Overcoming the Challenges of Long-Tail Distribution in Nighttime Vehicle Detection

Houwang Zhang and Leanne Lai Hang Chan, *City University of Hong Kong, Hong Kong, 999077, China.*

As the basic task of an intelligent transportation system, nighttime vehicle detection is associated with many challenges. Existing methods usually ignore the significant challenges that arise from imbalanced class distribution among vehicles, which always leads to poor detection for vehicles that belong to tail classes. By analyzing the existing solutions for long-tail object detection and considering the complex and diverse characteristics of nighttime traffic scenarios, we propose an enhanced detection approach based on anomaly detection. In addition, to tackle disturbance from complex lights, we reconstruct the loss function for background proposals, thus allowing the detector to pay more attention to hard-classified proposals and learn to distinguish vehicle lights from disturbed light resources. Comprehensive experiments prove that, compared with generic approaches, our proposed method can effectively solve the problem of long-tail distribution in nighttime vehicle detection and improve robustness in complex environments.

According to related research, most of the fatal vehicle accidents in recent decades were caused by rear-end collisions.[1] Hence, the intelligent traffic system (ITS) concept has emerged, with the aim of solving a series of road traffic problems, through the use of advanced driver assistance systems and autonomous driving systems.[2] As the basic component of an ITS, vehicle detection is at the root of the whole system, and a reliable and effective vehicle detection method is an important support and guarantee for subsequent automated traffic processing and operations.[3]

With the development of related theories in areas of image processing and computer vision, vehicle detection techniques have become mature and are now common in daily applications, such as widely used faster region-based convolutional neural networks (R-CNNs)[4] and you only live once (YOLO).[5] However, current vehicle detection methods are essentially designed for daytime scenes, whereas most of the traffic accidents occur at night.[6] Research reports also indicate that, compared with daytime conditions, it is more dangerous to drive at night and the possibility of traffic accidents increases in nighttime scenarios.[7] Hence, there is an urgent need for effective and accurate methods of nighttime vehicle detection.

Digital Object Identifier 10.1109/MIS.2024.3350628
Date of publication 1 August 2024; date of current version 10 April 2024.

Although state-of-the-art deep-learning-based object detectors can be directly applied to nighttime traffic scenes, they do not perform well for vehicle detection under low-light conditions because of two main problems: 1) the overall brightness and contrast are too low in these images and 2) vehicle distribution is always nonuniform. In regard to the former issue, due to poor lighting conditions, vehicle features such as shape, color, texture, and the gradient within the image are not salient and are always masked, which is also the reason why the performance of these detectors is worse in nighttime scenes than in daytime ones.[8] In recent years, many schemes have been developed for enhancing low-light traffic images, which may well solve the problems of low brightness and contrast, but researchers seldom pay attention to the latter challenge.

Figure 1 shows vehicle distribution in the widely used Berkeley Deep Drive (BDD) dataset.[9] Here, we selected traffic images captured under low-light conditions in two main types of scenes: city street and highway. As observed, in both scenes, certain vehicles, such as cars and trucks, account for nearly 99% of all the objects, and these can be taken as head classes. However, samples of other classes, such as buses, bikes,

