IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 May 2024, accepted 16 May 2024, date of publication 31 May 2024, date of current version 13 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3407955

RESEARCH ARTICLE

RAOD: A Benchmark for Road Abandoned Object Detection From Video Surveillance

YAJUN XU, HUAN HU, XIAOYA ZHU, YIBING NAN, KAI WANG, ZHAOXIANG LIU, AND SHIGUO LIAN, (Member, IEEE)
AI Innovation Center, China Unicom, Beijing 100013, China
Unicom Digital Technology, China Unicom, Beijing 100013, China

Corresponding authors: Yibing Nan (nanyb5@chinaunicom.cn) and Shiguo Lian (liansg@chinaunicom.cn)

ABSTRACT Road abandoned objects are potential safety hazards in modern traffic transport, especially in highway scenes. Promptly detecting such obstacles on the road is of great significance for driving safety and Intelligent Transportation Systems (ITS). Current research primarily focuses on developing diverse road anomaly detection approaches to discriminate the unknown objects regarded as abandoned ones. However, previous efforts have been largely inadequate due to the absence of abundant datasets. In addition, prevailing benchmarks mainly provide data pertinent to autonomous driving, which might not effectively generalize to highway scenarios owing to camera perspective and scope limitations. To address these challenges, we introduce a large-scale Road Abandoned Object Detection (RAOD) benchmark derived from video surveillance. First, we collect abundant real-world video clips containing various potential abandoned object categories on the road from our commercial ITS, then assemble a road abandoned object dataset comprising 557 video sequences and 18,953 images with pixel-level manual annotations. Second, we conduct exhaustive evaluation experiments employing a range of baseline models from mainstream algorithms on our dataset to illustrate the performance of different approaches. Third, we propose a novel image segmentation framework based on an area-aware attention mechanism. Experimental results reveal that our method outperforms the UNet-based model by nearly 9% in terms of dice score. Our dataset represents the most extensive open-source resource dedicated to road abandoned object detection, accessible publicly at https://github.com/yajunbaby/A-Benchmark-for-Road-Abandoned-Object-Detection-from-Video-Surveillance.

INDEX TERMS Road abandoned object detection, intelligent transportation system, video surveillance, area-aware attention mechanism.

I. INTRODUCTION

Road abandoned objects are treated as hazardous obstacles, threatening safe driving, especially for vehicles traveling at high speeds. It is crucial to accurately identify these small but potentially dangerous obstacles within the current framework of Intelligent Transportation Systems (ITS). Although remarkable achievements have been made in the field of road anomaly detection by both academic and industrial circles in the past few years, there are still numerous difficulties in widely applying these research results to highway scenarios. One of the key bottlenecks is that currently available datasets for road abandoned objects on highways are inadequate in numbers and homogeneous in categories, lacking sufficient diversity and complexity.

Currently, most existing related datasets [1], [2], [3], [4] are predefined closed-set of known categories including persons, vehicles, trees and other surroundings while abandoned objects are rarely included. Especially for highway environments, publicly available datasets are almost entirely related to autonomous driving, but datasets that focus on abandoned objects on highways are extremely scarce. Due to the limitations of camera perspectives and ranges, these autonomous driving datasets cannot be effectively applied to highway scenarios. Moreover, benchmarks related to road abandoned objects like LostAndFound [5], and

The associate editor coordinating the review of this manuscript and approving it for publication was Alessia Saggese.

RoadObstacle21 [6], suffer from limited diversity and small scales, and can not suit the complexities of highway environments. What's more, the focus on the road anomaly segmentation task methods [7], [8], [9], [10], [11], [12], [13], [14], heavily relying on mainstream datasets. This results in poor performance in identifying road obstacles in new environments.

To address the challenges, we have built a large-scale dataset of road abandoned objects collected from high-resolution videos captured by Closed-Circuit Television (CCTV) footage. In comparison to the existing datasets listed in Table 1, our dataset has the following characteristics: 1) It contains 557 video sequences with over 500 abandoned object sequences as positive samples, and 18,891 images with pixel-level annotations, which outperforms other datasets in terms of volume by a large margin. 2) It covers 10 categories of common abandoned objects on the highways, which can make it widely used for research and applications. 3) Given the special highway scenarios we focused on and various CCTV-footage perspectives, our dataset prominently enriches the data of the scenarios.

To effectively evaluate road abandoned object recognition algorithms in highway scenes and demonstrate several significant baselines with our dataset, we conduct comprehensive experiments in three related mainstream directions and present detailed comparisons and analyses.

