오픈 월드 객체 감지의 현재 트렌드에 대한 리뷰

이크발 무하마드 알리[O], 김수균[*]
[O]제주대학교 컴퓨터공학과,
[*]제주대학교 컴퓨터공학과
e-mail: kimsk@jejunu.ac.kr[*]

Unveiling the Unseen: A Review on current trends in Open-World Object Detection

MUHAMMAD ALI IQBAL[O], Soo Kyun Kim[*]
[O]Department of Computer Engineering, Jeju National University,
[*]Department of Computer Engineering, Jeju National University

● 요 약 ●

This paper presents a new open-world object detection method emphasizing uncertainty representation in machine learning models. The focus is on adapting to real-world uncertainties, incrementally updating the model's knowledge repository for dynamic scenarios. Applications like autonomous vehicles benefit from improved multi-class classification accuracy. The paper reviews challenges in existing methodologies, stressing the need for universal detectors capable of handling unknown classes. Future directions propose collaboration, integration of language models, to improve the adaptability and applicability of open-world object detection.

키워드: Open-world object detection, Uncertainty representation, Incremental learning

I. Introduction

When incorporating machine learning models into real-world applications, it's crucial to evaluate their efficiency and reliability. Our reliance on these models is growing, but they often face challenges like noise and errors when dealing with real-world data. To enhance AI systems, representing uncertainty is essential, involving aleatoric (data-related) and epistemic (knowledge-related) uncertainties. Aleatoric uncertainty pertains to inherent data characteristics, while epistemic uncertainty arises from insufficient data. In deep learning, uncertainty quantifiers (UQ) are vital for reliable predictions, especially in tasks like object detection. Traditional approaches struggle with open-world scenarios where new or unknown objects constantly appear. The open-world object detection strategy aims to effectively handle these challenges by introducing a novel technique operating within the object detector's latent space. Real-world object detection is intricate due to the vast number of diverse classes. Open-world scenarios make it impractical to predefine all possible classes in the training set. Our strategy addresses this by making the model capable of detecting, labeling, and learning unknown objects incrementally. Achieving accurate and reliable multi-class classification is crucial for applications like autonomous vehicles, surveillance systems, and robotics. Moreover, the model should accurately localize and classify all potential objects in an image or video frame. It must identify unknown objects and adapt to real-world scenarios incrementally. Despite advances in computer vision, ensuring complete reliability remains challenging, and addressing false positive detections is crucial for accurate outcomes in real-world applications. The capability to detect and identify unknown objects reinforces the model's adaptability and reliability.

Fig. 1. Performance of Open World Object Detector in Real World

The open-world objection detection is formulated as follows. An open-world object detector maintains a set of known object classes at a specific time instance $'t'$ i.e. $Ob^t = \{1, 2, \ldots, C\}$. In the dynamic real-world scenario, an additional set of unknown object classes may emerge during the testing phase, which is represented as set $U_k = \{C+1, \ldots.\}$. The classes recognized by the model are assumed to be labeled in the dataset $D^t = \{X, Y\}$,where X and Y denote input images and their corresponding labels, respectively. In the context of the real-world object detection scenario, the label Y_k signifies the bounding boxes, defined as $Y_k = \{x_k, y_k, w_k, h_k\}$ where x_k, y_k represents the coordinates of the bounding box with respect to the origin, and wk, and hk denotes the height and width of the bounding box.

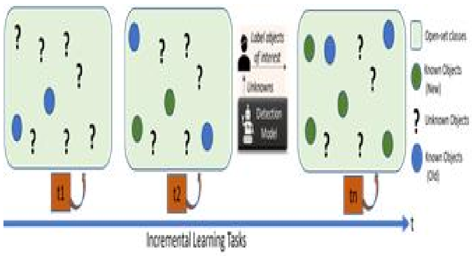

Fig. 2. Overview of Open World Object Detection

In the open-world object detector paradigm, model M is trained to detect all known classes C at a particular time instance $'t'$. The model is expected to recognize any new or unseen class that can appear during inference. The model is expected to accurately detect and classify all the objects appearing during inference; instances not known to the model are detected and labelled as an unknown class, creating the set of set of unknown instances U_t. The detected unknown instances in set U_t are then forwarded to a human annotator, to identify n new classes of interest and assign new labels to them. Active learning strategies can also be employed for the labelling task. Upon labelling the new instances the model's knowledge repository is updated incrementally with the newly identified classes, denoted as $M+n$, using specific incremental learning approaches without extensive retraining on the entire dataset from scratch. Simultaneously, the set of known classes is updated as $Ob_t + \{C+1, \ldots, C+n\}$. This iterative cycle persists throughout the lifetime of the object detector, ensuring a gradual and continuous update of its knowledge repository.