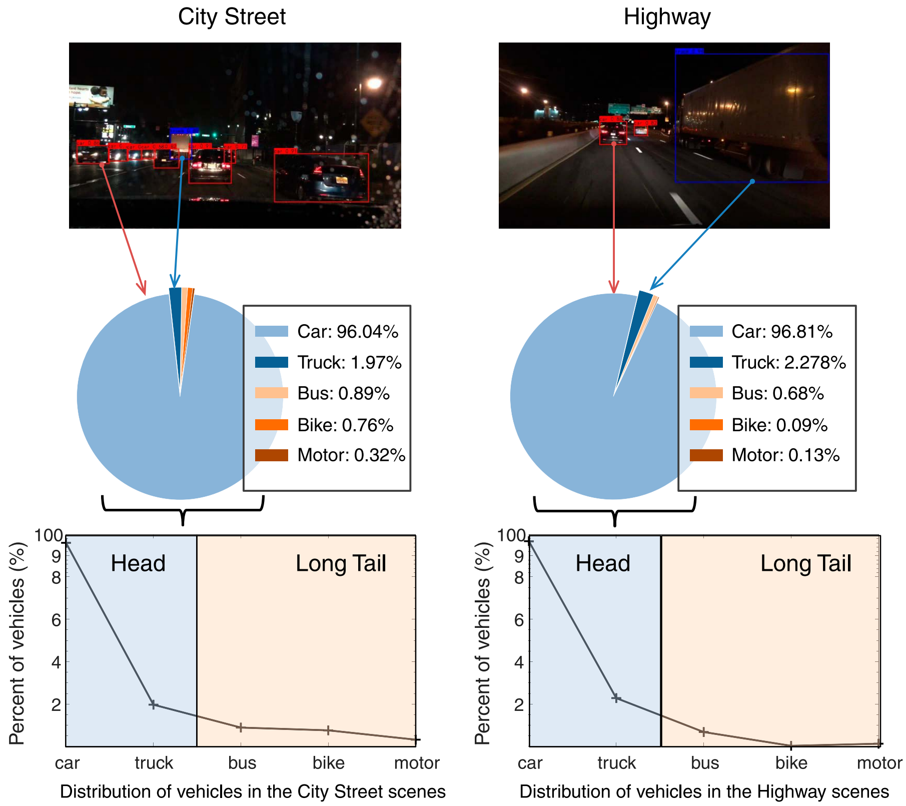

FIGURE 1. Illustration of the distribution of vehicles in two scenes: city streets and highways.

and motors, constitute a relatively small proportion of the overall targets and should be taken as tail classes. Vehicle distribution in nighttime traffic scenarios presents a long-tail distribution, and the extremely imbalanced class distribution among vehicles will always result in a performance gap between the head and tail classes.[10] More specifically, tail classes can easily be overwhelmed by head classes during training, and the detector may predict the samples in the tail classes to be head classes or background,[11] which greatly decreases average detection precision. In addition, due to disturbances from other lights (such as lamps and lighted buildings), some background proposals (i.e., candidate proposals generated from the background) may be misclassified as vehicles. These issues inevitably pose challenges for vehicle detection under dim-light traffic conditions.

Eliminating long-tail distribution from the task of general object detection has been a hot topic for researchers. Current research works are mainly divided into two categories: resampling strategies and reweighting methods. For resampling strategies, most of the works regard the problem of long-tail distribution as being influenced by a sampling imbalance in the training batch. Based on this assumption, resampling strategies usually involve the design of special sampling techniques.[11] However, this kind of approach inevitably results in a distortion of the original distribution, which impairs representation learning and causes problems such as overfitting or underfitting.[12] Hence, in recent years, researchers have turned to reweighting methods, in which the loss function is reconstructed to deal with the imbalance problem among different classes, and which can effectively weaken the long-tail distribution influence on natural scene datasets such as Large Vocabulary Instance Segmentation.[13]

Unlike general object detection tasks, there are unique challenges that make the long-tail distribution problem harder to solve in nighttime scenes. Existing methods for long-tail object detection are designed for large-scale datasets with thousands of categories, while for nighttime vehicle detection, there are only a few classes of vehicles. Although there is a fairly serious imbalance among the different classes of vehicles, class distribution is only similar to long-tail distribution, rather than perfectly fitting the shape. In addition, due to nighttime scene characteristics, there is always interference caused by light sources, which may affect vehicle detection and result in a high percentage of false positives. This means that current reweighting methods are not very effective for nighttime traffic scenes.

To overcome these challenges, we enhance the detection model with an anomaly detection scheme to address the issue of nighttime vehicle detection. To weaken the influence of extraneous lights in the image, we reconstruct the loss function for the background proposals based on intersection over union (IoU). Our key contributions can be summarized as follows:

- We unveil the long-tail distribution problem in nighttime vehicle detection, which is typically ignored by researchers, and greatly limits the overall performance of vehicle detection in low-light traffic conditions.
- By analyzing the unique challenges associated with nighttime vehicle detection, we show that existing state-of-the-art reweighting methods designed for large-scale datasets cannot be straightforwardly applied to low-light traffic scenes. We solve the problem of long-tail distribution among vehicles using anomaly detection. More specifically, in the training process, we regard the detected proposals that belong to tail classes as abnormal cases for the detection model.
- Considering the interference of lights, such as lamps or lights from buildings, we reconstruct the loss function of background proposals, through which the detection model can learn to distinguish vehicle lights from extraneous lights in environments containing multiple disturbances.

RELATED WORK

Resampling Methods

To tackle the problem of long-tail distribution, resampling methods are widely used in the early stage. Resampling methods usually oversample additional training data for the tail classes or undersample data for the head classes to generate more balanced samples.[11] Although resampling methods can be applied to the problem of long-tail distribution, they will inevitably cause problems as they change the distribution of the original data space.

