This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

Article Type: General Submission

Overcoming the Challenges of Long-tail Distribution in Night-time Vehicle Detection

Houwang Zhang & Leanne Lai Hang Chan, *Department of Electrical Engineering, City University of Hong Kong, Kowloon, Hong Kong*

Abstract—As a basic task of the intelligent transportation system, night-time vehicle detection is associated with many challenges. Existing methods usually ignore significant challenges arising from the imbalanced class distribution between vehicles, which always leads to poor detection for vehicles belonging to tail classes. By analyzing existing solutions for long-tail object detection and considering the complex and diverse characteristics of night-time traffic scenarios, we propose an enhanced detection approach based on anomaly detection. In addition, to tackle disturbance from complex lights, we re-construct the loss function for background proposals, thus allowing the detector to pay more attention to hard-classified proposals and to 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 night-time vehicle detection and improve the robustness in complex environments.

According to related research, most fatal vehicle accidents over recent decades have been caused by rear-end collisions [1]. Hence, the concept of the intelligent traffic system (ITS) 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 ITS, vehicle detection is at the root of the whole system, and a reliable and effective method of vehicle detection 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, techniques for vehicle detection have become mature and are now common in daily applications, such as the widely used Faster R-CNN [4] and YOLO [5]. However, current methods for vehicle detection are essentially designed for daytime scenes, whereas most 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 is increased in night-time scenarios [7]. Hence, there is an urgent need for effective and accurate methods of night-time vehicle detection.

XXXX-XXX © 2021 IEEE
Digital Object Identifier 10.1109/XXX.0000.0000000

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

Fig. 1 shows the distribution of vehicles in the widely used Berkeley Deep Drive (BDD) dataset [9]. Here, we have selected traffic images captured under low-light conditions in two main types of scenes: city street and highway. As we can observe, 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 and motors, constitute a relatively

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

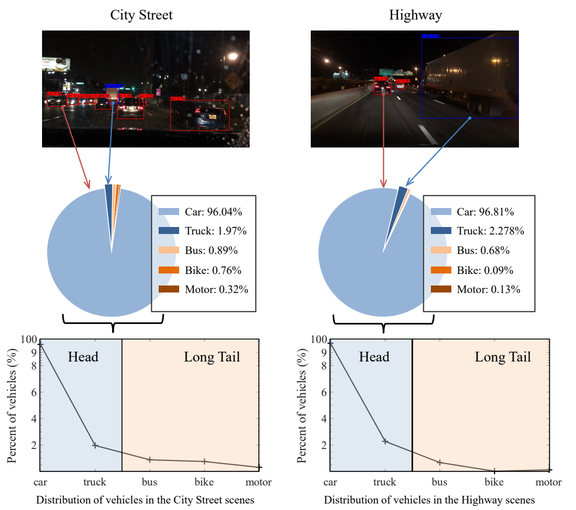


FIGURE 1: Illustration of the distribution of vehicles in two scenes: city streets and highways.

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

Eliminating the long-tail distribution in the task of general object detection has been a hot topic for researchers. Current research works can be mainly divided into two categories: re-sampling strategies and re-weighting methods. For the re-sampling strategies, most works regard the problem of long-tail distribution as an influence of the imbalance of sampling in the training batch. Based on this assumption, re-sampling strategies usually involve the design of special sampling techniques [11]. However, this kind of approach will inevitably result in distortion of the original distribution, which impairs the representation learning and causes problems such as over-fitting or under-fitting [12]. Hence, in recent years, researchers have turned to re-weighting methods, in which the loss function is re-construct to deal with the problem of imbalance between different classes, and which can effectively weaken the influence of the long-tail distribution on natural scene datasets such as LVIS [13].

