Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Doctoral Consortium Track

Object Detection in Real Open Environment

Xiaowei Zhao
Beihang University
xiaoweizhao@buaa.edu.cn

Abstract

Although object detection has achieved impressive progress and has a broad impact, it still encounters significant challenges in accurately detecting objects in open-world scenarios. The data in real-world open scenes often exhibit characteristics of limited annotations, such as very few annotated samples or even unknown classes with no annotations. Our studies mainly focus on applying detection to various open scenes and addressing the challenges of sparse samples and unknown classes. The comprehensive research aims to develop more powerful and efficient methods for object detection in the open world, making them more suitable for real-world applications.

1 Background

With the rapid development of artificial intelligence and computer vision, object detection algorithms based on deep learning have become prevalent in various fields, such as robot navigation, intelligent video surveillance, industrial detection, and aerospace due to their practical significance and application value. However, traditional object detection methods rely on closed, high-quality datasets with millions of annotated data. In real-world scenarios, the data often contains large amounts of background noise, a limited number of samples, and unknown classes. As a result, previous methods face significant challenges in real-world open scenarios.

Object detection in real open scenes with limited labelling faces significant challenges. Firstly, there may be a scarcity of samples, making it difficult to accurately describe the target sample distribution in the absence of sufficient labelled data. This can result in poor generalization and ineffective response to rare events. Secondly, there may be a lack of labelling information and unknown categories, which can lead to unsupervised targets being ignored or confused with known targets, causing misjudgments and negatively affecting reliable detection. Furthermore, the model may struggle to balance old and new knowledge, resulting in catastrophic forgetting.

The challenges mentioned above invalidate the independent and identical distribution assumption of big data learning, rendering traditional object detection algorithms ineffective. As a result, existing object detection in real scenes faces a significant impact on application effectiveness and safety guarantees. This underscores the urgent need for researchers to make breakthroughs in models, which can also help bridge the gap between humans and machines and facilitate practical application in the open world.

2 Our Studies

Our research mainly focuses on exploring object detection algorithms in real open-vision scenes, specifically addressing the challenges of sparse samples and unknown class. For the sparse samples, we study the few shot object detection (FSOD), which is an effective approach to solving the sparse sample problem. It leverages the knowledge of base classes which have an abundance of examples to solve new tasks with minimal support images of novel classes for training. And for the unknown classes, we explore Open World Object Detection (OWOD), whose objective is not only to make accurate predictions for fully annotated known classes, but also to detect and label unknown instances as the "unknown" class during the testing phase, and incrementally learn their labels as they are added.

We will introduce our studies for few-shot object detection and open-world object detection, in Section 2.1 and Section 2.2, respectively.

2.1 Few Shot Object Detection

Existing few-shot object detection (FSOD) approaches often struggle to generate sufficient high-quality positive region proposals, which are crucial for detection performance. This is mainly due to the lack of informative knowledge from base classes and non-specific alteration for novel classes. To overcome this limitation, we adopt an approach inspired by natural evolution, where the detection of novel classes is likened to the formation of new species that can better acclimate to a new environment by reorganizing and mutating existing populations.

In our research, we introduce a Temporal Speciation Network (TeSNet) [Zhao *et al.*, 2023] that uses the knowledge of base classes and adapts them to novel classes through evolving training. TeSNet consists of two main components: the Selective Recombination Module (SRM) and the Mutational Region Proposal Network (MRPN). The SRM enables a novel class to identify and learn from similar base classes through recombination, and generates a variety of novel individuals to produce positive proposals with diverse patterns. On the other hand, the MRPN adjusts the parameters trained on base classes based on the unique characteristics of novel samples, enabling accurate localization of positive proposals. The detector achieves satisfactory performance for novel classes by trained on diverse positive proposals .

2.2 Open World Object Detection

The OWOD setting poses new challenges, including (1) "Unknown" Objectness (UO), which involves distinguishing unknown instances from the background, (2) "Unknown" Discrimination (UD), which involves distinguishing unknown instances from known classes.