For example, typical oversampling methods, such as repeat factor sampling,[13] resample data in the tail classes during the training process using different sampling frequencies for different categories, although there will always be a high potential risk of overfitting. Unlike oversampling methods, undersampling methods, such as those in Drummond et al.,[14] aim to remove samples from the head classes to make the overall training samples uniform and balanced. However, undersampling methods will always cause a decrease in overall performance.

Reweighting Methods

Reweighting methods rely on the basic idea of assigning different weights to different samples that belong to special categories. Reweighting methods can effectively bring about improvement in tail classes, but they may introduce issues like optimization challenges[15] and lead to suboptimal overall performance.

In addition to methods that reweight samples at the class level, several recent studies have tried to adjust the weights of training data at the sample level. Focal loss[16] was proposed to tackle the class imbalance in an approach where the training samples were divided into well- and hard-classified categories. Equalization loss[11] (EQL) simply ignores sample gradients in the tail classes to avoid the proposals for the tail classes being oversuppressed by those of the head classes. Similarly, adaptive class suppression loss (ACSL) aims to estimate the suppression gradients of each sample adaptively from a statistic-free perspective,[17] through which the problem of long-tail distribution could be well solved.

Reweighting methods such as focal loss, EQL, and ACSL can effectively enhance general object detection tasks with long-tail distribution. However, most of these approaches need to estimate the frequencies of different classes, which may introduce inconsistency when applied to new scenes. Furthermore, these works focus only on the imbalance among foreground samples and ignore the imbalance of the background samples in nighttime traffic scenes. In this article, we propose a more general framework based on anomaly detection, which does not rely on previous category knowledge like frequency distribution or occupation ratio for each category. In addition, we also reconstruct the loss function for background samples to weaken the disturbance from extraneous lights for nighttime vehicle detection.

METHODOLOGY

Anomaly Detection for Tail Classes

An overview of the proposed pipeline, which is based on the widely used faster R-CNN, is given in Figure 2. The proposals generated during the detection progress can be divided into three categories: head classes (such as cars or trucks), tail classes (such as motors or bikes), and background classes. All of the foreground

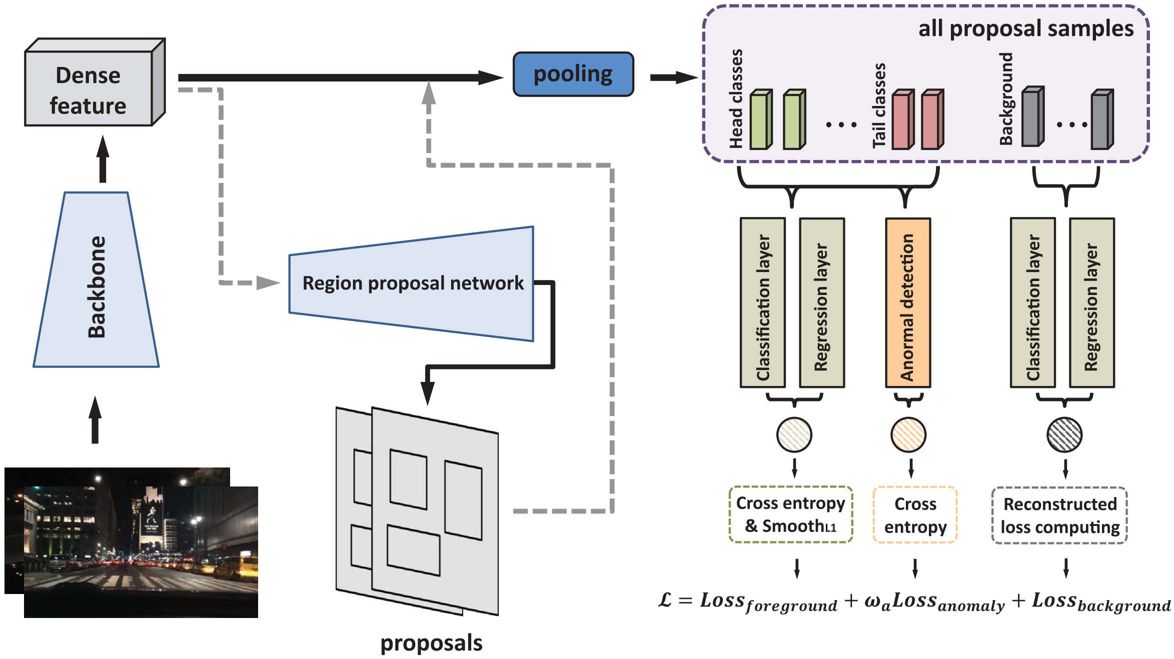

FIGURE 2. Overview of the proposed pipeline based on a faster R-CNN for nighttime vehicle detection.

proposals may contain head classes or tail classes; we regard the proposals that belong to tail classes as anomaly points, using a classification layer as the anomaly detection module to detect them. In this way, the detection model can learn to distinguish between proposals of tail classes and head classes and does not need to compute data distribution before training.

