Addressing Open-set Object Detection for Autonomous Driving perception: A focus on road objects

Corentin Bunel, Maxime Guériau, Alaa Daoud, Samia Ainouz-Zemouche, Gilles Gasso

Addressing Open-set Object Detection for Autonomous Driving perception: A focus on road objects

Corentin Bunel[1], Maxime Gueriau[1], Alaa Daoud[1], Samia Ainouz[1] and Gilles Gasso[1]

Abstract— **Autonomous Vehicles (AVs) are expected to take safe and efficient decisions. Hence, AVs need to be robust to real-world situations and especially to cope with open world setting i.e. the ability to handle novelties such as unseen objects. Classical object detection models are trained to recognize a predefined set of classes but struggle to generalize well to novel classes at inference stage. Open-Set Object Detection (OSOD) aims to address the challenge of correctly detecting objects from unknown classes. However, autonomous driving systems possess specific open-set characteristics that are not yet covered by OSOD methods. Indeed, a detection error could lead to catastrophic events, emphasizing the importance of prioritizing the quality of box detection over quantity. Also, the specific characteristics of objects encountered in road scenes could be leveraged to improve their detection in the open-world setting. In this vein, we introduce a new definition of objects of interest for autonomous driving perception, enabling the proposition of an AV specialized open-set object detector coined ADOS. The proposed model uses a new score, learnt with the background ground truth of the semantic segmentation. This On Road Object score measures whether the object is on drivable areas, enhancing the selection of unknown detection. Experimental evaluations are conducted on simulated and real world datasets and reveal that our method outperforms the baseline approaches in unknown object detection settings with the same detection performance on known objects as the closed-set object detector.**

I. INTRODUCTION

Autonomous vehicle (AV) systems mainly operate two tasks: perception and control [1]. The perception system elaborates a representation of the real world, called World Model [2], on which the control system hinges to make decisions. The effectiveness of the control system is heavily correlated to the relevance of information provided by the World Model. For instance, failure to detect some objects on the road may lead to detrimental consequences while conversely, over detecting objects can impede the user driving experience. Those shifts between reality and AV representation are deemed Corner Cases [3]. Specifically, the Unknown Objects [3], [4] represent object-level corner cases where the AV has to deal with a class of object that has not been encountered during training.

Currently, the best World Model algorithms are Deep Learning based. They learn known classes representation using large scale labeled datasets. Object detection is a major building block of the AV perception system that aims at providing object localization and semantics. However, the current state-of-the-art object detectors do not perform well on corner cases, as evidenced in CODA [5]. Because they are trained on a fixed set of classes, when exposed to open set settings with class instances outside the known set, those detectors fail to generalize well. Open-Set Recognition has emerged to address this problem [6], followed by open-world object detection [4], [7] that incorporates incremental learning into the detection of the unknown objects.

Despite this, open-set object detection for AVs poses specific challenges not yet fully addressed by the current trend of works. Indeed, on one hand perception systems are expected to yield high detection performance on known objects to ensure safe driving actions. For this, one can rely on different sensor modalities to trade off recall and precision contrary to the classical open-set settings. On the other hand, object misdetection, especially false detection may represent a significant issue for the driving decision-making process, making the precision of an object detector more important than recall. Moreover, the types of objects encountered and the diversity of backgrounds and environments in road scenes calls for dedicated open-set settings. The environment can range from a rainy night in a big city with roadworks to a snowy mountain at sunrise with bears and fallen trees [8].

In this paper, we propose a new open-set object detection approach for AVs. The following are the key contributions:

- A definition is introduced for the objects of interest to specify the characteristics of these objects the AV perception has to detect.
- Based on that, we propose ADOS, a new open-set object detector specialized for the autonomous vehicles perception task.
- A new score representing the objects on the road is presented. It is computed using semantic segmentation ground truth and allows to improve the detection of unknown objects.
- A benchmark with existing open-set object detectors that follows an open-set evaluation protocol for road scenes object detection using corner-case datasets illustrates the effectiveness of ADOS.

II. OVERVIEW OF OPEN-SET METHODS

Herein, we review the main open-set object detection methods and their application in AV context.

A trend of research uses the assumption that the unknown objects are close enough to known objects to apply unsupervised learning techniques like pseudo-labeling on them. Joseph *et al.* [4] introduced Open World Object

*This work is funded by the French National Research Agency as part of the MultiTrans project under reference ANR-21-CE23-0032.

[1]INSA Rouen Normandie, Univ Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, F-76000 Rouen, France corentin.bunel@insa-rouen.fr

Detector (ORE), which considers the best proposals from the Faster R-CNN regional proposal networks not overlapping with any ground truth as pseudo-labels. These are used as unknown objects pseudo-labels during training. Similarly, OW-DETR [9] uses attention-driven pseudo-labeling while Zhao *et al.* [10] rely on a Proposal Advisor with a classic non-parametric Selective Search to improve unknown pseudo-labels. These unknown proposals can then be used to lift the classification module to unknown classes. ORE uses contrastive learning with an energy-based classifier, OW-DETR considers a supplementary for novelty, and Zhao *et al.* [10] use a Class-specific Expelling Classifier.

