The bold represents the best performance on the metric.

Task IDs (→)	Task 1			Task 2			Task 3			Task 4
	K-mAP (↑)	U-mAP (↑)	U-Recall (↑)	K-mAP (↑)	U-mAP (↑)	U-Recall (↑)	K-mAP (↑)	U-mAP (↑)	U-Recall (↑)	K-mAP (↑)
Faster RCNN [57]	56.94	0	0	41.56	0	0	32.41	0	0	27.03
ORE [14]	56.49	0.71	5.72	39.64	0.14	2.66	20.17	0.12	3.34	25.95
SA [52]	55.56	0.20	1.93	39.02	0.03	0.79	31.54	0.003	0.12	26.42
D-DETR [16]	59.75	0	0	46.08	0	0	38.28	0	0	30.60
OW-DETR [17]	58.78	0.07	7.65	44.11	0.04	5.83	35.96	0.03	5.97	27.94
Ours	59.84	4.84	17.40	46.38	1.60	13.53	38.37	1.45	14.30	32.63

Table 2 further shows the comparison of our method respectively with the state-of-the-art methods according to common-used open-set evaluation metrics, such as WI, UDR, and UDP. From Table 2, it is evident that our method achieves similar WI scores to state-of-the-art methods while obtaining higher UDR and UDP scores. Specifically, our framework approaches or outperforms existing methods comprehensively in UDR at every stage, indicating that our framework helps discover more unknown objects. Moreover, our framework exhibits a significant improvement in UDP, suggesting that our framework has a more accurate classification capability for the discovered unknown objects.

Table 2. State-of-the-art comparison for OWOD according to open-set metrics. Our model achieves superior performance in terms of traditional evaluation metrics in most cases. Note that these metrics are only computed in the first three tasks because all classes are known in the task 4. The bold represents the best performance on the metric.

Task IDs (→)	Task 1			Task 2			Task 3		
	WI-0.8 (↓)	UDR (↑)	UDP (↑)	WI-0.8 (↓)	UDR (↑)	UDP (↑)	WI-0.8 (↓)	UDR (↑)	UDP (↑)
Faster RCNN [57]	0.0645	17.58	0	0.0273	16.32	0	0.0164	24.69	0
ORE [14]	0.0528	18.58	31.28	0.0315	17.30	15.37	0.0209	23.67	14.95
SA [52]	0.0563	8.51	22.73	0.0181	5.74	13.83	0.0136	9.12	1.30
D-DETR [16]	0.0600	20.74	0	0.0245	14.41	0	0.0187	34.48	0
OW-DETR [17]	0.0599	18.31	41.77	0.0319	16.24	35.88	0.0220	21.53	27.72
Ours	0.0553	20.00	86.97	0.0251	19.35	69.89	0.0179	22.59	63.30

Overall, our method has significantly improved its performance in unknown class detection, indicating that the objectness-driven discovery module we added has successfully suppressed the unknown discovery uncertainty. At the same time, the performance in known class detection has been more effectively maintained, indicating that our parallel decoupled architecture effectively suppresses the known discrimination uncertainty.

4.3. Ablation Study

As shown in Table 3, we investigate different components in our proposed framework. The table respectively lists the performance of our framework without the known object detector, our model without the objectness-driven discovery module, and the full model. The findings from the table indicate that the framework without the known object detector possesses only the capability to detect unknown targets, lacking the ability to discriminate known classes. The framework without the objectness-driven discovery module fails to detect unknown objects. In contrast, our framework possesses the combined abilities of known class discrimination and unknown class detection, achieving state-of-the-art performance. In conclusion, our full model achieves the best performance, and each module contributes to the proposed model.

Table 3. Ablation study of our model. “w/o KO detector” indicates our model without the known object detector. “w/o CTG detector” indicates our model without the objectness-driven discovery module. Our full model achieves the best performance, and each module contributes to the proposed model.