As discussed in the previous section, most of the reweighting methods for long-tail object detection rely on frequency estimations for all categories. However, for nighttime traffic scenes, vehicle and pedestrian distributions vary heavily at different times and locations; that is, situations during rush hour and late nighttime are totally different. Assigning new weights to changing distributions could be time consuming and laborious and may limit the applications of these schemes in real-world scenarios. Directly detecting the proposals of tail classes as anomaly points can avoid this problem and yields a better generalization ability for dynamic traffic scenes because it guides the detection model to learn a global classification for head and tail classes rather than several categories.

The $\mathrm{Loss}_{\mathrm{foreground}}$ for all the foreground proposals can be obtained from the specially designed functions of the detection model, such as cross entropy for classification and Smooth_{L1} for localization regression.

We can directly use cross entropy to compute the $\mathrm{Loss}_{\mathrm{anomaly}}$ for each foreground proposal x_f as

$$\mathrm{Loss}_{\mathrm{anomaly}}(x_f) = -\log(\hat{p_i})$$
$$\hat{p_i} = \begin{cases} p_i, & \text{if } y_i \in \text{tail classes} \\ 1 - p_i, & \text{otherwise} \end{cases} \quad (1)$$

where p_i denotes the output of the anomaly detection layer for the foreground proposal x_f, which indicates the probability of x_f belonging to tail classes, and y_i represents the corresponding ground-truth label of x_f.

Reconstructed Loss Function

For vehicle detection under nighttime traffic scenes, the headlights or taillights of vehicles are the most salient features and form the core target of traditional nighttime vehicle detection. When carried out in environments with complex lighting, such as urban roads, there are many other lights in addition to the headlights or taillights of vehicles, such as building lights, street lamps, reflected lights from car bodies and road reflectors, and so on, which increases the probability of false or missing detection. As shown in Figure 3(a) and (b), areas with lights such as street lamps or reflections appear very similar to vehicles in nighttime traffic scenarios and may be misclassified as vehicles.

To decrease such false-positive results, we need to analyze the background proposals generated during the training process. As shown in Figure 3(c), there are

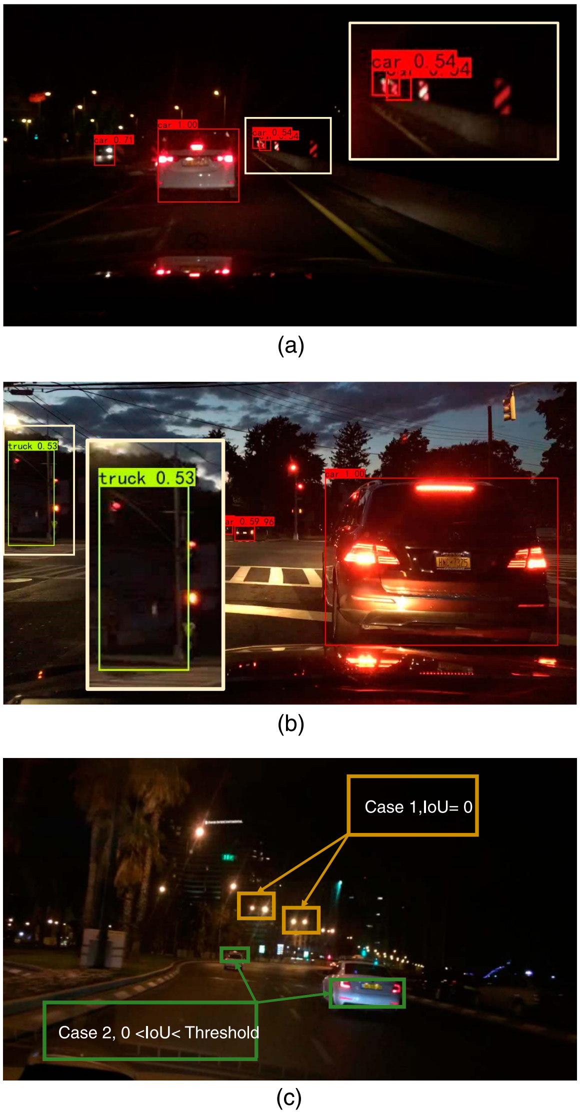

FIGURE 3. Illustration for the false detection cases. (a) and (b) Examples of false detection due to disturbance from interfering light sources. (c). Cases for the detected background proposals that may be misclassified during training.