Another way to detect unknown objects is to learn to detect them only using the information issued from the known ground truth. OpenDet [11] uses contrastive clustering to expand low-density latent space and employs an Unknown Probability Learner to extract unknown instances in that space. A standard approach is to use class-agnostic scores representing the quality of the proposal localization instead of binary classification. Many methods [12], [13], [14], [15] use IoU [12] or Centerness [16] as an objectness score. These objectness scores serves to filter the box proposals before classification. Using these scores makes proposals amenable to unknown objects instead of only a closed set of classes.

In order to classify object proposals into known, unknown, and background classes, OW-RCNN [14] and Unknown Sniffer [15] rely on two classifiers: one specialized in known classes and another dedicated to detecting unknown ones using class-agnostic scores. Instead of using a new classifier, PROB [17] classifies proposals into known classes and background as in closed settings. Then, it separates the unknown objects from the background classes using the objectness score.

When it comes to detection on road scenes, ORDER [18] proposed an open world object detector for road scenes where objects have high intra-class scale variation. The approach is based on ORE and considers a Feature-Mix module to improve unknown object identification. Using curriculum learning and a modified focal regression loss, ORDER improves the model capacity to detect small boxes and intra-class scale variation.

Finally, from an evaluation perspective, CODA [5] was proposed as a dataset of corner cases, including unknown objects on the road. Empirical evaluations of ORE on CODA show that a generic open world object detector is not sufficiently efficient to detect corner case objects on the road as ORE is based on the assumption of close similarity between known and unknown objects which might not hold true in road scene contexts. In this paper we leverage on the ground truth labels from the background in order to improve the detection of the unknown classes instances without probing any ground truth label on unknown objects. In the sequel we revisit the definition of open-set detection in AV settings and, then formulate our proposed method.

III. Problem definition

AV perception is composed of algorithms that infer high-level information from the raw data from sensors, such as object detection, localization, traffic sign detection, or forecasting. These algorithms all together should create a world model that provides relevant information for the control system to make safe and effective driving decisions [2]. The world model is conceptualized as a map containing all the different agents present around the vehicle (other vehicles, pedestrians, bicycles, etc.), along with their trajectories. It encompasses the space where the vehicle can drive, identifies the obstacles within the drivable area along with their semantics, and includes all information related to traffic rules [19]. However, the perception system is considered non-specifiable [2], implying that a complete concept specification may not be possible because the world model generally cannot cover all relations and properties in such a wide open context [20]. But we still need to detect unknown objects because they impact the decision of the AVs [21], [19]. In order to detect these corner case objects we need a clear distinction between the background and objects we need to detect, we will call them objects of interest.

A. *Objects of interest for AV*

Not every object present in the scene should be detected, as it would entail a tremendous quantity of information that is not necessarily useful to control systems. Furthermore, achieving such a result would be difficult, considering that humans cannot exactly define what constitutes an object [20].

To first approach what an object of interest is, we can base ourselves on the known objects. The main classes of objects occurring in road scenes datasets are Cars, Trucks, Pedestrians, Traffic Lights, Motorcycles, Traffic Signs, Buses, Bicycles, etc. Those are either traffic sights or objects that can provide information about traffic rules or road conditions, or agents defined by any object whose trajectory stems from an internal decision-making process. These two characteristics are actually used to make control decisions. Hence, we postulate that any object with those properties represents an object of interest. Because the control system needs to anticipate the agents' trajectories, it needs information on any present object and on the potential trajectories of the various agents. This property of being an obstacle can be attributed to object located within the drivable area. Drivable area is defined as the navigable space where the vehicle can operate [21]. For example, a ball is not an agent; if it is only placed on a balcony, it is not of interest, but if the ball crosses the road, then it is of interest. Similarly, a construction cone is an object of interest because it provides information about the road and is likely to be on the trajectories of agents. Thereon, we assume that an object of interest presents at least one of these properties: being an agent, an obstacle, or providing traffic rules information.

B. Defining open-set settings for AV

In a closed-set setting, known objects and the background are well defined. Known objects encompass all instances of the set of classes available at training stage, while the background comprises everything else. However, in an open-set setting, the boundaries between unknown and background remain somewhat ambiguous [22], [9]. This ambiguity arises because the definition of an object depends on the purpose for which the object detector is employed. For instance, a car is an object, a wheel is an object, and a nut is also considered an object. However, the question arises: do we want to detect every nut that attaches a wheel to a car? One approach outlined in [22] is to categorize unknown objects into a set of super classes. Yet, the precise definition of what constitutes a super class remains unclear.

In this paper, we define a super class as an object with a set of properties. Therefore, we have one super class as an object of interest that makes a clear separation between the background and the unknown objects. The background is simply defined by anything but an object of interest, and the unknown objects are all the objects of interest that are not an instance of the known classes.

IV. Proposed method

We aim to have an object detector capable of detecting all the known objects of interest with their correct classes. It should also detect all the other objects of interest and classify them as unknown, without proposing boxes that correspond to the background.

We chose Faster R-CNN as the base detector, as the literature [6], [22], [4], [12] showed that it has better performances on open-set tasks. The separation between classification and localization inside its two-stage architecture and the use of an additional class detection as background make it more stable for open-set settings. Firstly, we will describe the basic Faster R-CNN and then explain how we lift Faster R-CNN to our purpose of open-set object detection for autonomous driving.