	WI-0.8	K-mAP	U-mAP	U-Recall
w/o KO detector	0	0	5.74	17.79
w/o CTG detector	0.0553	59.84	0	0
Ours	0.0553	59.84	4.84	17.40

4.4. Parameter Analysis

As shown in Figure 4, we delve into a comprehensive analysis of our model’s behavior under varying IOU threshold P_t in the collaborative open-world classifier. The performance of our framework in terms of unknown class mAP and Recall for different P_t values is shown in Figure 4a, while UDP and UDR metrics are illustrated in Figure 4b. From the figures, it can be observed that when P_t is set to 0 or 0.2, the framework exhibits lower UDR, UDP, and U-Recall values. As the value of P_t increases, U-Recall and UDR increase accordingly. This phenomenon can be attributed to the larger number of retained unknown-class candidate boxes with high Intersection over Union (IOU) scores. However, it also implies a higher occurrence of known-class detections being misclassified as unknown-class, resulting in fluctuations in UmAP and UDP with the increasing value of P_t . In general, when P_t takes other values, UDR, UDP, and U-Recall exhibit relatively stable performance, indicating that our method can detect many unknown objects with small overlaps with known objects, and these detections are less affected by changes in P_t . When P_t is set to 0.4, the framework achieves its highest U-mAP value of 4.83. Therefore, our method effectively achieves the detection of unknown categories with good stability and fewer uncertainties.

4.5. Visualization

Figure 5 displays a visualization comparison between our model and other state-of-the-art models. It is evident that our model demonstrates the capability to simultaneously detect both known and unknown objects with a higher degree of accuracy in bounding box localization and classification labeling. In addition, ORE exhibits an inclination to over-generate bounding boxes with the goal of localizing more unseen objects. Conversely, SA tends to under-generate bounding boxes to minimize misclassification. While OW-DETR is capable of detecting some unknown classes, its precision in detection is relatively poor, often resulting in misalignment of bounding box positions or misclassifying certain

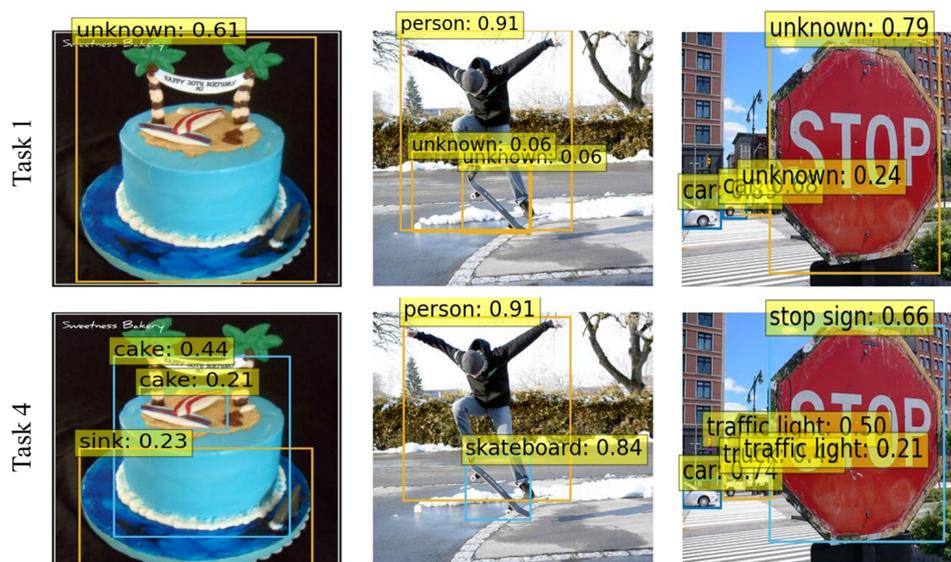


Figure 6. Visualization of our model in the task 1 and the task 4. After the initial training, our model recognizes the cake, the skateboard, and the stop sign as unknown. After the incremental learning, they can be labeled correctly.

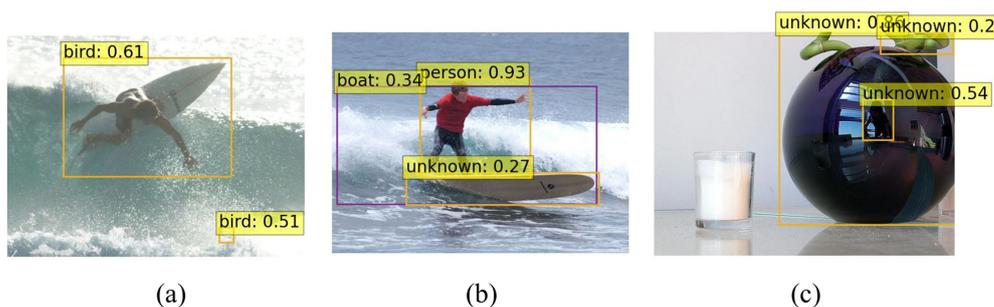


Figure 7. Failure cases of our model. (a) The model mistakenly detected the combination of a person and a surfboard as a bird. (b) Even though people and unknown objects (surfboards) have been detected, their combination is still mistakenly detected as a boat. (c) Part of this unknown object (upper right part) is mistakenly detected as a separate unknown object.

