1st Place Solution for SSLAD Challenge 2022: Corner Case Detection

Jiawei Zhao*, Yiting Duan*, Jinming Su, Wangwang Yang, Tingyi Guo, Junfeng Luo and Xiaolin Wei

Meituan Vision AI Department
{zhaojiawei12,duanyiting,sujinming,yangwangwang,guotingyi, luojunfeng,weixiaolin02}@meituan.com

Abstract. In this report, we introduce the technical details of our solution for ECCV 2022 Workshop SSLAD Track 3 – Corner Case Detection. For obstacle detection in road scenes, it is very challenging to detect novel instances that are not seen or barely seen during training. To address this issue, we propose an efficient pipeline for obstacle detection in road scenes based on large-scale unlabeled data. Specifically, we use large-scale unlabeled data to train a closed-set model and an open-set model respectively in a pseudo-supervised learning manner, and further refine the predictions of novel classes to iterative improve the performance. Finally, we achieve 3.09 on the test set and win the 1st prize.

Keywords: corner case, pseudo-supervised learning, open-set detection

1 Introduction

Effective obstacle detection in the road scene is crucial for reliable autonomous-driving perception systems. In recent years, deep learning has achieved remarkable success in detecting common traffic obstacles (*e.g.*, cars, pedestrians, cyclists, etc.). However, such detectors are often incapable of detecting novel objects that have not been seen or are rarely seen in the training data. These novel objects are called corner cases, which include new instances of the common class (*e.g.*, an overturned truck), and instances of the novel class (e.g., a cone bucket). In this task of corner case detection, the goal is to detect common classes and the rest of the novel classes in the real world.

For obstacle detection of common classes, close-set object detection methods (*e.g.*, YOLO series [1, 2], FCOS [3], Faster RCNN [4], Cascade RCNN [5]) are generally applied. However, close-set methods heavily rely on well-annotated data and perform poorly on large-scale unlabeled collected data from different scenes (*e.g.*, weather, cities, periods). Notably, these methods quickly lose efficiency in the detection of novel classes which are not seen during training. Although some methods of open-set detection have been proposed to detect

* denotes equal contribution.

novel categories that have not been seen before, such as ORE [6], OWDERT [7], OpenDet [8], these methods perform poorly in the real world.

In order to solve these problems, we propose an effective pipeline for obstacle detection. First, we introduce an iterative pseudo-supervised learning method to train the close-set model for known class object detection. Then we train the open-set model (*i.e.*, OpenDet [8]) in the road scene for unknown corner-case detection. At last, some useful strategies are adopted to boost the performance of both common and corner-case object detection. More details of our solution will be published soon on arxiv.com.

2 Approach

We will introduce two independent settings for corner case detection, depending on whether to use annotated corner case labels. In Section 2.1, assuming the novel classes of objects are totally unlabeled, we propose an unsupervised corner case detection method to detect obstacles in road scenes. In Section 2.2, we propose a supervised corner case detection method when there are some labeled data about corner cases of novel classes (the organizer of the competition clarifies that all participants are allowed to use the validation data freely).

2.1 Unsupervised Corner Case Detection

In this section, in order to exploit both labeled common class data and unlabeled novel class data, we propose an effective pipeline that jointly uses a close-set detection model and an open-set detection model to improve detection results for common and novel cases. As shown in Fig. 1, we separately train a close-set detection model for known classes and an open-set detection model for unknown classes, and then combine these two models to train a powerful detection model via a joint training manner.

Close-Set Detection Model. For the close-set detection model of known class objects, we apply cascade RCNN [5] as our baseline architecture and swin-base [9] as our backbone. To perform more robustly on novel instances of common classes, we apply multiple augmentation methods (*e.g.*, AutoAugmentV2 [10], MixUp, Albu) to enrich the samples during training. To exploit the large-scale unlabeled images, we propose an effective pseudo-supervised learning method to iterative train the network on generated pseudo-labels as pseudo-pretrain for the supervised learning stage. More details could be found in our report on Track 1. Specifically, we apply the pseudo-supervised learning method on the 1000w unlabeled images in SODA10M [11] and fine-tune with pseudo-pretrain on annotated SODA [11] and ONCE [12] to detect common classes in the supervised learning stage.