Auxiliary Proposal Based OWOD

To tackle the challenging distinctions between the unknown classes and the background (UO) or known classes (UD), we propose a simple and effective OWOD framework, consisting of an auxiliary Proposal ADvisor (PAD) module and a Class-specific Expelling Classifier (CEC)[Zhao *et al.*, 2022]. As the conventional RPN is prone to be biased by the supervision of known classes, the non-parametric PAD is introduced to assist RPN in identifying unknown instances more confidently due to the class-agnostic property of the object. And the CEC adjusts the over-confident activation boundary and eliminates confusing predictions using a class-specific expelling function. The class-specific expelling function reallocates the class predictions of a predicted bounding box to prevent the detection model from overconfidently classifying unknown instances into known classes, thus benefiting the UD challenges.

Objectness-Centric OWOD

Previous approaches often overlook the significance of distinguishing between unknown instances and backgrounds, or similar known classes, resulting in a drop in performance.

To address these issues, we propose a two-branch objectness-centric framework for open world detection that fully exploits the generalized objectness of both known and unknown classes, ensuring open world recognition performance [Wu *et al.*, 2022]. Our framework is designed to effectively capture the objectness of both known and unknown instances, localizing unknown instances while accurately recognizing multiple classes. Specifically, our framework comprises a bias-guided detector that is trained using known annotations to make precise detections for known classes and an objectness-centric calibrator that learns to generalize the localization ability of objects, irrespective of their categories, without compromising the performance of known classes. During inference, we employ an objectness-centric confirmation to validate the predictions of known classes and detect unknown instances.

Causal Debiasing Learning for OWOD

For the Unknown Discrimination, previous methods rely heavily on information from known classes to extract unknown features, resulting in inaccurate unknown features biased towards the known ones. The biased feature extraction will further affect the performance of both known and unknown classes: the unknown classes tend to be misclassified as the known ones or even suppressed, and the precision of known classes will be degraded.

To solve the above problems, we first construct a Structural Causal Model (SCM) for the OWOD task based on the causality theory, which shows that the feature of known classes is essentially a confounder that causes spurious known-unknown correlations, and the causality from the images to unknown features as well as unknown features to predictions is weak. Then, we propose a novel causal debiasing framework for open world object detection, consisting of a semantic causal intervention module to remove the confounder via backdoor adjustment with the semantics, and an unknown causality enhancement module to strengthen the causal effect of the unknown objects and their predictions through self-supervised adjustments [Zhao, 2023].

3 Discussion

Our long-term research interest is to investigate much more strategies to address the challenges of limited annotations encountered by object detection in real-world open scenarios. Currently, we are attempting to explore open vocabulary object detection, where the challenge of limited annotations exists similarly without labelling information. However, we can utilize large-scale pre-trained models to align the relationship between vision and language, thereby aiding object detection. This approach helps us to better understand the correspondence between objects in vision and language, thereby improving the accuracy and efficiency of object detection.

Precise object detection methods in open scenarios can greatly expand the deployment potential of advanced neural networks. We believe that these works can provide inhibition lines for real-world AI applications and promote to the design of stronger DNN models.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62206010 and 62022009.

References

[Wu *et al.*, 2022] Yan Wu, Xiaowei Zhao, Yuqing Ma, Duorui Wang, and Xianglong Liu. Two-branch objectness-centric open world detection. In *Proceedings of the 3rd International Workshop on Human-Centric Multimedia Analysis*, pages 35–40, 2022.

[Zhao *et al.*, 2022] Xiaowei Zhao, Xianglong Liu, Yifan Shen, Yuqing Ma, Yixuan Qiao, and Duorui Wang. Revisiting open world object detection. *arXiv preprint arXiv:2201.00471*, 2022.

[Zhao *et al.*, 2023] Xiaowei Zhao, Xianglong Liu, Yuqing Ma, Shihao Bai, Yifan Shen, Zeyu Hao, and Aishan Liu. Temporal speciation network for few-shot object detection. *IEEE Transactions on Multimedia*, 2023.

[Zhao, 2023] Xiaowei Zhao. Open world detection with causal learning. In *Under Review*, 2023.