5. Conclusions and Future Work

In conclusion, our parallel OWOD framework presented effectively addresses the challenges of uncertainties in open campus scenarios, enabling the detection of known and unknown objects. The objectness-driven discovery module alleviates the unknown discovery uncertainty by driving the model to learn the objectness derived from the deterministic supervision of known classes. The known object detector and the collaborative open-world classifier mitigate the known discrimination uncertainty and accomplish collaborative detection. This framework has the potential to transform campus monitoring and classroom behavior detection in intelligent educational settings.

It should be pointed out that there are still limitations of our parallel OWOD framework. While our framework has superior performance compared to state-of-the-art methods and notably enhances the detection of unknown classes, it still demonstrates apparent shortcomings in unknown object detection compared to known classes, with one of the obvious shortcomings lying in a noticeable decrease in precision. To address this issue, a substantial improvement in the model’s unsupervised learning capabilities by exploring more effective methods to enhance the model’s ability to detect unknown classes through learning is essential.

In addition, in the context of future prospects, several promising research directions emerge for the refinement and extension of the parallel OWOD framework. Subsequent

investigations may center on augmenting the framework's adaptability across a spectrum of diverse campus scenarios, expanding its capability to detect a broader array of unknown objects, and minimizing the incidence of false positives. Moreover, the exploration of the framework's applicability in domains beyond education, such as healthcare, surveillance, and autonomous systems, represents an intriguing avenue for future exploration. The development of more efficient training strategies and the exploration of advanced neural network architectures may also make valuable contributions to the ongoing advancement of OWOD technology.

Author Contributions: Conceptualization, J.D., Z.Z. and Y.L.; methodology, J.D., Z.Z. and Y.M.; software, J.D. and Z.Z.; validation, J.D., S.H., R.Z. and Y.L.; formal analysis, J.D., S.H. and J.Y.; investigation, S.H., Y.L. and J.Y.; resources, Y.M., J.Y. and B.L.; data curation, Z.Z., S.H. and R.Z.; writing—original draft preparation, J.D., Z.Z. and S.H.; writing—review and editing, Y.M., R.Z. and B.L.; visualization, Z.Z., S.H. and R.Z.; supervision, J.Y., Y.M. and B.L.; project administration, J.D., Y.M. and B.L.; funding acquisition, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62206010, 62022009, and the National Key R&D Program of China (2021ZD0110603).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <http://host.robots.ox.ac.uk/pascal/VOC/> (accessed on 24 October 2023) and <https://cocodataset.org/#home> (accessed on 24 October 2023).

Acknowledgments: We thank the State Key Lab of Software Development Environment and the China Electronics Standardization Institute for providing the experimental environment and equipment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
2. Arcucci, R.; Zhu, J.; Hu, S.; Guo, Y.K. Deep data assimilation: Integrating deep learning with data assimilation. *Appl. Sci.* **2021**, *11*, 1114. [CrossRef]
3. Li, F.; He, F.; Wang, F.; Zhang, D.; Xia, Y.; Li, X. A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning. *Appl. Sci.* **2020**, *10*, 1605. [CrossRef]
4. Shieh, C.S.; Lin, W.W.; Nguyen, T.T.; Chen, C.H.; Horng, M.F.; Miu, D. Detection of unknown ddos attacks with deep learning and gaussian mixture model. *Appl. Sci.* **2021**, *11*, 5213. [CrossRef]
5. Chiu, M.T.; Xu, X.; Wei, Y.; Huang, Z.; Schwing, A.G.; Brunner, R.; Khachatrian, H.; Karapetyan, H.; Dozier, I.; Rose, G.; et al. Agriculture-vision: A large aerial image database for agricultural pattern analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2828–2838.
6. Ni, J.; Chen, Y.; Chen, Y.; Zhu, J.; Ali, D.; Cao, W. A survey on theories and applications for self-driving cars based on deep learning methods. *Appl. Sci.* **2020**, *10*, 2749. [CrossRef]
7. Rehman, A.; Iqbal, M.A.; Xing, H.; Ahmed, I. COVID-19 detection empowered with machine learning and deep learning techniques: A systematic review. *Appl. Sci.* **2021**, *11*, 3414. [CrossRef]
8. Zhou, H.; Jiang, F.; Shen, R. Who are raising their hands? Hand-raiser seeking based on object detection and pose estimation. In Proceedings of the Asian Conference on Machine Learning, PMLR, Beijing, China, 14–16 November 2018; pp. 470–485.
9. Zhao, J.; Zhu, H. CBPH-Net: A Small Object Detector for Behavior Recognition in Classroom Scenarios. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2521112. [CrossRef]
10. Liu, H.; Yu, Y.; Liu, S.; Wang, W. A Military Object Detection Model of UAV Reconnaissance Image and Feature Visualization. *Appl. Sci.* **2022**, *12*, 12236. [CrossRef]
11. Park, Y.; Shin, Y. Applying Object Detection and Embedding Techniques to One-Shot Class-Incremental Multi-Label Image Classification. *Appl. Sci.* **2023**, *13*, 10468. [CrossRef]
12. Miao, B.; Chen, Z.; Liu, H.; Zhang, A. A target re-identification method based on shot boundary object detection for single object tracking. *Appl. Sci.* **2023**, *13*, 6422. [CrossRef]
13. Xu, H.; Zhang, J.; Sun, H.; Qi, M.; Kong, J. Analyzing students' attention by gaze tracking and object detection in classroom teaching. *Data Technol. Appl.* **2023**, *57*, 643–667. [CrossRef]