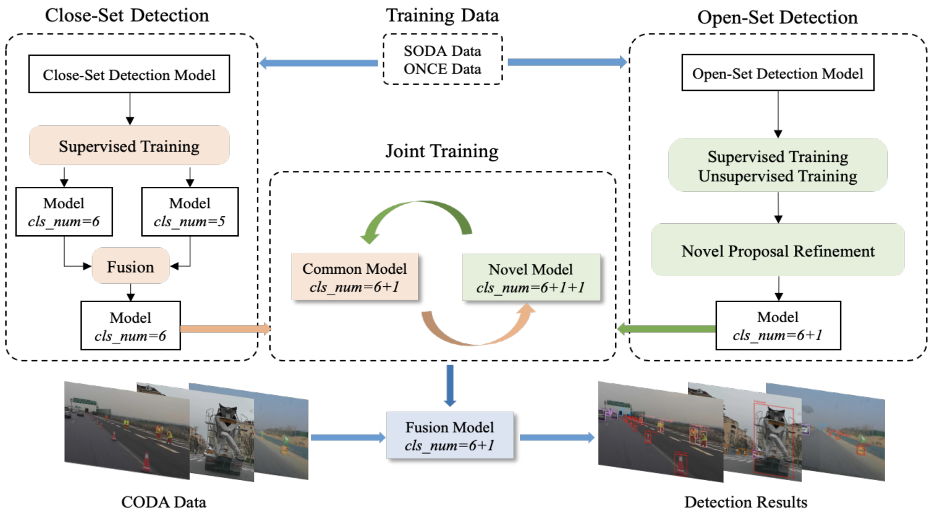

Fig. 1. Framework of our solution for unsupervised corner case detection. We propose an effective pipeline to respectively train the close-set model for known classes and the open-set model for unknown classes, then iterative mine and refine the novel proposals via a joint training manner.

Open-Set Detection Model. For the open-set detection model of unknown class objects, we use the OpenDet Detector [8] as our baseline architecture, a state-of-the-art method for open-set detection by expanding low-density latent regions. OpenDet consists of two well-designed learners, CFL and UPL, where CFL performs instance-level contrastive learning to learn more compact features and UPL learns the unknown probability that serves as a threshold to further separate known and unknown classes. Based on the OpenDet method, We train an open-set model on labeled data of common classes, the backbone is ResNet50 [13] with ImageNet-1K pre-trained weights.

Novel Proposal Refinement. Since the detection performance of our trained OpenDet model is not strong, there are some false positive detections about corner cases. In order to remove unreliable corner-case predictions, we utilize point-cloud annotations in ONCE [12] dataset to conduct several strategies for proposal refinement, which could filter novel proposals and obtain high-quality proposals of novel classes.

With refined novel proposals, we retrain the common classes and novel proposals mined from open-set detection model in a joint training manner. Finally, a powerful corner case detector could be obtained. Due to time constraints, this section only gives a rough introduction to the proposed refinement method in unsupervised corner case detection. For further research on corner case detection, we will release a technical report with more details soon.

2.2 Supervised Corner Case Detection

As the organizer of the competition clarify that all participants are allowed to use the validation data freely, we follow other teams and apply the close-set detection method mentioned above on the CODA [14] validation set.

Firstly, we load the best close-set model in Section 2.1 for initialization and retrain the network on the CODA validation set. Then we obtain the reliable pseudo-labels on unlabeled SODA [11] images and retain the images having novel-class predictions for pseudo-supervised learning. Iterative pseudo-supervised learning on more reliable pseudo-labels and more unlabeled images could obtain a more robust representation for novel instances. At last, we train the CODA [14] validation set with obtained pseudo-pretrained model and achieve the best performance on both common classes and novel classes.