A. Faster RCNN

Faster R-CNN works in three steps. The first step is a backbone, acting as a feature pyramid network, that takes the raw image and attempts to express it in a latent space. The idea is to extract as many characteristics as possible from the image. The second step is a Regional Proposal Network (RPN) that takes those features and provides proposals of possible bounding boxes representing the objects in the image. RPN uses anchor boxes, which are pre-defined bounding

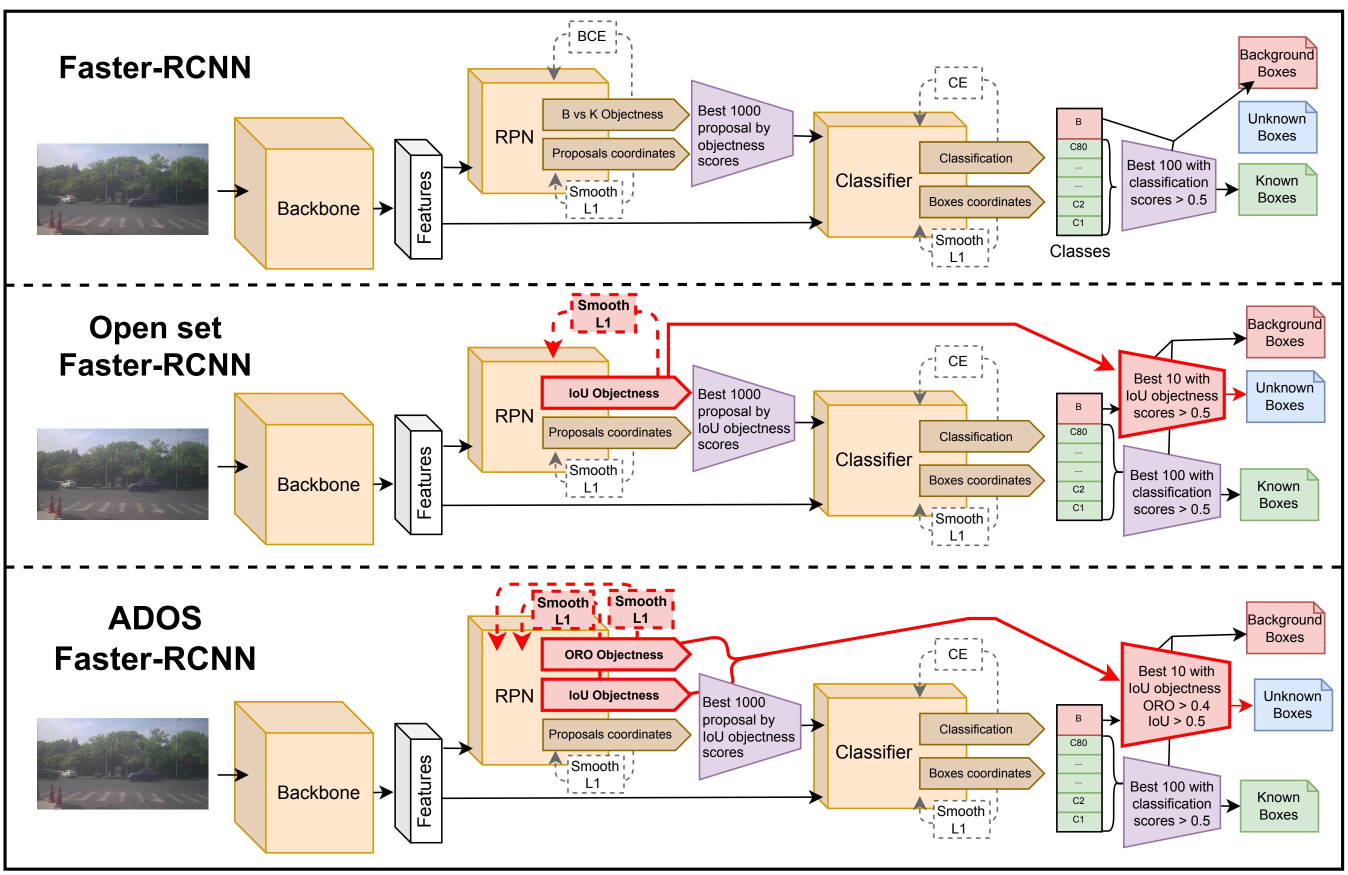

Fig. 1. **Proposed OS Faster-RCNN and ADOS architecture.** The differences between our models and Faster-RCNN are highlighted in red. OS Faster-RCNN changed the objectness head of the RPN by an IoU regression. It then extracts the unknown detection from the background one using this new objectness head. Autonomous Driving Open Set Faster-RCNN (ADOS) keeps the same modification and adds a new head calculating the ORO score, which is used to filter the unknown detection.

boxes placed at various locations and scales across the image. For each of those anchors, the RPN gives an objectness score and adjustments for the anchor box's coordinates. Then, these proposals are filtered by their score and subjected to Non-Maximum Suppression (NMS). The best proposals are then given to the last step, the classification network.

The classifier uses the features from the backbone corresponding to the proposals, thanks to the Regions Of Interest (ROI) pooling layer. The classes of the classifier include the known ones from the dataset, in addition to one class background to exclude every other possible object. It provides, for each proposal and each class, a classification score and regression on the coordinates. The output boxes consist of the top 100 best classification scores with their regressed box coordinates after removing the background classes, the small boxes, and applying NMS.