14. Joseph, K.; Khan, S.; Khan, F.S.; Balasubramanian, V.N. Towards open world object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 5830–5840.
15. Zhao, X.; Liu, X.; Shen, Y.; Ma, Y.; Qiao, Y.; Wang, D. Revisiting open world object detection. *arXiv* **2022**, arXiv:2201.00471.
16. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J.F.D.D. Deformable transformers for end-to-end object detection. In Proceedings of the 9th International Conference on Learning Representations, Virtual Event, Austria, 3–7 May 2021; OpenReview.net.
17. Gupta, A.; Narayan, S.; Joseph, K.; Khan, S.; Khan, F.S.; Shah, M. OW-DETR: Open-world detection transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9235–9244.
18. Chen, C.; Seff, A.; Kornhauser, A.; Xiao, J. Deepdriving: Learning affordance for direct perception in autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2722–2730.
19. Guo, S.; Wang, S.; Yang, Z.; Wang, L.; Zhang, H.; Guo, P.; Gao, Y.; Guo, J. A Review of Deep Learning-Based Visual Multi-Object Tracking Algorithms for Autonomous Driving. *Appl. Sci.* **2022**, *12*, 10741. [[CrossRef](#)]
20. Lee, Y.; Park, S. A deep learning-based perception algorithm using 3D lidar for autonomous driving: Simultaneous segmentation and detection network (ssadnet). *Appl. Sci.* **2020**, *10*, 4486. [[CrossRef](#)]
21. Smistad, E.; Falch, T.L.; Bozorgi, M.; Elster, A.C.; Lindseth, F. Medical image segmentation on GPUs—A comprehensive review. *Med. Image Anal.* **2015**, *20*, 1–18. [[CrossRef](#)] [[PubMed](#)]
22. Nagi, A.T.; Awan, M.J.; Mohammed, M.A.; Mahmoud, A.; Majumdar, A.; Thinnukool, O. Performance analysis for COVID-19 diagnosis using custom and state-of-the-art deep learning models. *Appl. Sci.* **2022**, *12*, 6364. [[CrossRef](#)]
23. Qureshi, S.A.; Raza, S.E.A.; Hussain, L.; Malibari, A.A.; Nour, M.K.; Rehman, A.U.; Al-Wesabi, F.N.; Hilal, A.M. Intelligent ultra-light deep learning model for multi-class brain tumor detection. *Appl. Sci.* **2022**, *12*, 3715. [[CrossRef](#)]
24. Peng, T.; Liang, Y.; Wu, W.; Ren, J.; Pengrui, Z.; Pu, Y. CLGT: A graph transformer for student performance prediction in collaborative learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 15947–15954.
25. Tsiakmaki, M.; Kostopoulos, G.; Kotsiantis, S.; Ragos, O. Transfer learning from deep neural networks for predicting student performance. *Appl. Sci.* **2020**, *10*, 2145. [[CrossRef](#)]
26. Liang, Y.; Peng, T.; Pu, Y.; Wu, W. HELP-DKT: An interpretable cognitive model of how students learn programming based on deep knowledge tracing. *Sci. Rep.* **2022**, *12*, 4012. [[CrossRef](#)]
27. Si, J.; Lin, J.; Jiang, F.; Shen, R. Hand-raising gesture detection in real classrooms using improved R-FCN. *Neurocomputing* **2019**, *359*, 69–76. [[CrossRef](#)]
28. Ma, Y.; Liu, X.; Bai, S.; Wang, L.; Liu, A.; Tao, D.; Hancock, E.R. Regionwise generative adversarial image inpainting for large missing areas. *IEEE Trans. Cybern.* **2022**, *53*, 5226–5239. [[CrossRef](#)]
29. Hu, S.; Ma, Y.; Liu, X.; Wei, Y.; Bai, S. Stratified rule-aware network for abstract visual reasoning. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 1567–1574.
30. Li, F.; Wechsler, H. Open set face recognition using transduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1686–1697. [[PubMed](#)]
31. Heflin, B.; Scheirer, W.; Boulton, T.E. Detecting and classifying scars, marks, and tattoos found in the wild. In Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 23–27 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 31–38.
32. Pritsos, D.A.; Stamatos, E. Open-set classification for automated genre identification. In Proceedings of the European Conference on Information Retrieval, Moscow, Russia, 24–27 March 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 207–217.
33. Scherrek, M.D.; Rigling, B.D. Open set recognition for automatic target classification with rejection. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 632–642. [[CrossRef](#)]
34. Fei, G.; Liu, B. Breaking the closed world assumption in text classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 506–514.
35. Vareto, R.; Silva, S.; Costa, F.; Schwartz, W.R. Towards open-set face recognition using hashing functions. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 634–641.
36. Liu, Z.; Miao, Z.; Zhan, X.; Wang, J.; Gong, B.; Yu, S.X. Large-scale long-tailed recognition in an open world. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2537–2546.
37. Bendale, A.; Boulton, T.E. Towards open set deep networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1563–1572.
38. Perera, P.; Morariu, V.I.; Jain, R.; Manjunatha, V.; Wigington, C.; Ordonez, V.; Patel, V.M. Generative-discriminative feature representations for open-set recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11814–11823.
39. Yoshihashi, R.; Shao, W.; Kawakami, R.; You, S.; Iida, M.; Naemura, T. Classification-reconstruction learning for open-set recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4016–4025.
40. Dhamija, A.; Gunther, M.; Ventura, J.; Boulton, T. The overlooked elephant of object detection: Open set. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1021–1030.