3 Experiments

3.1 Datasets

In the ECCV2022 SSLAD Challenge Track3, two large-scale real-world autonomous-driving datasets, SODA10M [11] and ONCE [12] are provided as the official dataset. SODA10M is a 2D object detection dataset, which contains 10 million unlabeled images and 10k fully-annotated images with 6 representative categories (pedestrian, cyclist, car, truck, tram, tricycle). ONCE is a 2D and 3D object detection dataset that contains 1 million LiDAR frames, 7 million camera images, and 15k fully-annotated scenes with 5 categories (car, bus, truck, pedestrian, cyclist). Besides, CODA2022 [14] is provided for evaluation, which is collected from SODA10M and ONCE . SODA10M and the validation set of CODA2022 are used to train our detectors. ONCE has also been used in some of our attempts.

3.2 Implementation Details

In unsupervised corner case detection, we implement the close-set model and open-set model respectively. For close-set detection, we adopt multi-scale training and AdamW optimizer with an initial learning rate of 1e-4, the training epoch is set to 36 with the learning rate decayed by a factor of 10:1 at epochs 27 and 33, and the image size ranges from (648, 1920) to (1080, 1920). For open-set detection, we adopt SGD optimizer with an initial learning rate of 0.001, a momentum of 0.9, weight decay of 0.0001, and train 60k iterations with the learning rate decayed by a factor of 10:1 at 30k and 50k.

In supervised corner case detection, we re-implement the close-set model with an initial learning rate of 1e-5 and train 73 epochs. We adopt multi-scale augmentation and soft-NMS [15] to generate submissions. All experiments are conducted on 16 NVIDIA A100 GPUs. Our implementation is based on the open-source object detection toolbox MMDetection [16].

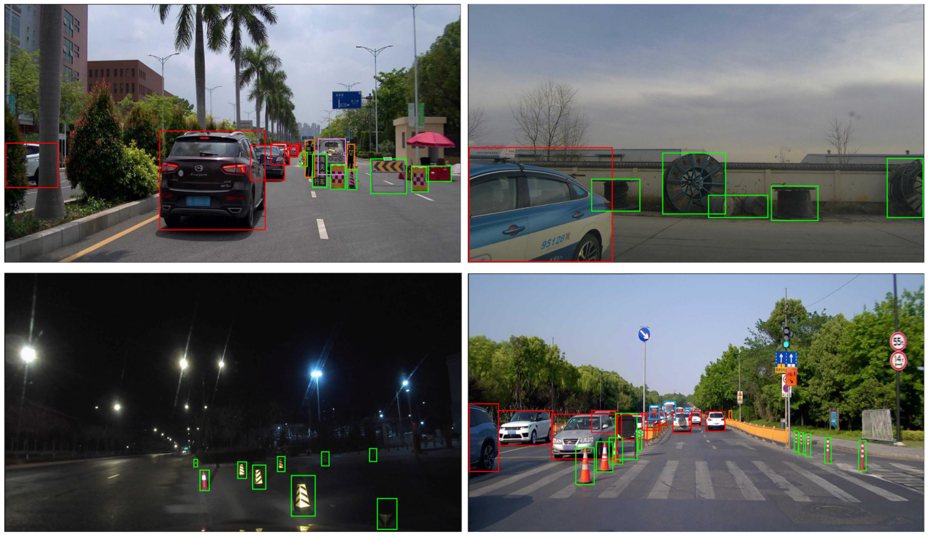

Fig. 2. Representative visual examples from the proposed solution. Noted that the novel objects are drawn in green, while the common objects are drawn in other colors.

3.3 Performance Analysis

The final submissions on the test leaderboard are shown in Tab. 1, our approach achieved the best performance on all metrics and won the first prize. Since the methods in Section 2.1 and Section 2.2 are different, we conduct ablation studies on unsupervised and supervised settings respectively.

Table 1. Final performance on the test leaderboard.