The Objectness Score from the RPN is learned with Binary Cross Entropy (BCE), where the corresponding anchor is considered a positive sample if it has an Intersection over Union (IoU) with a ground truth of more than 70 percent; otherwise, it is a negative sample. The regression of coordinates is only learned on the positive samples using Smooth L1 loss. The sampling strategy of the RPN involves randomly selecting 256 anchors, with half of them as positive and the other half as negatives.

The Classification Score is learned using Cross Entropy loss. There is a sampling of 512 boxes, with 25 percent being positive and the rest as negatives. Boxes are considered positive if their corresponding anchors have an IoU greater than 70 percent; otherwise, they are considered negative and assigned the background class.

B. Custom Open-set Faster RCNN

In Faster R-CNN, unknown objects are not detected because they are filtered by the RPN and by the classifier. Firstly, they are not proposed by the RPN because BCE loss of the objectness score enforces a strict binary classification between known and everything else. The RPN tends to overfit on the known labeled objects while excluding other possible objects. We make the RPN open to unknown proposals in the same way as in [12], [13], [14], [15]. We replace the Binary Cross-Entropy (BCE) loss with a smooth L1 loss that regresses the Intersection over Union (IoU) [23] with the corresponding known ground truth. Contrary to classification, regressing a score can be transposed to any object making it a suitable choice to enable unknown proposals in the RPN.

Secondly, unknown objects are filtered in the classification step within the background class. The classifier is optimized to separate known classes from the rest of the proposals. Introducing the task of distinguishing unknown classes could impact known detection performance. Given that achieving high performance on known classes is paramount for autonomous vehicles, we keep the classifier unmodified, in contrast to previous work that modifies the classifier to classify as unknown [14], [15]. We follow PROB [17] by using the objectness score to extract unknown objects from those classified as background. The classification score of the background classes is replaced with the objectness score of the RPN. Unknown objects are the objects classified as background with a score above a specified threshold.

In summary, the RPN makes a selection of boxes that look like objects. Then the classifier either discriminates those proposals as into either a known class or a background class. Finally, we separate the propositions from the background classes into unknown objects and background. Our final outputs are separated between known, unknown and background, answering the open-set settings. Open set architecture is summarized in Figure 1, with its differences from Faster-RCNN highlighted in red.

C. Autonomous driving special scores

This custom Faster R-CNN works for an open-set setting where known objects share characteristics with unknown objects. In the case of autonomous driving, the objects of interest, as defined above, could be very different from the known objects. For example, road debris or traffic cones do not have much in common with cars, pedestrians, or traffic lights. To achieve better performance, the model needs more information than that taken from the known classes only. However, we do not have any ground truth on those unknown objects.

Above, we defined an object of interest as any object present on the potential trajectories of the agents. These trajectories occur on the road, sidewalk, and terrain. The ground truth for these drivable areas is present in the semantic segmentation of our datasets, providing pixel information corresponding to the region where objects are of interest. To provide more relevant information to the model about what constitutes an object of interest, we introduce a new score indicating whether an object is on a drivable area or not and is called ORO for On Road Object. Drawing inspiration from shortcut learning [24], we aim to associate this score with the context of where the object is and what is around it.

We have the ground truth of semantic segmentation, where each pixel p has a class C_p. These classes are grouped into three distinct super-classes SC_p: $Background$, $Objects$, and $Drivable$. The $Objects$ super class includes the same object classes as those of the bounding boxes ground truth, in addition to the "static" and "dynamic" classes of the semantic segmentation. The $Drivable$ super class comprises RoadLines, Roads, Sidewalks, Ground, and Terrain. The $Background$ super class includes all the other classes.

The score ORO of a box B is calculated as follows. We defined the area under the box B^{under} as the lower third of B enlarged by λ percent, with B excluded. ORO represents the proportion of pixels classified as $Drivable$ in B^{under}. Making B^{under} proportional to the box size using λ at 30 percent has empirically shown a robust representation of the area under the box. We formally define the ORO score in equation 1.

$$ORO = \frac{\sum_{p \in B_{under}} \delta(I_p \in Drivable)}{\sum_{p \in B_{under}}} \tag{1}$$

Propositions with a high score should represent objects that can be in the trajectory of any agents, and thus are considered objects of interest. This is a class-agnostic score, as it does not depend on the class of the object. We will use it as another objectness score similarly to the RPN IoU score. We set a new threshold on the ORO score that filters once more the unknown from the background, improving the results on unknown objects detection.

D. Implementation details

To adapt Faster-RCNN to open-set settings, we update the objectness head of the RPN. Instead of a BCE loss learning to distinguish objects from background, we use a Smooth L1 loss to regress the IoU score between the anchor boxes and the known ground truth. In this work, we use the inferring IoU score instead of the BCE classification logits for the objectness score. The objectness score is still capped between 0 and 1, ranging from 0 when there is no object overlapping with the anchor box and up to 1 where the anchor box overlaps perfectly with a possible object. Changes only affect the computation of the objectness, not the way it is used in the model. Therefore, the objectness still helps in filtering the proposals from the RPN that are then forwarded to the classifier. The sampling strategy needs to be adujsted to obtain the IoU ground truth for the loss function. Both positive and negative samples are required to limit the model's overfitting toward 0 and 1 values. Positive samples are anchors boxes that have enough overlap with any ground truth, and negative samples can be all the others. However, as all objects are not labelled, making everything else negative will prevent the model from giving high IoU for unknown objects. Approaches from previous work [12], [14] solve this problem by restricting the loss computation to the fraction of positive samples with an IoU above 30 percent. The boxes coordinates regression head is the same as in the original implementation.