two main kinds of background proposals that may be misclassified at the training stage. Case 1 (the yellow boxes) contains regions with interference lights (especially paired lights like street lamps), which are easily misclassified as vehicles. Case 2 (the green boxes) includes the regions that capture only certain parts of vehicles; because of setting the IoU threshold (such as 0.3 or 0.5), they will still be classified as negative proposals during the training process. Case 1 arises due to the interference caused by light sources, while case 2 emerges from the IoU threshold configuration. Different from case 1, case 2 manages to capture a segment of the vehicle, signifying that the false detection proposals originating from light-source interference require enhanced sensitivity in the detector. As a result, the loss function should assign greater penalties to the misclassified proposals that closely resemble case 1, which effectively addresses this issue.

To determine whether the generated proposal is similar to case 1 or case 2, the IoU between the predicted box with the target box can be used as a criterion: a higher IoU value means that the generated proposal is closer to case 2, while proposals with a lower IoU is similar to case 1. Hence, for each background or negative proposal x_b generated in the training process, by referencing the focal loss[16] and the IoU-based loss[18,19] we compute the cost as

$$\begin{aligned}
&\mathrm{Loss}_{\mathrm{background}}(x_b) = \mathrm{loss}_{\mathrm{cls}}(x_b) + \omega_l \mathrm{loss}_{\mathrm{loc}}(x_b)\\
&\mathrm{loss}_{\mathrm{cls}}(x_b) = -\sum_{i=1}^{C} \alpha (1 - C_{\mathrm{IoU}})^{\beta} \log(\hat{q_i})\\
&\hat{q_i} = \begin{cases} q_i, & \text{if } \ y_i = 1\\ 1 - q_i, & \text{otherwise}\end{cases}\\
&\mathrm{loss}_{\mathrm{loc}}(x_b) = C_{\mathrm{IoU}}
\end{aligned} \tag{2}$$

where q_i is the output of the classification layer for x_b, and the total number of categories is C. C_{IoU} denotes the IoU of x_b with the closest target box in the image, and y_i represents the corresponding ground-truth label of x_b. α and β are used as adjustment parameters for the background proposal weights. ω_l is a parameter for the localization loss $\mathrm{loss}_{\mathrm{loc}}$.

In (2), we also compute the localization loss $\mathrm{loss}_{\mathrm{loc}}$ for background proposals, which are usually ignored in a normal detection stream. A standard detection model pays attention only to the localization of foreground proposals, while here we want to make the network assign higher costs to background proposals like case 1. In this way, the network will decrease the ratio of background proposals like case 2 and learn to be more sensitive to hard-to-classify proposals like case 1.

After computing all the losses for the foreground proposals, background proposals, and anomaly detection, the sum of losses $\mathcal{L}$ can be computed as

$$\begin{aligned}
\mathcal{L} = &\ \mathrm{Loss}_{\mathrm{foreground}} + \omega_a \mathrm{Loss}_{\mathrm{anomaly}}\\
&+ \mathrm{Loss}_{\mathrm{background}}
\end{aligned} \tag{3}$$

where ω_a is the anomaly detection weight of tail classes.

EXPERIMENTS

Dataset

For experiments, we adopted the widely used BDD dataset[9] and Hong Kong nighttime vehicle detection

(HKNVD) dataset[6] for comparisons with related schemes. For the BDD dataset, we chose images captured under nighttime conditions and divided them into two main scenes: highways and city streets. Specifically, for city street scenes, 10,000 images are used for training and 4945 images are used for testing. For highway scenes, we selected 4000 images for training and 2019 images for testing. For the HKNVD dataset, 500 images and 336 images are chosen for training and testing, respectively.

For two subdatasets in the BDD dataset, we chose five categories of annotated vehicles: cars, buses, trucks, motors, and bikes. For the HKNVD dataset, cars, taxis, buses, and minibuses are annotated with labels. Empirically, we set buses, motors, and bikes as the tail classes for the BDD dataset and, similarly, we set buses and minibuses as the tail classes for the HKNVD dataset.