User	Team	Sum	AR-agnostic-corner	AR-agnostic	AP-agnostic	AP-Common
gavin	**MTCV**	**3.09**	**0.80**	**0.85**	**0.78**	**0.66**
IPIU-XDU	IPIU-XDU	3.06	0.79	0.85	0.77	0.64
haoooooooo	edl	2.83	0.76	0.81	0.70	0.55

Unsupervised Corner Case Detection. The detailed ablation studies in unsupervised corner case detection are shown in Tab. 2. We first train a strong baseline detector with ImageNet-1K pre-trained model on the SODA train set. Adopting multiple augmentation methods could effectively improve the score from 1.64 to 1.77. Moreover, our proposed pseudo-supervised learning method could boost the performance from 1.77 to 1.93 only with 10w unlabeled images. We further train with a 1000w pseudo-pretrained model on the SODA train set and validation set. Besides, we adopt some refinement to remove unreliable novel proposals, which achieves a score of 2.33. Notably, we find that the 'tram' in SODA is the same as the 'bus' in CODA, hence we modify the 'tram' to 'bus' in submissions and achieve an improvement of 0.13.

Table 2. Ablation study of unsupervised corner case detection on the validation leaderboard and test leaderboard.

Aug	Common pseudo-pretrain	tram->bus	Novel	Dataset	Val	Test
					1.64	-
✓				SODA5k	1.77	-
✓	10w				1.93	-
✓	1000w				2.07	-
✓	1000w		refine	SODA1w	2.33	-
✓	1000w	✓	refine		2.46	2.40

Supervised Corner Case Detection. The detailed ablation studies in supervised corner case detection are shown in Tab. 3. We load the best close-set model as the pre-trained model and fine-tune the CODA validation set, which could greatly improve the performance to 3.07 on the CODA test set. Adopting 11w and 40w pseudo-labels containing novel classes from 100w and 300w unlabeled images respectively for pseudo-supervised learning, we further improve the better performance to 3.08 and 3.09.

Table 3. Ablation study of supervised corner case detection on the test leaderboard.

Pretrain	Train Set	Test
best close-set model		3.07
+ novel pseudo 11w	CODA val	3.08
+ novel pseudo 40w		3.09

4 Conclusions

In this paper, we introduce an effective pipeline for unsupervised and supervised corner case detection. Especially, we propose a large-scale pseudo-supervised method for both close-set detection and open-set detection, then we iterative refine the mined novel proposals in a joint training manner. With these methods mentioned above, we achieve great performance improvement in both unsupervised and supervised settings. Finally, our proposed approach achieves the best performance on all metrics and wins the 1st prize with a clear margin in the corner case detection challenge.

References

1. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
2. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
3. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. (2019) 9627–9636
4. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
5. Cai, Z., Vasconcelos, N.: Cascade r-cnn: high quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence **43**(5) (2019) 1483–1498
6. 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. (2021) 5830–5840
7. 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. (2022) 9235–9244
8. 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. (2022) 9591–9600
9. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 10012–10022
10. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 113–123
11. Han, J., Liang, X., Xu, H., Chen, K., Hong, L., Mao, J., Ye, C., Zhang, W., Li, Z., Liang, X., Xu, C.: Soda10m: A large-scale 2d self/semi-supervised object detection dataset for autonomous driving (2021)
12. Mao, J., Niu, M., Jiang, C., Liang, H., Chen, J., Liang, X., Li, Y., Ye, C., Zhang, W., Li, Z., et al.: One million scenes for autonomous driving: Once dataset. arXiv preprint arXiv:2106.11037 (2021)
13. 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. (2016) 770–778
14. Li, K., Chen, K., Wang, H., Hong, L., Ye, C., Han, J., Chen, Y., Zhang, W., Xu, C., Yeung, D.Y., et al.: Coda: A real-world road corner case dataset for object detection in autonomous driving. arXiv preprint arXiv:2203.07724 (2022)
15. Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: Soft-nms–improving object detection with one line of code. In: Proceedings of the IEEE international conference on computer vision. (2017) 5561–5569
16. Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al.: Mmdetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)