# Loop Mining Large-Scale Unlabeled Data for Corner Case Detection in Autonomous Driving

Jiawei Zhao*, Yiting Duan*, Jinming Su, Wangwang Yang, Tingyi Guo, Xingyue Chen, Junfeng Luo

Meituan
{zhaojiawei12,duanyiting,sujinming,yangwangwang,guotingyi, chenxingyue02,luojunfeng}@meituan.com

**Abstract.** For obstacle detection in road scenes, it is very challenging to detect novel objects 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 a open-set model separately in a pseudo-supervised learning manner, and then iteratively improve the performance of both models through the proposed loop-optimization strategy, which employs some useful tricks to remove false positive detections about corner cases. Experimental evidence demonstrates that our approach achieves new state-of-the-art on the popular CODA dataset.

**Keywords:** corner case, pseudo-supervised learning, close-set detection, open-set detection, loop-optimization

## 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 the recognition [1–3], segmentation [4–6], detection [7–9] tasks for common traffic obstacles (*e.g.*, cars, pedestrians, cyclists.). However, such detectors are often incapable of detecting novel objects that have not been seen or rarely seen in the training data. These novel objects are called corner cases, which include new instance of common class (*e.g.*, an overturned truck), instance of 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 [7, 10], FCOS [8], Faster RCNN [9], Cascade RCNN [11]) are generally applied. However, close-set methods heavily rely on well-annotated

* Equal contribution

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 novel categories that have not been seen before, such as ORE [12], OWDERT [13], OpenDet [14], these methods perform poorly in the real world.

In order to solve these problems, we mainly propose three aspects for optimization. First, we introduce an iterative pseudo-supervised learning method to obtain the close-set model for known class detection. Second, based on the OpenDet [14], we train the open-set model in the road scene for unknown corner case detection. And, some critical processes including point-cloud based depth estimation, ground mask segmentation, and clustering are applied to remove unreliable corner-case predictions. Last but not least, we introduce a loop-optimization strategy to boost the performance of both common and corner-case object detection. Our proposed approach achieved the-state-of-art in the CODA test leaderboard and won first prize in the 2022 SSLAD Corner Case Detection Challenge, this paper is an extended version of our approach [15, 16], including more implementation details and experiments.

## 2 Approach

We introduce two independent settings for corner case detection, depending on whether to use annotated corner case labels. Considering whether there are labeled data for corner cases, our method includes two independent settings. 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.

### 2.1 Unsupervised Corner Case Detection

Assuming the novel classes of objects are unlabeled, we propose an effective pipeline that jointly uses a close-set detection model and an open-set detection model to exploit both labeled common class data and unlabeled novel class data. As shown in Fig. 1, we separately train a close-set detection model for objects of known-classes and an open-set detection model for objects of unknown-classes, then iteratively optimize the two models to obtain a powerful detection model.

**Close-Set Detection Model.** For the close-set detection model of known class objects, we apply cascade RCNN with FPN as our baseline architecture and swin-base as our backbone.

**(1) Data Augmentation**: To perform more robustly on novel instances of common classes, we apply multiple augmentation methods (*e.g.*, AutoAugmentV2, MixUp, Albu) to enrich the samples during training. We first use color/bright augmentations (ColorJitter, RandomBrightnessContrast, RGBShift)