For ADOS Faster-RCNN[1] we added another head to the RPN that infers the ORO score as calculated in equation 1. It is identical to the IoU head with a smooth L1 loss. We calculate the ORO score on each positive anchor using the semantic segmentation labels divided into the three super-classes previously defined. We use the same anchors that were sampled for the IoU. ORO is not directly used in the RPN, instead, it helps in extracting unknown boxes after the classification stage from those classified as background. The ORO score is inferred in the RPN, ensuring that the classification task is not impacted. Our assumption is that it may help the RPN to distinguishing objects from the background.

All the proposals boxes are separated using scores into Known, Unknown and Background objects after the classification. We start by separating the proposals classified as known using their classification score. This score is the logits generated by the classifier for each class and each proposal. Known objects are all the boxes classified within the known classes with a classification score over 0.5. The final known detections generated by the model are the best one hundred using their classification scores. There are still some unknown objects within the proposals classified as known. We select the subset of proposals that shows a classification score between 0.2 and 0.5. Proposals with a score below 0.2 are considered as background. After the proposals classified within the known classes are extracted as known object, unknown objects should be distinguished from background. We use the ORO and IoU scores from the RPN proposals to separate them. Proposals with an IoU over 0.5 and an ORO over 0.4 are considered as unknown objects otherwise they are considered as background. IoU signifies the presence of objects, while ORO represents the aspect of being situated on the road. Because we consider that the scene should not contain more than 10 unknown objects, we limit the selection to the best 10 based on the IoU score. Many valid detections with high score were in or near known objects, preventing other unknown objects from being detected by lower scored detections. Therefore, we removed any unknown detections with more than 50% overlap with a known detection.

[1]Our code is available at `https://github.com/leenheart/ADOS_object_detection`

V. Experiments

A. Open challenges regarding evaluation

The evaluation of models in open-set settings is a complex task due to the lack of fully annotated dataset and the complexity of unknown objects definition [22], [25]. Related work selected the COCO [26] dataset that offers a diversity of labelled classes and is considered as a meaningful baseline for object detection benchmarks. To obtain the ground truth for unknown objects, certain known classes are treated as unknown and excluded from the training process. The literature creates multiple tasks, each with unique separation, to mitigate the impact of the choice of the known set considered as unknown.

Dhamija *et al.* [6] introduce a novel error type to evaluate object detectors within open-set settings. This error involves identifying unknown objects as known objects. It occurs because the classifier has learned to filter out anything not present in the known ground truth. However, when the model encounters unknown objects closely related in the feature space to the known classes, it may misclassify them as known instead of unknown. The quantification of this error is conducted using A-OSE [27], which measures the total count of unknown objects classified as known. The objective is thus to minimize this indicator.

In this work, we evaluate the ability of an object detector to detect every object of interest in road scenes. To make a complete evaluation [22], [25], we need to measure the mean average precision (mAP) on known objects but also on unknown objects. Contrary to previous works, comparing models performance on a subset of known classes is not possible considering the importance of detecting known objects in AV perception. Consequently, the usual road scenes