Implementation Details

To evaluate the effectiveness of our proposed anomaly detection scheme and the reconstructed loss function for nighttime vehicle detection, we used classical detectors faster R-CNN and YOLOv5 as the baseline models. In both the training and test streams, we resized the input images to 600 $\times$ 600. At the training stage, we applied the Adam method as the optimizer with an initialized learning rate le^{-5} and weight decay $5le^{-4.}$ For the loss function weights, ω_a was set to three for anomaly detection, and ω_l was set to five for background proposal locations. The parameters for adjusting the weights of background proposal classifications, α and $\beta,$ were set to 1.75 and 1.2, respectively.

For the highway and city street scenes in the BDD dataset, we set the training epochs at 50 and 25, respectively. For the HKNVD dataset, we set the training epochs at 200. When testing, non-maximum suppression with an IoU threshold of 0.5 was adopted to remove overlapping proposals. Other hyperparameters, such as the size of anchors, number of foreground and background proposals, and so on, were set at default values. More details can be found at https://github.com/ProfHubert/Imbalanced-Detection.

Comparison With the Baseline

Results of the baseline model comparison are given in Table 1(a). We used mean average precision with an IoU threshold of 0.5 to evaluate the detection results from each model on each category. As shown in the table, compared with the original detector, a faster R-CNN or YOLOv5 with the proposed strategies achieved great improvements on all the highway, city street, and HKNVD dataset. Generally speaking, the detector's performance on the HKNVD dataset is better than on the two BDD subdatasets, which can be attributed to the imbalance being not so severe and annotations being not so dense and elaborate on the HKNVD dataset.

Moreover, by observing Table 1(a), although head class performances (such as cars) may be slightly weaker than in the original model, the overall performance is still much better compared with the original results. The tail class performances (especially bikes and minibuses) improved significantly. For instance, on a faster R-CNN with ResNet50, our approach boosted the average precision of the tail classes by roughly 37% (29.58 versus 21.54) on the city street scenes for bikes, 29% (22.97 versus 17.67) on the highway scenes for buses, and 4% (97.42 versus 93.07) on the HKNVD dataset for minibuses.

Comparison With Other Solutions for Long-Tail Detection

To demonstrate the effectiveness of our proposed approach, we compared it with existing state-of-the-art algorithms for long-tail object detection, such as focal loss,[16] EQL,[11] ACSL,[17] and effective class-margins loss (ECL),[20] using the same baseline faster R-CNNs with ResNet50 model. For these methods, the vehicles belonging to head or tail classes were defined using the same rule stated in the previous section, and hyperparameters for each approach were set as default values.

The experimental results from these methods are listed in Table 1(b), and it can be seen that compared with these methods, the detector based on our proposed approach achieved obviously better results. Although these methods can slightly improve vehicle detection performance for head classes such as cars, their overall average precision was roughly three points lower than our methods. As stated in the previous section, most of the existing methods for long-tail object detection can be effectively applied to natural datasets but do not perform very well in nighttime traffic scenarios. Take the city street scenes, for example: our approach outperformed ACSL by 3.58 points (34.78 versus 31.2), EQL by 3.13 points (34.78 versus 31.65), ECL by 3.55 points (34.78 versus 31.23), and focal loss by 3.94 points (34.78 versus 30.84), which further proved the effectiveness and superiority of our proposed methods.

We also presented a visual comparison of these methods in Figure 4. It can be observed that the detector based on our proposed method can precisely detect objects in all the categories. For instance, as shown in Figure 4(a), the detector based on our

TABLE 1. Quantitative comparison results on the BDD dataset and HKNVD dataset. (a) Comparison of results from the baseline models. (b) Comparison of results from the proposed methods with other strategies using the baseline model. (c) Results of ablation experiments for the proposed strategies on the city street scenes of BDD dataset, A represents anomaly detection used for tail classes, B denotes the reconstructed loss function for background proposals. Faster R-CNN* or YOLOv5*: the original detector with anomaly detection and reconstructed loss function $\mathrm{Loss}_{\mathrm{background}}$.