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

To overcome these challenges, we enhance the detection model with an anomaly detection scheme to address the issue of night-time vehicle detection. To weaken the influence of extraneous lights in the image, we re-construct the loss function for the background proposals based on IoU (Intersection over Union). Our key contributions can be summarised as follows:

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

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

Related work

Re-sampling methods

To tackle the problem of long-tail distribution, re-sampling methods are widely used in the early stage. Re-sampling methods usually over-sample additional training data for the tail classes or under-sample data for the head classes, in order to generate more balanced samples [11]. Although re-sampling 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] re-sample the data in the tail classes during the training process using different sampling frequencies for different categories, while there will always be a high potential risk of over-fitting. Unlike over-sampling methods, under-sampling methods such as that in [14] aim to remove samples from the head classes to make the overall training samples uniform and balanced. However, under-sampling methods will always cause a loss of overall performance.

Re-weighting methods

Re-weighting methods rely on the basic idea of assigning different weights to different samples belonging to special categories. Re-weighting 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 re-weight 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-classified and hard-classified categories. Equalization loss [11] (EQL) simply ignores the gradients of samples in the tail classes to avoid the proposals for the tail classes being over-suppressed 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.

Re-weighting methods such as Focal loss, EQL and ACSL can effectively enhance general tasks of object detection with the 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 between foreground samples and ignore the imbalance of the background samples in night-time traffic scenes. In this paper, we propose a more general framework based on anomaly detection that does not rely on prior category knowledge like frequency distribution or occupation ratio for each category. Besides, We also reconstruct the loss function for the background samples to weaken the disturbance from extraneous lights for night-time 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 Fig. 2. The generated proposals 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 proposals may contain head classes or tail classes; we regard the proposals belonging 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 the data distribution before training.

As discussed in the previous section, most re-weighting methods for long-tail object detection rely on the estimation of the frequencies for all categories. However, for night-time traffic scenes, the distribution of vehicles and pedestrians varies heavily at different times and locations, such as the situations during rush hour and late night time 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, since it guides the detection model to learn a global classification for head and tail classes rather than several categories.

The $Loss_{foreground}$ 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 $Loss_{anomaly}$ for each foreground proposal x_f as follows:

$$Loss_{anomaly}(x_f) = -log(\hat{p}_i)$$
$$\hat{p}_i = \begin{cases} p_i, & if\ y_i \in Tail\ Classes \\ 1 - p_i, & otherwise \end{cases} \quad (1)$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