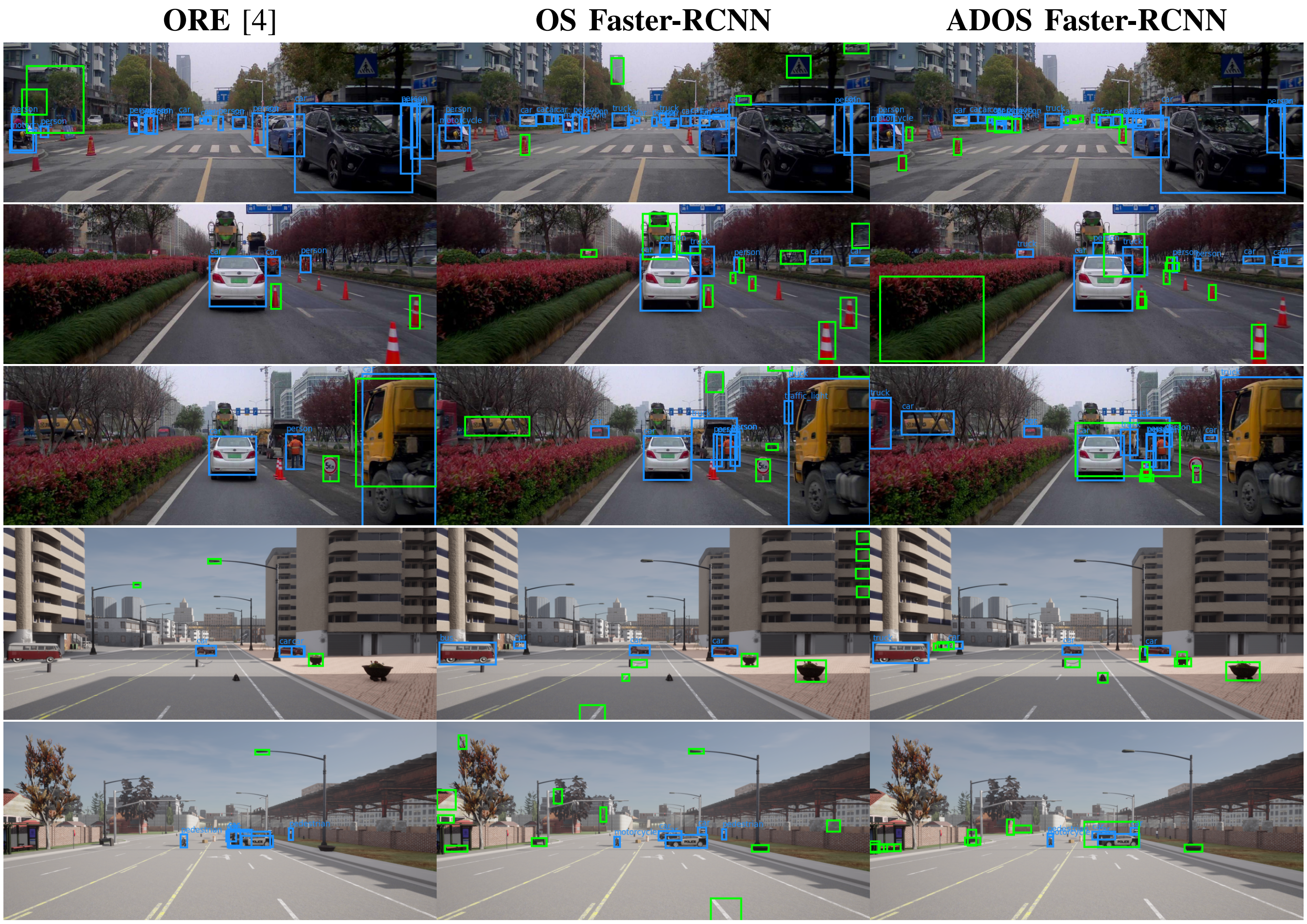

Fig. 2. Qualitative results on example images from CODA dataset (first 3 rows) and CARLA (last 2 rows) dataset. Known boxes are in blue and unknown boxes are in green. Some of the detection obtains by OS Faster-RCNN are outside of the drivable area contrary to ADOS that only have detection on the road thanks to the ORO scores. ORE has also good detection with some outside the road, but is shows worst performance than the two other models. In three models, inadequate detections persist, indicating that despite satisfactory detections, there is room for further improvement.

dataset can not be used in our benchmark as they do not provide the ground truth for unknown classes.

As far as our knowledge can go, the only real world dataset for AV perception containing unknown objects ground truth is CODA [5]. It is an object-level corner case dataset with 34 classes like barrier, debris, dustbin, traffic cone or trailer. However the objects are not exhaustively annotated, preventing the computation of precision metrics of the objects. Instead, CODA is useful to verify if our model is able to detect unknown objects in real road scene scenarios. The recall metric shows the number of annotated unknown objects that are detected by the model. It is a first way to evaluate unknown objects detection performance considering the unavailability of annotated data. However, recall is less important than the precision metric when the object detector is part of an autonomous vehicle perception system. Indeed, an AV can rely on other sensors and associated perception algorithms to complete the world model. If the object detector gives too much inaccurate information, it may not be trusted by the control system, thus limiting the reliability of unknown objects detection. Using a simulator to get an exhaustive ground truth of unknown objects will enable the evaluation of the precision metric. CARLA [28] is an open-source simulator for autonomous driving research. Thanks to the CornerSim framework [29] built upon CARLA, we generate a corner-case dataset of 2000 road scene images. Each image consists of a random number of known and unknown objects disposed randomly on the drivable area. The dataset contains complete labels for known and unknown objects such as construction cones, trash bags, street barriers, etc. Examples of images can be seen in the last two rows of Figure 2. We split this dataset into train, validation, and testing sets with 1300, 300, and 400 images, respectively.

B. Models

Faster-RCNN was selected as the closed set baseline. This two-stage detector is known to perform well in open world setting [6]. We select ORE [4] as the open-set baseline, as it has been identified as a suitable baseline for open-set object detection in the literature [9], [17]. Our comparison is limited to these models due to the difficulty of integrating reproducible models into our benchmarks. These baselines are compared against our two models: (i) Open-Set Faster-RCNN relying on an IoU regression head for the RPN objectness score and (ii) our model that leverages additionally

TABLE I

COMPARISON OF FOUR MODELS ON SIMULATED AND REAL DATASETS WITH OR WITHOUT ENABLING UNKNOWN DETECTION IN THE BACKGROUND CLASS.