The characteristics of road abandoned objects, namely their small size and high diversity, notably undermine the performance of baseline algorithms. In mainstream works, researchers tend to care about the object itself no matter in category or location. Small objects are often misclassified in a complicated background, or the background is wrongly identified as part of an object due to similar appearances. To alleviate background interferences, most works [15], [16] devise an extra model to locate road region at first, which significantly increases computational costs and time consumption while getting poor performance within our dataset. To address this issue, we propose a novel image segmentation framework that adaptively tends to relevant regions considering the surroundings. Our framework follows an encoder-decoder architecture with soft dice loss [17] and plugs into an elaborately designed area-aware module with heatmap loss to facilitate joint training of the segmentation network. Experiments show that the accuracy of small road abandoned objects can be largely improved and our method can generalize well in the proposed dataset.

In Summary, the contributions of this paper are as follows:

We have built a large-scale dataset of road abandoned objects, the first one publicly available from video surveillance, to advance the field of modern traffic transportation.

We conduct evaluation experiments on mainstream algorithms to construct practical baselines on our dataset.

We introduce an area-aware attention-based segmentation framework, which can notably improve the recognition accuracy of road abandoned objects in response to the challenges posed by their small size and high diversity.

TABLE 1. Comparison of publicly available datasets and our dataset for road abandoned objects in the number of images, positive videos and types.

Dataset	#images	#positives	#types
LostAndFound [5]	2,104	112	9
RoadObstacle21 [6]	327	-	-
Ours	**18,891**	**502**	**10**

TABLE 2. Overview of road abandoned object detection methods.

Method Category	Pros	Cons	Examples
Generic Small Object Recognition	Effective for small objects	May struggle with extreme sizes, cluttered scenes	SSD, YOLO series, etc.
Abandoned Object Detection	Utilizes background subtraction, tracks objects	Prone to false positives, complex backgrounds	Gaussian models [18], etc.
Road Anomaly Detection	Diverse approaches	Insufficient labeled datasets, varying definitions of 'anomalous'	Bayesian learning [7], etc.

II. RELATED WORK

In this section, we first briefly overview existing datasets relevant to road abandoned objects. Then we introduce some methods of road abandoned object detection from three directions, specifically encompassing generic small object recognition, abandoned object recognition and road anomaly detection. Table 2 showcases the classic algorithms in the three research directions aforementioned, along with their respective advantages and disadvantages.

A. DATASETS AND BENCHMARKS

In recent years, public datasets and benchmarks have been proposed. The WildDash benchmark [19] has unified the label strategy of several mainstream datasets including Cityscapes [1], Apollo Scape [2], KITTI [3], Mapillary [4], concentrating on the challenging scenes being full of dangerous factors while there is no clearly annotated abnormal object on roads. LostAndFound [5] is the first proposed dataset related to road abandoned objects, which collects some small accidentally appeared obstacles from various streets in German, comprising 112 videos and 2,102 frames. Additionally, the unknown category object is regarded as out-of-distribution sample sets to explore uncertainty estimation research in Fishyscapes benchmark [20]. CAOS benchmark [21] filtered out two categories as anomaly objects from BDD100K datasets [22]. Recently, a novel dataset, RoadObstacle21 constructed in segmentation benchmark [6], consisting of 321 images with pixel-level annotation information, which views the road as interests of the region to narrow the range of detection. However, these relevant benchmarks and datasets have defects in some aspects, e.g., small scale, inadequate diversity, and single background from the perspective of autonomous driving.

B. ROAD ABANDONED OBJECT DETECTION METHODS

1) GENERIC SMALL OBJECT RECOGNITION

Given that the road abandoned object usually occupies a small part of the regions in the picture, we predefine its category as a specific class within the closed sets in the method of detecting or segmenting small objects. Typical object detection frameworks, such as Faster-RCNN, YOLO series, and SSD series, have acquired impressive performance in detecting small objects. These works [23], [24] devised novel methods based on FasterRCNN, focusing on the stage of anchor proposals generation and pyramid feature extraction respectively. YOLO series-based methods [25], [26], [27], [28], [29], [30] are designed for small object detection have achieved quite effective performance. Researchers also adopt UNet-based segmentation frameworks to segment small objects in the field of medical imaging. UNet [31] is the representative literature in the segmentation algorithm. Many prompted works [32], [33], [34], [35], [36] are proposed in recent years. For example, UNet++ [34] is a newer and more robust network architecture that builds upon UNet by incorporating a series of nested dense skip connections. UNet3+ [35] is designed with full-scale connections to acquire more details from medical images, enhancing diagnostic precision.