II. Related works

The investigation into open-world object detection encompasses three key domains: open-set classification, open-world classification, and open-set detection. The open-set classification paradigm acknowledges the potential emergence of novel unknown classes during testing, considering the information acquired from the training set as incomplete. In addressing this, [1] introduced a one vs. rest paradigm to minimize the risk of samples from open space, followed by [2], which proposed a hybrid model (PI-SVM) combining probability estimation and discrimination with known negative classes. Further contributions include the Weibull-calibrated SVM (W-SVM) and a Compact Ablating Probability model, which assigns low likelihood values to instances moving toward open space.

In open-set object detection, [3] revealed limitations in deep neural networks, introducing the Open-Max layer to estimate the likelihood of input belonging to an unknown class. Temperature scaling and small perturbations were added, and presented a probabilistic methodology to determine inlier odds using a novel network architecture. Generative OpenMax (G-OpenMax) [3] explicitly models and calculates scores for unknown classes, leveraging generative adversarial networks. Open-world classification introduced a flexible configuration where known and unknown classes coexist, adapting to new labels for unknown objects. Open set detection [4] explored object detectors' responses to open-set settings, revealing challenges in efficiently separating known and unknown classes. Recent approaches include a classification-free Object Localization Network (OLN) [5], proposing a network for calculating objectness scores, and a probabilistic approach for objectness estimation [6]. A two-stage object detector [7] deploys strategies for detecting known and unknowns, and [8] employs convolutional operations for knowledge distillation in unknown recognition. However, these methodologies face challenges in retaining information as new classes are added to the model's knowledge repository, making it difficult to distinguish between known and unknown objects in the feature space. Addressing this, incremental learning in open-world object detection remains an open research problem requiring an optimal solution.

III. Future Direction

Future advancements should prioritize the development of universal object detectors that can leverage images from multiple sources and accommodate heterogeneous label spaces during training. Large-scale collaborative efforts in classification and localization are essential. However, addressing challenges associated with noisy annotations in large vocabulary datasets and inconsistencies in specialized datasets is crucial. Future Solutions should ensure that detectors learn from a wide range of images, covering diverse categories and scenes. To enhance the generalization ability of object detectors, exploring the integration of language models holds promise. Language models can play a pivotal role in labeling and categorizing images without relying solely on visual information. By incorporating textual descriptions or context, detectors may overcome the limitations of purely visual learning. This approach could enable detectors to predict category tags for novel classes not annotated during training, contributing to improved open-world generalization. Efforts should be directed towards enabling detectors to effectively handle novel classes that were not annotated during the training phase. Overcoming the requirement for fully-supervised learning by human annotations is crucial. Future research may explore semi-supervised or unsupervised approaches that leverage contextual information, language cues, or other innovative strategies to predict category tags for unseen classes.

IV. Conclusion

Advancing open-world object detection necessitates the development of universal detectors capable of adapting to diverse environments. The crucial role of large-scale collaboration in training from heterogeneous sources cannot be overstated. The integration of language models emerges as a promising solution to address challenges in visual learning. Future research should prioritize exploring innovative techniques, potentially incorporating semi-supervised approaches, to effectively handle unseen classes. The overarching goal is the evolution of open-world object detection into a versatile and adaptive technology with widespread applicability.

ACKNOWLEDGEMENT

“이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2021R1I1A3058103)“

REFERENCES

[1] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, ‘‘Toward open set recognition,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 7, pp. 1757–1772, Jul. 2013.

[2] L. P. Jain, W. J. Scheirer, and T. E. Boult, ‘‘Multi-class open set recognition using probability of inclusion,’’ in Computer Vision—ECCV 2014. Cham, Switzerland: Springer, 2014, pp. 393–409.

[3] S. D. Z. Ge and R. Garnavi, ‘‘Generative OpenMax for multi-class open set classification,’’ in Proc. Brit. Mach. Vis. Conf., Sep. 2017, pp. 42.1–42.12.

[4] A. R. Dhamija, M. Günther, J. Ventura, and T. E. Boult, ‘‘The overlooked elephant of object detection: Open set,’’ in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2020, pp. 1010–1019

[5] O. Zohar, K.-C. Wang, and S. Yeung, ‘‘PROB: Probabilistic objectness for open world object detection,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 11444–11453.

[6] Z. Wu, Y. Lu, X. Chen, Z. Wu, L. Kang, and J. Yu, ‘‘UC-OWOD: Unknown-classified open world object detection,’’ in Computer Vision—ECCV 2022, S. Avidan, G. Brostow, M. Cissé, G. M. Farinell and T. Hassner, Eds. Cham, Switzerland: Springer, 2022, pp. 193–210.

[7] Y. Ma, H. Li, Z. Zhang, J. Guo, S. Zhang, R. Gong, and X. Liu, ‘‘Annealingbased label-transfer learning for open world object detection,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 11454–11463

[8] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, ‘‘Towards open world object detection,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 582