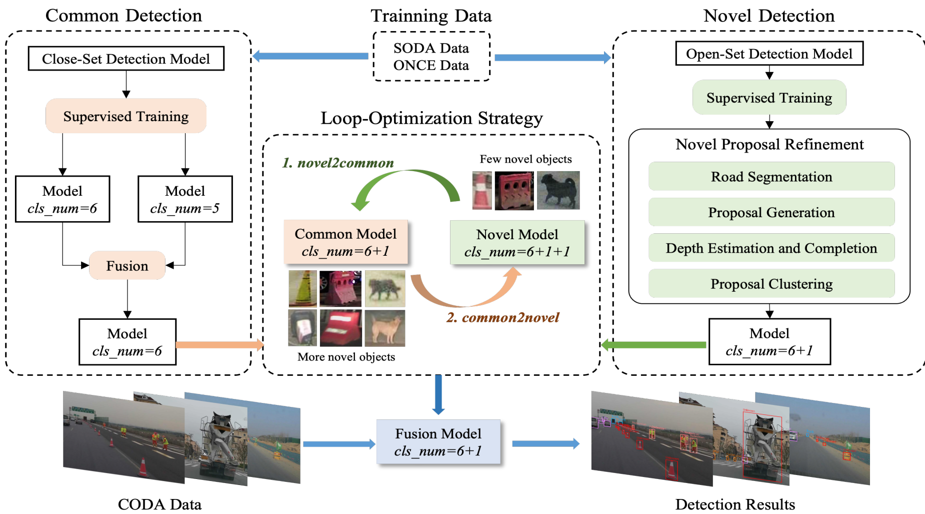


**Fig. 1.** Framework of our proposed approach.

and weather augmentations (RandomFog, Random Rain) implemented in Albu. Then two random samples are selected to fuse with a random weighted summation. At last, we apply the v2 policy implemented in AutoAugment.

**(2) Pseudo-supervised learning**: To exploit the large-scale unlabeled images in SODA10M [17], we propose a simple but effective pseudo-supervised method to iterative train the pretrain weights with more accurate pseudo labels and more unlabeled images as shown in Fig. 2. In supervised learning, large-scale pseudo labels would introduce more noise and limit performance improvements. To solve this problem, we utilize pseudo labels to train a more robust representation as the pretrain weights for supervised learning. We first train a strong two-stage detector with ImageNet-1K [18] pretrain as Fig. 2 (a), then we iterative predict the unlabeled images and retrain the detector with only pseudo labels as Fig. 2 (b) to conduct better pretrain weights, which plays a key role in our pipeline. To avoid the huge time consumption of training large-scale unlabeled images, we first resize images to a smaller resolution (360, 640) to train 50w/500w/1000w pseudo labels for one epoch respectively. Notably, we filter the pseudo labels with confidence 0.8.

**Open-Set Detection Model.** For the open-set detection model of unknown class objects, we use the OpenDet Detector [14] as our baseline architecture, a sota 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

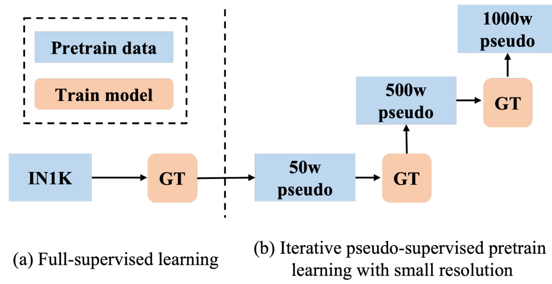


**Fig. 2.** Pipeline of our solution.

open-set model with data from road scenes, where the backbone is ResNet50 with ImageNet-1K pretrained 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 refine novel proposals from four aspects.

**(1) Point-cloud based road segmentation**: there is a fact that only those obstruct or has a potential to obstruct the road are annotated for objects of novel categories. Therefore, we extract the road segmentation mask to refine proposals of novel categories, ensure that novel objects are not far from the ground, and suppress false proposals on the ground. Specifically, we regard the ground as a plane and cluster the point clouds of ground based on RANSAC [19] to obtain the results of point-cloud road segmentation. With the point-cloud road segmentation, we project the point clouds of the road onto the image, and then calculate the concave hull polygon of these projected points to get the segmentation mask of the road in the image. Because only the ONCE [20] dataset provides the point cloud data, we conduct above operation on the ONCE and retrain a road segmentation model based on the OCRNet [21].

**(2) Point-cloud based proposal generation**: in order to increase the recall of objects of novel categories, we conduct DBSCAN [22] to cluster the point clouds near the ground with ground-level points removed. In this manner, we get get a number of novel proposals, which is helpful to the recall of novel objects but also introduces some false positive proposals. In the final model, we didn't adopt this strategy, but it is worth exploring further.

**(3) Depth estimation and completion**: small objects in the distance are usually detected accurately, so the extracted proposals from small objects in the distance is usually inaccurate. We filter out small objects in the distance by the depth and size of the bounding box. In order to get the depth of the image, we calculate the depth with cloud points on the ONCE dataset, and then use a traditional depth completion method IP-Basic [23] to get the pixel-wise dense depth map. Therefore, the distant noise proposals can be filtered according to the depth map with the size of bounding box.