2) ABANDONED OBJECT DETECTION

Abandoned object detection from video surveillance mainly relied on the strategy of splitting the foreground object from the background to trace the potential lost items [18], [37], [38]. Reference [18] uses Gaussian background modeling to locate foreground objects and further judge possible abandoned objects by introducing edge statistics into the elimination of noise interference. Three mixed Gaussian models are adopted to remove complicated backgrounds in [37], which takes advantage of travel tracks, target size and temporal to find abandoned ones. In highway scenes, the abandoned objects may be the ones lost by vehicles or persons. The process of these anomalies on roads is usually incomplete due to the constant camera calibration and movement and it is hard to robustly make analysis according to the method of continuous inter-frame difference. In summary, various abandoned objects in complicated environments make it difficult to segment by traditional methods.

3) ROAD ANOMALY DETECTION

The types of road abandoned objects tend to be diverse and unpredictable. Currently, there are few large-scale and high-quality relevant datasets, therefore, it is hard to obtain a closed set with abundant data samples. The majority of related works concentrate on road anomaly detection tasks, which can be roughly divided into three groups: category-based methods, uncertainty prediction-based methods, and generation model-based methods.

For category-based methods, most research focuses on the samples that are absent from training data yet yield high confidence through the classification network. The exploration of these hard samples is equivalent to employing detection methods for abnormal road objects. Reference [7] designs a simple and effective pipeline for detecting samples out of the data distribution. Meanwhile, [39] introduces a novel optimization technique, applying similar adversarial training to improve the learning ability of abnormal samples, and [8] proposes an effective evaluation index as a standard for out of distribution detection. Similarly, the techniques and methods of the classification task are also suitable for the pixel-level image segmentation task [9], [20].

For uncertainty prediction-based methods, abnormal objects are directly categorized as uncertain samples, and the probability score of abnormal objects is estimated by designing the method of uncertainty estimation. Initially, such methods primarily rely on the distribution of the Bayesian learning [40], [41]. Over time, there have been relevant improvements and advancements made to these methods. For example, [42] put forward the solution to the problem of computing complexity. To model the uncertainty estimation at the pixel level, [43] uses Bayesian inference to decode networks in image segmentation tasks. Methods [11] and [44] deeply explore the uncertainty in Bayesian learning, including aleatoric and epistemic uncertainty, to more accurately detect abnormal objects. Some methods directly introduce additional data sets to expose uncertain samples and significantly improve the learning of uncertain samples.

For generation model-based methods, the methods of this group are to reconstruct and generate images containing unknown objects as training data of unknown-class models based on image synthesis. The synthesized data is usually low-quality, thus many works exploit different strategies to improve the ability of data synthesis. Reference [45] reconstructs the roadblock data based on the RBM network, [12] introduces uncertainty maps to help the model improve the discrimination of input data and synthetic data, and [13] improves the prediction ability of unknown categories by identifying the image regions with poor synthetic effect.

III. ROAD ABANDONED OBJECT DETECTION (RAOD) DATASET

In this section, we first present the proposed dataset, which contains 557 videos and 10 types of abandoned objects. Then we describe how to construct the dataset and make an analysis by statistic-based comparison in detail.

A. DATASET CONSTRUCTION

To collect high-quality videos, we have accumulated video clips of real road abandoned object accidents as an initial data source from our commercial traffic intelligent analysis platform over nearly two years. These filtered video datasets cover more than 70 distinct highway scenes across various periods and distances. To obtain abundant samples with different appearances of abandoned objects in pictures, we extract video frames at dense intervals. To facilitate