Method	Backbone	City Street						Highway						Hong Kong				
		Car	Bus	Truck	Motor	Bike	Average	Car	Bus	Truck	Motor	Bike	Average	Car	Taxi	Bus	Minibus	Average
Faster R-CNN	Res50	56.36	36.46	31.85	5.66	21.54	30.37	43.01	17.67	32.8	0.75	9.09	20.66	88.91	90.64	88.35	93.07	90.24
Faster R-CNN*	Res50	56.19	**42.83**	35.31	**9.97**	**29.58**	**34.78**	44.18	22.97	35.01	**6.08**	17.39	25.12	89.72	92.44	91.36	97.42	92.74
Faster R-CNN	Res101	55.1	35.61	35.1	7.14	20.64	30.71	44.09	23.37	**35.03**	0.85	4.17	21.5	89.14	91.4	88.84	93.43	90.7
Faster R-CNN*	Res101	56.63	39.79	**38.24**	8.66	27.21	34.11	43.08	**24.41**	34.55	3.56	**24.78**	**26.07**	90.99	**92.5**	**91.65**	**98.05**	93.3
YOLOv5	Darknet	60.11	11.1	33.33	5.41	20.9	26.17	42.18	4.16	25.5	0.27	0.29	14.48	93.61	88.33	85.98	53.16	80.27
YOLOv5*	Darknet	**60.32**	28.21	36.34	6.75	21.91	30.7	**45.25**	11.85	30.96	2.92	7.17	19.63	**94.41**	91.76	91.12	97.28	**93.64**

Strategy	City Street						Highway						Hong Kong				
	Car	Bus	Truck	Motor	Bike	Average	Car	Bus	Truck	Motor	Bike	Average	Car	Taxi	Bus	Minibus	Average
None	56.36	36.46	31.85	5.66	21.54	30.77	43.01	17.67	32.8	0.75	9.09	20.66	88.91	90.64	88.35	93.07	90.24
Ours	56.19	**42.83**	35.31	**9.97**	**29.58**	**34.78**	44.18	22.97	35.01	**6.08**	**17.39**	**25.12**	**89.72**	92.44	**91.36**	**97.42**	**92.74**
ACSL[17]	**57.13**	35.28	34.34	5.71	23.53	31.2	**44.68**	19.14	**36.09**	0.31	10.77	22.2	88.23	**92.98**	87.76	94.82	90.95
EQL[11]	56.93	33.63	35.38	7.35	24.94	31.65	43.47	18.18	35.55	1.04	9.52	21.53	88.87	91.71	86.48	94.89	90.49
ECL[20]	56.89	36.13	**35.57**	7.29	20.27	31.23	43.67	24.2	35.22	0.71	0.14	20.79	88.98	92.91	89.24	93.9	91.26
Focal loss[16]	56.43	36.43	33.08	6.39	21.89	30.84	43.98	**23.81**	35	0.96	6.04	21.96	88.4	92.35	86.95	93.34	90.26

A	B	Car	Truck	Bus	Motor	Bike	Average
—	—	56.36	31.85	36.46	5.66	21.54	30.77
✓	—	56.32	**35.67**	40.76	**10.39**	27.02	34.03
—	✓	**56.96**	35.41	37.61	8.04	25.68	32.74
✓	✓	56.19	35.31	**42.83**	9.97	**29.58**	**34.78**

FIGURE 4. Visual comparison of the results from the baseline model using each method for nighttime vehicle detection.

approach can detect the truck from the crowd scene even though its size is very small, whereas the other detectors just ignored it. In Figure 4(b), the minibus is ignored or misclassified as a car or bus, while the detector based on our method can precisely detect it. In addition, for the region with complex lights, the original detector and those based on ACSL or EQL generated false-positive results, which can be effectively avoided by our proposed method.

Ablation Studies

To validate the contributions of our proposed anomaly detection scheme for tail classes and the reconstructed loss function for background proposals in nighttime vehicle detection, we carried out several ablation experiments using the baseline faster R-CNN model with a backbone of ResNet50. For convenience, we conducted experiments only on the city street scenes of the BDD dataset, which are larger and more complex than other scenes.

The ablation experiment results are listed in Table 1(c). Compared with the original detector without the proposed strategies, the anomaly detection and reconstructed loss $\mathrm{Loss}_{\mathrm{background}}$ approaches can both contribute—to varying degrees—to the improvement of nighttime vehicle detection. Of these two strategies, anomaly detection works more effectively for tail classes, while resulting in a slight loss of detection accuracy for head classes such as cars. $\mathrm{Loss}_{\mathrm{background}}$ can improve detection precision on all the categories, from which it can be concluded that the detector learns to be more sensitive to disturbance from other types of lights and reduces the number of false-positive proposals through the reconstructed loss for background proposals. Furthermore, compared with other state-of-the-art methods for long-tail object detection, the detector with a single strategy (i.e., anomaly detection or $\mathrm{Loss}_{\mathrm{background}}$ alone) still achieved higher improvements, which confirms the effectiveness of our proposed method for vehicle detection in low-light traffic scenarios.

After combining these two strategies, the detector's best performances exhibit some variations across several categories. Nevertheless, the detector with these two strategies achieves the highest overall performance.