41. Miller, D.; Nicholson, L.; Dayoub, F.; Sünderhauf, N. Dropout sampling for robust object detection in open-set conditions. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 3243–3249.
42. Miller, D.; Dayoub, F.; Milford, M.; Sünderhauf, N. Evaluating merging strategies for sampling-based uncertainty techniques in object detection. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2348–2354.
43. Hall, D.; Dayoub, F.; Skinner, J.; Zhang, H.; Miller, D.; Corke, P.; Carneiro, G.; Angelova, A.; Sünderhauf, N. Probabilistic object detection: Definition and evaluation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1031–1040.
44. Han, J.; Ren, Y.; Ding, J.; Pan, X.; Yan, K.; Xia, G.S. Expanding Low-Density Latent Regions for Open-Set Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9591–9600.
45. Bendale, A.; Boulton, T. Towards open world recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1893–1902.
46. Xu, H.; Liu, B.; Shu, L.; Yu, P. Open-world learning and application to product classification. In Proceedings of the The World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 3413–3419.
47. Willes, J.; Harrison, J.; Harakeh, A.; Finn, C.; Pavone, M.; Waslander, S. Bayesian embeddings for few-shot open world recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, 1–16. [[CrossRef](#)] [[PubMed](#)]
48. Mancini, M.; Naeem, M.F.; Xian, Y.; Akata, Z. Open world compositional zero-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 5222–5230.
49. Zhang, S.; Li, Z.; Yan, S.; He, X.; Sun, J. Distribution alignment: A unified framework for long-tail visual recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2361–2370.
50. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2020**, *53*, 1–34. [[CrossRef](#)]
51. Xian, Y.; Schiele, B.; Akata, Z. Zero-shot learning—the good, the bad and the ugly. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4582–4591.
52. Yang, S.; Sun, P.; Jiang, Y.; Xia, X.; Zhang, R.; Yuan, Z.; Wang, C.; Luo, P.; Xu, M. Objects in semantic topology. *arXiv* **2021**, arXiv:2110.02687.
53. Wu, Z.; Lu, Y.; Chen, X.; Wu, Z.; Kang, L.; Yu, J. UC-OWOD: Unknown-Classified Open World Object Detection. *arXiv* **2022**, arXiv:2207.11455.
54. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5999–6009.
55. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
56. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
57. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.