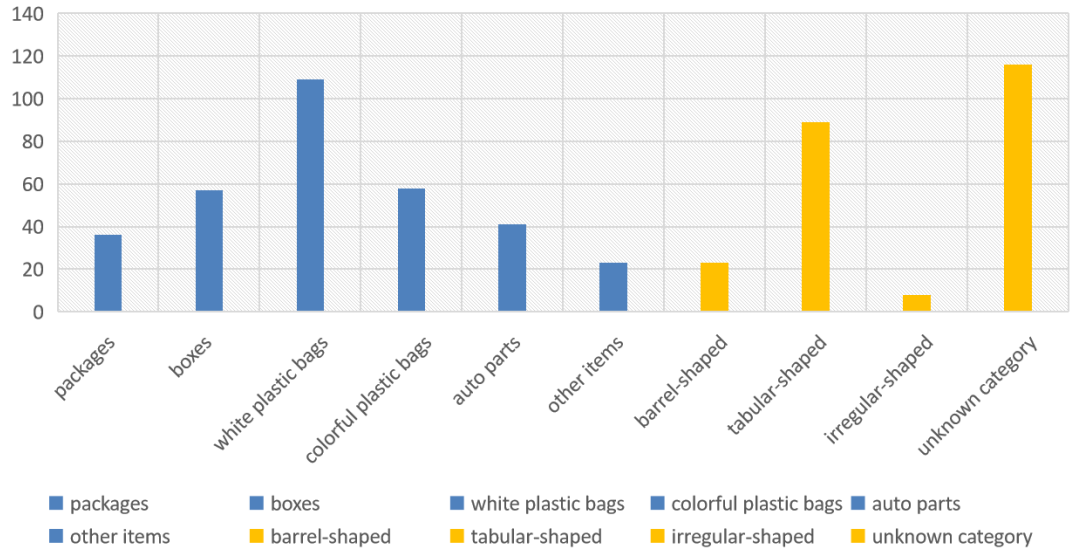

FIGURE 1. The count distribution in different categories. The left blue ones represent recognizable objects, and the right orange ones represent illegible objects.

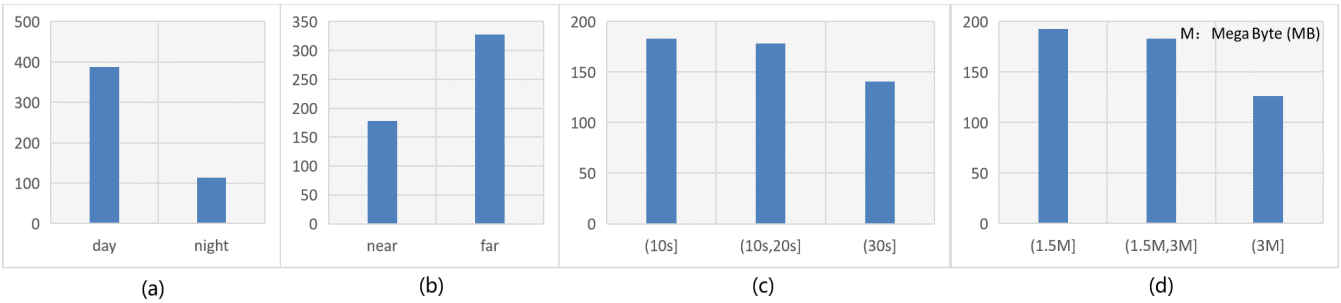

FIGURE 2. The count distributions of positive videos including time between day and night (a), the distance between near and far (b), playback duration (c) and video size (d).

experiments with various algorithms, each extracted frame is human-annotated at the pixel level. Specifically, due to the extreme diversity and uncertainty of road-abandoned objects, almost any object, except for people and vehicles, can be abandoned on the road. We cannot exhaustively list all possible categories of abandoned objects and annotate them individually. Moreover, detailed categorization of each abandoned object would lead to the class imbalance of data, which poses difficulties for the model's training and learning. To avoid these issues, we opted to categorize all abandoned objects as a unified category of "abandoned".This simplifies the problem, enabling the model to focus more on learning the common features of abandoned objects rather than getting caught up in subtle differences among various categories.

B. DATASET STATISTICS

To support the study of road abandoned object detection, the details about the proposed dataset will be presented in the subsequent statistical analysis.

1) SIZE

Our dataset consists of 502 abandoned object videos, 55 normal videos, and over 18,891 abandoned object frames. Figure 3 illustrates the count distributions of positive videos: time between day and night (a), the distance between near and far (b), playback duration (c) and video sizes (d). We observe that the distribution of the number of videos is relatively balanced in all aspects.

2) TYPES

Our dataset possesses characteristics of small size and high diversity, and it encompasses 10 classifications depicted in Figure 1. The left blue ones represent recognizable objects, and the right orange ones belong to illegible objects. For the branch of recognizable objects, the most prevalent one is white plastic bags, followed by colorful plastic bags, boxes, auto parts, packages and other items. In particular, some objects with small quantities, e.g., chairs, tree branches, and animals are classified as other items. For the other illegible branch, the most common one is an unknown category with blurred appearances, and the remaining objects can be grouped into barrel-shaped, tabular-shaped, and irregular-shaped. Moreover, an overview of sample images of our abandoned object dataset is shown in Figure 3, with the upper left corner presenting the enlarged result of the abandoned object area within the dotted box.