**(4) Proposal clustering**: The open-set model generally miss some corner case objects, and there are few corner case objects in real road scenes, both cause the class imbalance problem during the training of common detector. Besides, there are some wrong results in corner case proposals due to low confidence. In order to alleviate the above problems, we clustered the novel class proposals mined from the training dataset, selected some high confidence novel class objects as template for copy-paste data augmentation in the training of common detector. Specifically, we use ResNet50 trained on ImageNet as the encoder to extract the features of the proposals, and then use the K-means method for clustering.

**Loop-Optimization Strategy.** Considering that the trained open-set model generally miss detecting some corner cases, we introduce a loop-optimization strategy to boost the performance of both common and corner-case object detection. Specifically, on the one hand, we obtain unknown but reliable corner-case objects through the open-set model, and label these objects as the known corner class to retrain the close-set model; on the other hand, we get known corner cases by the retrained close-set model, and use them as known corner class to retrain the new open-set model. After several iterations of this loop, a powerful detector can be obtained.

### 2.2 Supervised Corner Case Detection

Assuming there are some labeled novel classes, we train the close-set detection model on the CODA validation set. Then we obtain the reliable pseudo-labels on unlabeled SODA [17] images and retain the images having novel-class predictions for pseudo-supervised learning. At last, we train the CODA [24] validation set with obtained pseudo-pretrained model and achieve the best performance on both common classes and novel classes.

## 3 Experiments

### 3.1 Datasets

SODA10M [17] is a 2D object detection dataset , which contains 10 million unlabeled images and 20k images fully-annotated with 6 representative categories (pedestrian, cyclist, car, truck, tram, tricycle). ONCE is a 2D and 3D object detection dataset which contains 1 million LiDAR frames, 7 million camera images, and 15k fully-annotated scenes with 5 categories (car, bus, truck, pedestrian, cyclist). Besides, CODA which collected from SODA10M and ONCE is provided for evaluation. SODA10M and the validation set of CODA 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 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, 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, 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. All experiments are conducted on 16 NVIDIA A100 GPUs. Our implementation is based on the open-source object detection toolbox MMDetection [25].

### 3.3 Performance Analysis

The performance on the test leaderboard is 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.** Performance on the CODA 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 |
| haoooooo | 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 pretrain 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 1000w pseudo-pretrain on the SODA train set and validation set.

To detect the novel classes in an end-to-end manner, we utilize the open-set model to generate initial novel-class predictions on the SODA train and validation set. For the novel proposal refinement, we tried four versions successively. In refine v1, the strategies include removing false novel proposals by ground mask model, and filtering overlapped proposals by NMS with common class and novel class respectively. Based on v1, in refine v2, we adopt a stronger ground mask model and filter those proposals that have too larger or too small areas. Further, in refine v3, we remove those novel proposals with low confidence and filter false proposals which nearly locate on the edges of images or inside common boxes. In refine v4, we adopt copy-paste by novel proposal clustering based on v3.

**Table 2.** Ablation study of unsupervised corner case detection on CODA dataset.

| Aug | Common pseudo-pretrain | tram->bus | Novel | Dataset | Val | Test |
|---|---|---|---|---|---|---|
| | | | | | 1.64 | - |
| ✓ | | | | SODA5k | 1.77 | - |
| ✓ | 10w | | | | 1.93 | - |
| ✓ | 1000w | | | | 2.07 | - |
| ✓ | 1000w | | refine v1 | | 2.31 | - |
| ✓ | 1000w | | refine v2 | SODA1w | 2.33 | - |
| ✓ | 1000w | ✓ | refine v2 | | 2.46 | 2.40 |
| ✓ | 1000w | ✓ | refine v3 | | 2.47 | - |
| ✓ | 1000w | ✓ | refine v4 | | 2.51 | - |

**Supervised Corner Case Detection.** We first apply ablation studies of pseudo-supervised learning on common case detection dataset as shown in Tab. 3. In the pseudo-supervised learning stage, we first obtain reliable pseudo-labels on 50w unlabeled images by the best-performed model and train the network with a small resolution (360, 640). The model trained on pseudo labels is further used as the pretrain weights of supervised learning. We iterative generate more accurate pseudo-labels on more unlabeled images and train better pretrain weights for supervised learning. Notably, our proposed pseudo-supervised learning could effectively improve performance by almost 10%.