	Known					Unknown					
	COCO	CODA	CARLA			CODA *(5000 imgs)*		CARLA *(400 imgs)*			
	mAP	Recall	mAP	Precision	Recall	Recall	A-OSE ↓	mAP	Precision	Recall	A-OSE ↓
Without unknown in the background class											
Faster-RCNN	0.43	0.44	0.29	0.75	0.50	0.01	2663	0.01	0.38	0.04	78
OS Faster-RCNN	0.40	0.40	0.26	0.88	0.35	0.01	2339	0.01	0.43	0.02	30
ADOS	0.24	0.39	0.36	0.89	0.40	0.01	2032	0.01	0.01	0.01	35
With unknown in the background class											
Faster-RCNN	0.43	0.44	0.29	0.75	0.50	0.04	2663	0.00	0.13	0.07	78
ORE	-	0.26	0.06	0.64	0.19	0.03	**431**	0.01	0.10	0.01	195
OS Faster-RCNN	0.40	0.40	0.26	0.88	0.35	**0.10**	2339	**0.02**	0.16	**0.11**	**30**
ADOS	0.24	0.39	0.36	0.89	0.40	0.08	2032	**0.02**	**0.21**	0.10	35

a new objectness score specialized for autonomous driving perception task. ADOS Faster-RCNN is the Open-Set Faster-RCNN with an additional head in combination with the IoU head in the RPN. This head infers the ORO score of each anchors (as illustrated in Figure 1).

All models were trained on the COCO [26] dataset. The Faster-RCNN models use the PyTorch implementation and their pre-trained associated weights, ensuring reproducibility. We selected the COCO dataset as it features an interesting diversity among the represented classes, enabling for more open training setup. Also, both ORE and Faster-RCNN have available weights that were pre-trained on COCO allowing for a fair comparison. Only ADOS Faster-RCNN is fine-tuned on CARLA in order to learn the ORO score. To monitor the impact of our changes from the original implementations on the performance, we evaluate the Faster-RCNN models in two modes: first, when removing detections classified as background, and second, when extracting unknown detections from those classified as background.

C. Results

We compared the proposed methods on both real and simulated datasets with the baseline. Table I displays the performance metrics: mAP, Recall, Precision, and Open Set Error (A-OSE) when calculable. The top panel of the table shows the performance results when the unknown selection process from the background class is deactivated. The only unknown detections obtained are those classified within the classes learned by the model, but where the class is not present in the dataset ground truth. Since models trained on COCO have 80 classes, we keep them in the classifier, enabling the possibility of having detection with classes missing in the dataset known classes. Consequently, they are considered as unknown. We present these results with two complementary objectives. The first one is to demonstrate that the quantitative aspect of these types of unknowns is very low compared to the unknowns from the background class. ADOS exhibits lower performance in this scenario because the fine-tuned RPN on CARLA does not provide detections for classes that are not represented in the CARLA dataset.

These results are also useful for comparing the known detection performances with and without the unknown detections. Since mAP, Recall, and Precision are exactly the same between the first and second panels of the table with the same models, we can conclude that we have achieved the objective of not impacting the known classes detection while opening the model to unknown ones.

The OS Faster-RCNN demonstrates a 10 percent improvement in recall and a 6 percent improvement in precision on the CARLA dataset. Additionally, it shows a 6 percent improvement in recall on the CODA dataset. This highlights that our Open Set Faster-RCNN, trained on the same real-world dataset COCO as ORE, obtains superior performance in Autonomous Driving situations.

A qualitative example in Figure 2 illustrates the ability for the models to detect unknown objects. The OS Faster-RCNN provides more unknown detections than ORE. However a significant number of these unknown detections are outside the scope of the road. We obtain the same recall value for ADOS and OS Faster-RCNN and improve precision by 5 percent. The ORO score allows the model to focus unknown detections on the road, as defined in Section III-A. Figure 2 effectively demonstrates this capability when contrasted with the outputs of ORE and OS Faster-RCNN.

ORE exhibits five times less Open Set Error in the real-world dataset than ADOS but has five times more in the simulated dataset. ADOS decreases the open set error compared to OS Faster-RCNN and Faster-RCNN, indicating an improvement in the model. However, as discussed in the introduction, there is a need for a model that ensures very low open set error.

VI. CONCLUSION

This paper addresses the significant challenge of developing AV perception systems that are robust in open-world scenarios. By introducing a novel definition for the object of interest, we adapt open-set object detection methods to road scene scenarios. Our proposed model incorporates the On Road Object score that aims at improving the model precision when detecting unknown objects of interest. Through the evaluation on both real and simulated corner case datasets, we showcase the performance improvement compared to open-set and closed-set baselines.

However, it is essential to acknowledge that ADOS has not yet reached the performance level required for a system that involves potential safety implications. This indicates the need for further research to address these challenges. We believe that our work can serve as a valuable foundation, and we encourage future work to build upon the insights presented in this paper, recognizing the ongoing efforts required in this field. Future work directions involve exploring new scores that capture the semantic meaning of the scene and developing underlying rules to assist the model in inferring unknown objects without relying on their costly ground truth.

References

[1] Ö. Ş. Taş, F. Kuhnt, J. M. Zöllner, and C. Stiller, "Functional system architectures towards fully automated driving," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, pp. 304–309, June 2016.

[2] M. Hoss, M. Scholtes, and L. Eckstein, "A Review of Testing Object-Based Environment Perception for Safe Automated Driving," *Automotive Innovation*, vol. 5, pp. 223–250, Aug. 2022.

[3] J. Breitenstein, J.-A. Termohlen, D. Lipinski, and T. Fingscheidt, "Systematization of Corner Cases for Visual Perception in Automated Driving," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, (Las Vegas, NV, USA), pp. 1257–1264, IEEE, Oct. 2020.

[4] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, "Towards Open World Object Detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Nashville, TN, USA), pp. 5826–5836, IEEE, June 2021.