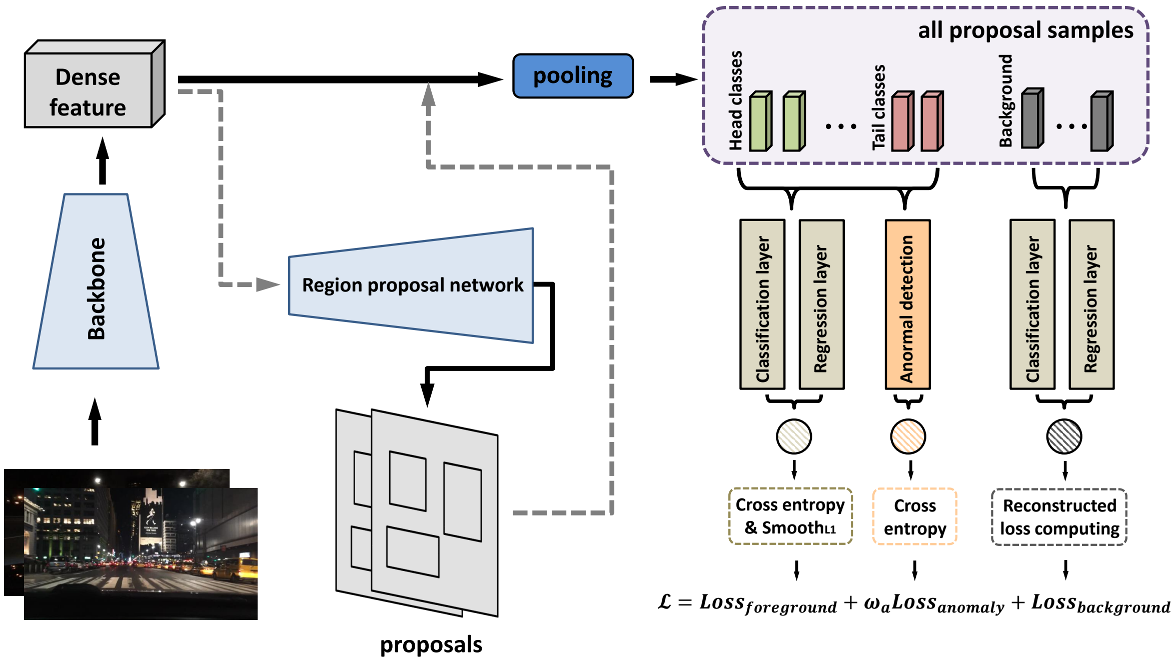


FIGURE 2: Overview of the proposed pipeline based on Faster R-CNN for night-time vehicle detection.

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.

Re-constructed loss function

For the detection of vehicles under night-time traffic scenes, the headlights or taillights of vehicles are the most salient features and form the core target of traditional night-time 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, etc., which increases the probability of false or missing detections. As shown in Fig. 3. (a-b), areas with lights such as street lamps or reflections appear very similar to vehicles in night-time traffic scenarios, and may be misclassified as vehicles.

To decrease such false positive results, we need to analyze the background proposals generated in the training process. As shown in Fig. 3. (c), there are two main kinds of background proposals that may be misclassified at the training stage. Case 1 (such as the yellow boxes) contains regions with interference lights (especially paired lights like street lamps), which are easily misclassified as vehicles. Case 2 (such as green boxes) includes regions that capture only certain parts of vehicles, due to the setting of the IoU threshold (such as 0.3 or 0.5), they will still be determined 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. Distinct from Case 1, Case 2 manages to capture a segment of the vehicle, signifying that 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 misclassified proposals that closely resemble Case 1, effectively addressing this issue.

To distinguish 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 larger value of the IoU means the generated proposal is closer to Case 2, while proposals with a lower IoU should be 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 follows:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

(a) (b) (c)

FIGURE 3: Illustration for the false detection cases. (a-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.

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

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, y_i represents the corresponding ground-truth label of x_b. α and β are used as adjustment parameters for the weights of the background proposals. ω_l is a parameter for the localization loss $loss_{loc}$.

In Eq. (2), we also compute the localization loss $loss_{loc}$ for the 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 the 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 follows:

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

where ω_a is the weight for the anomaly detection of tail classes.

Experiments

Dataset

For experiments, we adopted the widely used BDD dataset [9] and Hong Kong night-time vehicle detection (HK) dataset [6] for comparisons with related schemes. For BDD dataset, we chose images captured under night-time conditions, and divided them into two main scenes: highways and city streets. Specifically, for city street scenes, 10000 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 HK dataset, 500 images and 336 images are chosen for training and testing, respectively.

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

Implementation details

To evaluate the effectiveness of our proposed anomaly detection scheme and the reconstructed loss function for night-time vehicle detection, we used the classical detectors Faster RCNN and YOLOv5 as the baseline

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

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 weights in the loss functions, ω_a for the anomaly detection was set to three, and ω_l for the location of background proposals was set to five. The parameters for adjusting the weights of classification of background proposals, α and β, were set to 1.75 and 1.2, respectively.

For the highway and city street scenes in BDD dataset, we set training epochs to 50 and 25, respectively. For HK dataset, we set training epochs as 200. When testing, Non-Maximum Suppression with an IoU threshold of 0.5 was adopted to remove overlapped proposals. Other hyper-parameters such as the size of anchors, the number of foreground proposals and background proposals, etc. were set to the default values.

Comparison with the baseline

The results of comparison with the baseline models are given in Table. 1. (a). We used the mean average precision (mAP) 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, Faster R-CNN or YOLOv5 with the proposed strategies achieved great improvements on all of highway, city street and Hong Kong scenes. Generally speaking, the detector's performance on HK dataset is better than on the two BDD sub-datasets, which can be attributable to the imbalance is not so severe and annotations are not so dense and elaborate on HK dataset.