CONCLUSION

In this article, we unveiled the long-tail distribution problem in nighttime vehicle detection. To tackle this, we proposed combining anomaly detection with proposal classification in the stream of vehicle detection, which could greatly improve the sensitivity of detection models to vehicles belonging to tail classes. In view of the disturbance from extraneous lights in low-light traffic scenarios, we further reconstructed the loss function for background proposals. Validation on the BDD and HKNVD datasets proved that our proposed methods could greatly improve the baseline model and outperform other state-of-the-art solutions for long-tail object detection in low-light traffic scenes.

REFERENCES

1. Y. Li, D. Wu, Q. Chen, J. Lee, and K. Long, "Exploring transition durations of rear-end collisions based on vehicle trajectory data: A survival modeling approach," *Accident Anal. Prevention*, vol. 159, Sep. 2021, Art. no. 106271, doi: 10.1016/j.aap.2021.106271.
2. P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, and A. Mouzakitis, "A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6206–6221, Jul. 2022, doi: 10.1109/TITS.2021.3084396.
3. A. Bell, T. Mantecón, C. Díaz, C. R. del Blanco, F. Jaureguizar, and N. García, "A novel system for nighttime vehicle detection based on foveal classifiers with real-time performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5421–5433, Jun. 2022, doi: 10.1109/TITS.2021.3053863.
4. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, vol. 28, pp. 91–99.
5. G. Jocher et al., 2022, "ultralytics/yolov5: v7.0 — YOLOv5 SOTA realtime instance segmentation," *Zenodo*, doi: 10.5281/zenodo.7347926.
6. H. Kuang, L. Chen, F. Gu, J. Chen, L. Chan, and H. Yan, "Combining region-of-interest extraction and image enhancement for nighttime vehicle detection," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 57–65, May/Jun. 2016, doi: 10.1109/MIS.2016.17.
7. X. Shao, C. Wei, Y. Shen, and Z. Wang, "Feature enhancement based on CycleGAN for nighttime vehicle detection," *IEEE Access*, vol. 9, pp. 849–859, 2021, doi: 10.1109/ACCESS.2020.3046498.
8. P. Tao, H. Kuang, Y. Duan, L. Zhong, and W. Qiu, "BITPNet: Unsupervised bio-inspired two-path network for nighttime traffic image enhancement," *IEEE Access*, vol. 8, pp. 164,737–164,746, 2020, doi: 10.1109/ACCESS.2020.3022393.
9. F. Yu et al., "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2636–2645.

10. C. Feng, Y. Zhong, and W. Huang, "Exploring classification equilibrium in long-tailed object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3417–3426.
11. J. Tan et al., "Equalization loss for long-tailed object recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11,662–11,671.
12. B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9719–9728.
13. A. Gupta, P. Dollar, and R. Girshick, "LVIS: A dataset for large vocabulary instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5356–5364.
14. C. Drummond et al., "C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling," in *Proc. Workshop Learn. Imbalanced Datasets II*, Washington, DC, USA, 2003, vol. 11, pp. 1–8.
15. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, vol. 26, pp. 3111–3119.
16. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
17. T. Wang, Y. Zhu, C. Zhao, W. Zeng, J. Wang, and M. Tang, "Adaptive class suppression loss for long-tail object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3103–3112.
18. J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "UnitBox: An advanced object detection network," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 516–520.
19. S. Jiang, H. Qin, B. Zhang, and J. Zheng, "Optimized loss functions for object detection and application on nighttime vehicle detection," *Proc. Inst. Mech. Engineers D, J. Automobile Eng.*, vol. 236, no. 7, pp. 1568–1578, 2022, doi: 10.1177/09544070211036366.
20. J. H. Cho and P. Krähenbühl, "Long-tail detection with effective class-margins," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer-Verlag, 2022, pp. 698–714, doi: 10.1007/978-3-031-20074-8_40.

HOUWANG ZHANG is pursuing his Ph.D. degree in electrical engineering with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong, 999077, China. His research interests include deep learning, image processing, and intelligent transportation systems. Zhang received his master's degree in control science and engineering from the School of Automation, China University of Geosciences. He is a Graduate Student Member of IEEE. Contact him at houwang.zhang@my.cityu.edu.hk.

LEANNE LAI HANG CHAN is an associate professor of electrical engineering at City University of Hong Kong, Hong Kong, 999077, China. Her research interests include retinal prosthetics, neurostimulation, and computer vision. Chan received her Ph.D. degree in biomedical engineering from the University of Southern California. She is a Senior Member of IEEE. Contact her at leanne.chan@cityu.edu.hk.