[5] K. Li, K. Chen, H. Wang, L. Hong, C. Ye, J. Han, Y. Chen, W. Zhang, C. Xu, D.-Y. Yeung, X. Liang, Z. Li, and H. Xu, "CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving," *European Conference on Computer Vision*, pp. 406–423, Sept. 2022.

[6] A. R. Dhamija, M. Gunther, J. Ventura, and T. E. Boult, "The Overlooked Elephant of Object Detection: Open Set," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (Snowmass Village, CO, USA), pp. 1010–1019, IEEE, Mar. 2020.

[7] A. Bendale and T. Boult, "Towards Open World Recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1893–1902, June 2015.

[8] F. Heidecker, J. Breitenstein, K. Rösch, J. Löhdefink, M. Bieshaar, C. Stiller, T. Fingscheidt, and B. Sick, "An Application-Driven Conceptualization of Corner Cases for Perception in Highly Automated Driving," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, pp. 644–651, July 2021.

[9] A. Gupta, S. Narayan, K. J. Joseph, S. Khan, F. S. Khan, and M. Shah, "OW-DETR: Open-world Detection Transformer," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (New Orleans, LA, USA), pp. 9225–9234, IEEE, June 2022.

[10] X. Zhao, Y. Ma, D. Wang, Y. Shen, Y. Qiao, and X. Liu, "Revisiting Open World Object Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2023.

[11] J. Han, Y. Ren, J. Ding, X. Pan, K. Yan, and G.-S. Xia, "Expanding Low-Density Latent Regions for Open-Set Object Detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (New Orleans, LA, USA), pp. 9581–9590, IEEE, June 2022.

[12] D. Kim, T.-Y. Lin, A. Angelova, I. S. Kweon, and W. Kuo, "Learning Open-World Object Proposals Without Learning to Classify," *IEEE Robotics and Automation Letters*, vol. 7, pp. 5453–5460, Apr. 2022.

[13] S. Konan, K. J. Liang, and L. Yin, "Extending One-Stage Detection with Open-World Proposals," *arXiv preprint arXiv:2201.02302*, Jan. 2022.

[14] D. Pershouse, F. Dayoub, D. Miller, and N. Sünderhauf, "Addressing the Challenges of Open-World Object Detection," *arXiv preprint arXiv:2303.14930*, Mar. 2023.

[15] W. Liang, F. Xue, Y. Liu, G. Zhong, and A. Ming, "Unknown Sniffer for Object Detection: Don't Turn a Blind Eye to Unknown Objects," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Vancouver, BC, Canada), pp. 3230–3239, IEEE, June 2023.

[16] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: A Simple and Strong Anchor-Free Object Detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 1922–1933, Apr. 2022.

[17] O. Zohar, K.-C. Wang, and S. Yeung, "PROB: Probabilistic Objectness for Open World Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11444–11453, 2023.

[18] D. K. Singh, S. N. Rai, K. J. Joseph, R. Saluja, V. N. Balasubramanian, C. Arora, A. Subramanian, and C. V. Jawahar, "ORDER: Open World Object Detection on Road Scenes," *Proc. NeurIPS Workshops*, vol. 1, no. 2, p. 3, 2021.

[19] A. Furda and L. Vlacic, "An object-oriented design of a World Model for autonomous city vehicles," in *2010 IEEE Intelligent Vehicles Symposium*, pp. 1054–1059, June 2010.

[20] B. Spanfelner, D. Richter, S. Ebel, and U. Wilhelm, "Challenges in applying the ISO 26262 for driver assistance systems," *Tagung Fahrerassistenz, Munchen*, p. 16, Aug. 2018.

[21] C. Sanchez, *A World Model Enabling Information Integrity for Autonomous Vehicles*. PhD thesis, Université de Technologie de Compiègne, May 2022.

[22] Y. Hosoya, M. Suganuma, and T. Okatani, "Rectifying Open-set Object Detection: A Taxonomy, Practical Applications, and Proper Evaluation," *arXiv preprint arXiv:2207.09775*, Nov. 2022.

[23] B. Jiang, R. Luo, J. Mao, T. Xiao, and Y. Jiang, "Acquisition of Localization Confidence for Accurate Object Detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 784–799, 2018.

[24] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, "Shortcut Learning in Deep Neural Networks," *Nature Machine Intelligence*, vol. 2, pp. 665–673, Nov. 2020.

[25] Z. Ma, Y. Yang, G. Wang, X. Xu, H. T. Shen, and M. Zhang, "Rethinking Open-World Object Detection in Autonomous Driving Scenarios," in *Proceedings of the 30th ACM International Conference on Multimedia*, (Lisboa Portugal), pp. 1279–1288, ACM, Oct. 2022.

[26] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 740–755, Springer International Publishing, 2014.

[27] D. Miller, L. Nicholson, F. Dayoub, and N. Sünderhauf, "Dropout Sampling for Robust Object Detection in Open-Set Conditions," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3243–3249, May 2018.

[28] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, pp. 1–16, PMLR, Oct. 2017.

[29] A. Daoud, C. Bunel, and M. Guériau, "CornerSim: A Virtualization Framework to Generate Realistic Corner-Case Scenarios for Autonomous Driving Perception Testing," in *The 15th International Conference on Ambient Systems, Networks and Technologies (ANT)*, Apr. 2024.