Moreover, by observing Table. 1. (a), although performance on the head classes (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 performance in the tail classes (especially bikes and minibuses) has improved significantly. For instance, on Faster RCNN with ResNet50, our approach boosted the average precision of the tail classes by about 37% (29.58 vs. 21.54) on the city street scenes for bikes, 29% (22.97 vs. 17.67) on the highway scenes for buses, 4% (97.42 vs. 93.07) on the Hong Kong scenes 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 ECL [20], using the same baseline model Faster R-CNN with ResNet50. For these methods, vehicles belonging to head classes or tail classes were defined with the same rule stated in the previous section, 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 the vehicle detection performance for head classes such as cars, their overall AP was about three points lower than our methods. As stated in the previous section, most existing methods for long-tail object detection can be effectively applied to natural datasets, but do not perform so well in night-time traffic scenarios. Take the city street scenes for example, our approach outperformed ACSL by 3.58 points (34.78 vs. 31.20), EQL by 3.13 points (34.78 vs. 31.65), ECL by 3.55 points (34.78 vs. 31.23), and focal loss by 3.94 points (34.78 vs. 30.84), which further proved the effectiveness and superiority of our proposed methods.

We also presented a visual comparison of these methods in Fig. 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 Fig. 4. (a), the detector based on our approach can detect the truck from the image even though its size is very small, whereas the other detectors just ignored it. In Fig. 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.