**Table 3.** Ablation study of pseudo-supervised learning on SODA val set.

| Pseudo-pretrain | Pretrain img size | Train img size | SODA Val mAP |
|---|---|---|---|
| 50w | (360,640) | (1080,1920) | 76.45 |
| 500w | (360,640) | (1080,1920) | 79.28 |
| 1000w | (360,640) | (1080,1920) | 81.44 |

The detailed ablation studies in supervised corner case detection are shown in Tab. 4. We utilize the pseudo-pretrain weight based on SODA to obtain the best close-set model, then iteratively load the best close-set model as pretrain 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.

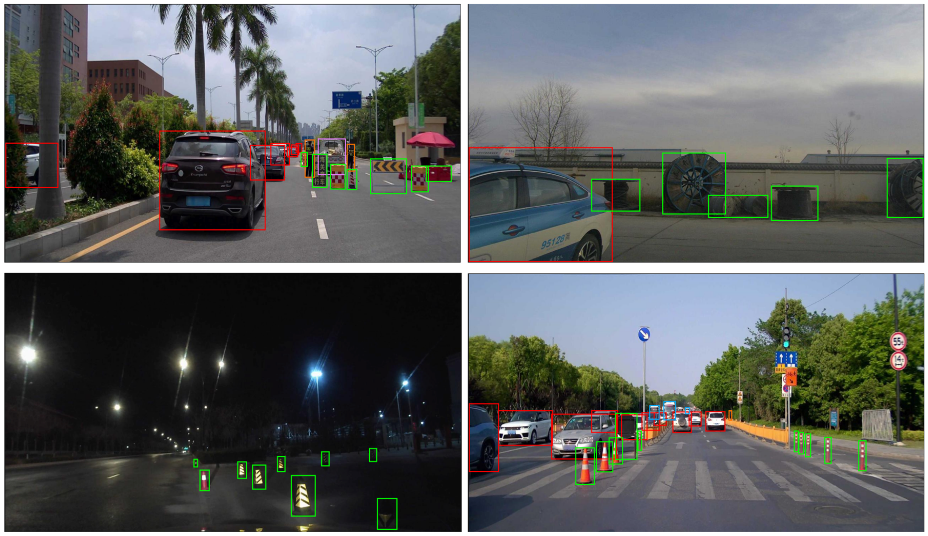

**Fig. 3.** Representative visual examples from the proposed solution. Noted that the novel objects are drawn in green, while the common objects are drawn in other colors.

**Table 4.** Ablation study of supervised corner case detection.

| 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 2022 SSLAD Corner Case Detection Challenge.

## 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. (2016) 770–778
2. Zhao, J., Yan, K., Zhao, Y., Guo, X., Huang, F., Li, J.: Transformer-based dual relation graph for multi-label image recognition. In: Proceedings of the IEEE/CVF international conference on computer vision. (2021) 163–172
3. Zhao, J., Zhao, Y., Li, J.: M3tr: Multi-modal multi-label recognition with transformer. In: Proceedings of the 29th ACM international conference on multimedia. (2021) 469–477
4. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4) (2017) 834–848
5. Zhao, J., Zhao, Y., Li, J., Chen, X.: Is depth really necessary for salient object detection? In: Proceedings of the 28th ACM international conference on multimedia. (2020) 1745–1754
6. Zhao, Y., Zhao, J., Li, J., Chen, X.: Rgb-d salient object detection with ubiquitous target awareness. IEEE Transactions on Image Processing **30** (2021) 7717–7731
7. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
8. 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
9. 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)
10. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
11. 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
12. 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
13. 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
14. 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
15. Zhao, J., Duan, Y., Su, J., Yang, W., Guo, T., Luo, J., Wei, X.: 1st place solution for sslad challenge 2022: Corner case detection
16. Zhao, J., Chen, X., Cui, Z., Li, X., Luo, J., Wei, X.: 1st place solution for sslad challenge 2022: 2d object detection
17. 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)

18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6) (2017) 84–90
19. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6) (1981) 381–395
20. 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)
21. Yuan, Y., Huang, L., Guo, J., Zhang, C., Chen, X., Wang, J.: Ocnet: Object context for semantic segmentation. International Journal of Computer Vision **129**(8) (2021) 2375–2398
22. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. Volume 96. (1996) 226–231
23. Ku, J., Harakeh, A., Waslander, S.L.: In defense of classical image processing: Fast depth completion on the cpu. In: 2018 15th Conference on Computer and Robot Vision (CRV), IEEE (2018) 16–22
24. 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)
25. 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)