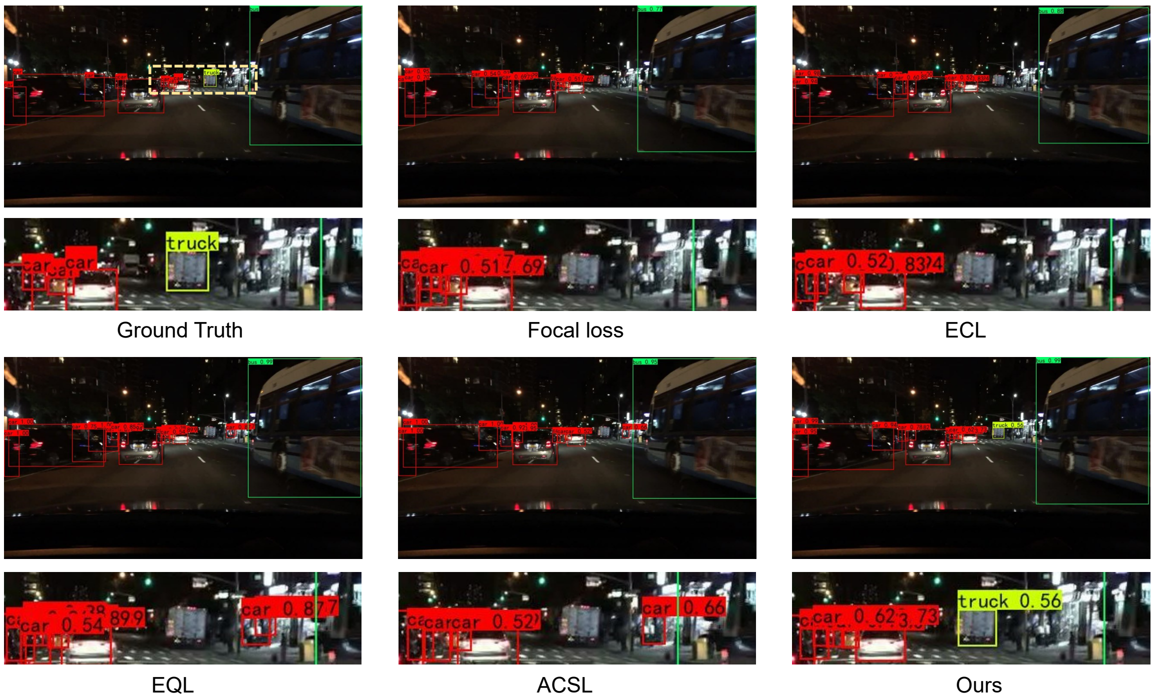
Ablation studies

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

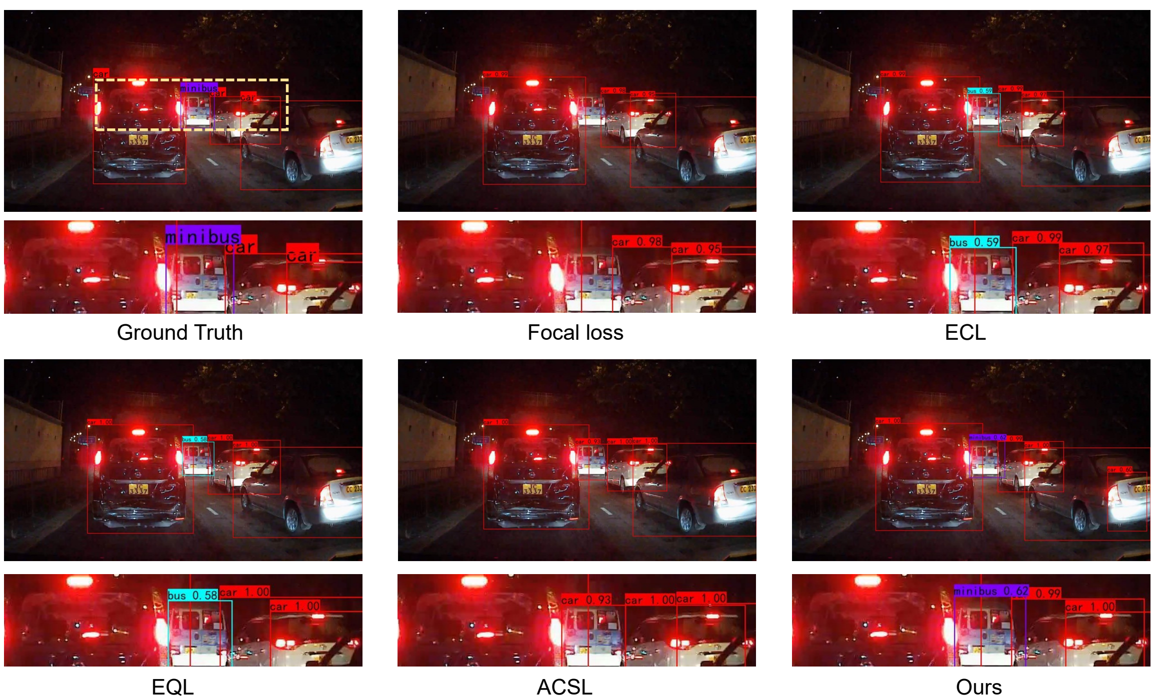
The results of these ablation experiments are shown in Table. 1. (c). Compared with the original detector without the proposed strategies, the anomaly detection and reconstructed loss $Loss_{background}$ approaches can both contribute to a varying degree to the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628



(a) Visualization comparison on the city street scenes of BDD dataset



(b) Visualization comparison on the HK dataset

FIGURE 4: Visual comparison of the results from the baseline model using each method for night-time vehicle detection.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

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	R50	56.36	36.46	31.85	5.66	21.54	30.37	43.01	17.67	32.80	0.75	9.09	20.66	88.91	90.64	88.35	93.07	90.24
Faster R-CNN*	R50	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	R100	55.10	35.61	35.10	7.14	20.64	30.71	44.09	23.37	**35.03**	0.85	4.17	21.50	89.14	91.40	88.84	93.43	90.70
Faster R-CNN*	R101	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	**92.50**	**91.65**	**98.05**	93.30
YOLOv5	Darknet	60.11	11.10	33.33	5.41	20.90	26.17	42.18	4.16	25.50	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	30.70	**45.25**	11.85	30.96	2.92	7.17	19.63	**94.41**	91.76	91.12	97.28	**93.64**

(a) Comparison of results from the baseline models, Faster R-CNN* or YOLOv5* means the original detector with anomaly detection and reconstructed loss function $Loss_{background}$.

Stategy	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	32.80	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	31.20	**44.68**	19.14	**36.09**	0.31	10.77	22.20	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	24.20	35.22	0.71	0.14	20.79	88.98	92.91	89.24	93.90	91.26
Focal loss [16]	56.43	36.43	33.08	6.39	21.89	30.84	43.98	**23.81**	35.00	0.96	6.04	21.96	88.40	92.35	86.95	93.34	90.26

(b) Comparison of results from the proposed methods with other strategies using the baseline model.

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**

(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.

TABLE 1: Quantitative comparison results on the BDD dataset and HK dataset.

improvement of night-time vehicle detection. Of these two strategies, anomaly detection works more effectively for the tail classes, while resulting in a slight loss of the detection accuracy for the head classes such as cars. $Loss_{background}$ 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 re-constructed loss for the 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 $Loss_{background}$ alone) still achieved higher improvements, which confirms the effectiveness of our proposed method for night-time vehicle detection in low-light traffic scenarios.

After combining these two strategies, the detector's best performances exhibit some variations across

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

several categories. Nevertheless, the detector with these two strategies achieves the highest overall performance.

CONCLUSION

In this paper, we unveiled the long-tail distribution problem in night-time vehicle detection. To tackle this, we proposed to combine 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 re-constructed the loss function for the background proposals. Validation on the BDD dataset and HK dataset proved that our proposed methods could greatly improve the baseline model and outperform other state-of-the-art solutions for long-tail object detection on 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 Analysis & Prevention*, vol. 159, p. 106271, 2021.
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 Transactions on Intelligent Transportation Systems*, 2021.
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 Transactions on Intelligent Transportation Systems*, 2021.
4. S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.
5. G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, J. Fang, Z. Yifu, C. Wong, D. Montes *et al.*, “ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation,” *Zenodo*, 2022.
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 Intelligent systems*, vol. 31, no. 3, pp. 57–65, 2016.
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, 2020.
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.
9. F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, “Bdd100k: A diverse driving dataset for heterogeneous multitask learning,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
10. C. Feng, Y. Zhong, and W. Huang, “Exploring classification equilibrium in long-tailed object detection,” in *Proceedings of the IEEE/CVF International conference on computer vision*, 2021, pp. 3417–3426.
11. J. Tan, C. Wang, B. Li, Q. Li, W. Ouyang, C. Yin, and J. Yan, “Equalization loss for long-tailed object recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9719–9728.
13. A. Gupta, P. Dollar, and R. Girshick, “Lvis: A dataset for large vocabulary instance segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5356–5364.
14. C. Drummond, R. C. Holte *et al.*, “C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling,” in *Workshop on learning from imbalanced datasets II*, vol. 11. Citeseer, 2003, 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,” *Advances in neural information processing systems*, vol. 26, 2013.
16. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3103–3112.
18. J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, “Unitbox: An advanced object detection network,” in *Proceedings of the 24th ACM international conference on 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,” *Proceedings of the Institution of Mechanical Engineers, Part D: Journal*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3350628

of Automobile Engineering, vol. 236, no. 7, pp. 1568–1578, 2022.

20. J. Hyun Cho and P. Krähenbühl, “Long-tail detection with effective class-margins,” in *European Conference on Computer Vision*. Springer, 2022, pp. 698–714.

Houwang Zhang is currently pursuing the Ph.D. degree in electrical engineering with the Department of Electrical Engineering, City University of Hong Kong. His current research interests include deep learning, image processing, and intelligent transportation systems. Contact him at houwang.zhang@my.cityu.edu.hk.

Leanne Lai Hang Chan is currently an Associate Professor of Electrical Engineering with the City University of Hong Kong. Her current research interests include retinal prosthetics, neurostimulation, neural recording, and computer vision. Contact her at leanne.chan@cityu.edu.hk.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/