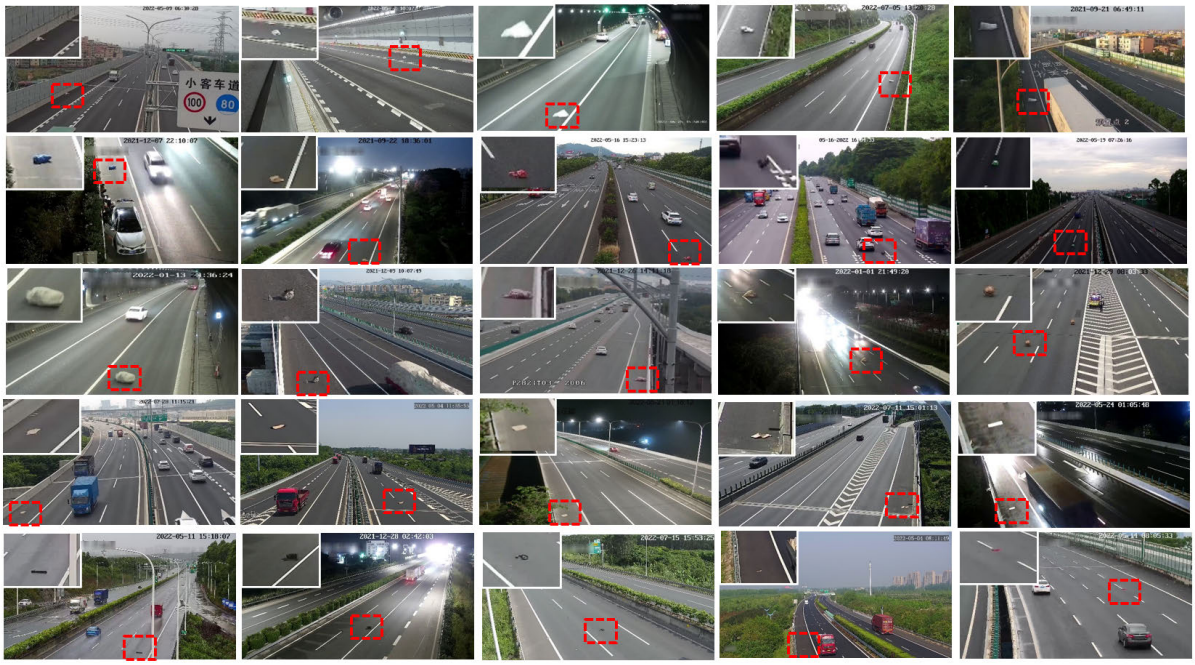

FIGURE 3. An overview of sample images of our abandoned object dataset. The red dotted boxes locate road abandoned objects and the upper left corner is the enlarging results.

IV. THE PROPOSED IMAGE SEGMENTATION METHOD

The pipeline of our proposed method is illustrated in Figure 4. The highway image is fed into the Encoder-Decoder Module based on UNet to derive feature maps with high-level and low-level semantics. Then our area-aware module processes these feature maps to generate area-aware weighted feature maps by adaptively focusing on important regions related to abandoned objects. Finally, we perform pixel-level classifiers to get a binary segmentation mask.

A. AREA-AWARE MODULE

As depicted in Figure 4, the area-aware module receives semantic feature maps and outputs mixed ones by fusing the weighted parts, which perform adaptive awareness in relevant regions. The middle K branch indicates that feature maps of reduced dimension can be generated as weight maps responding to the active score of each predefined division block in two-dimensional image space. For this branch, we first reduce the channel number of feature maps into the same output category from the segmentation network. Then we utilize the Global Average Pool function to aggregate feature blocks in two-dimensional space with a predefined number. Continuing with Conv with 1×1 and Sigmoid function, we can acquire an area-aware heat map consisting of probability score block by block. In the down Q branch, we perform matrix multiplication with heat maps and feature maps of reduced dimension. The mixed feature maps can be obtained by adding another feature map of reduced dimension in the top V branch.

B. TRAINING AND LOSS DESIGN

Our goal is to enhance relevant area response and segment abandoned objects in weighted areas. For this purpose, we designed two loss functions corresponding to two subtasks. Firstly, to tackle the challenges of small object segmentation, we employed the soft Dice loss, aiming to enable the segmentation network to pay more attention to the learning of foreground objects, thereby alleviating the issue of class imbalance. Secondly, to reduce the interference caused by similar-shaped objects in complex backgrounds, we creatively broadened the recognition scope by the information of surroundings around the abandoned objects. By jointly training the network with a specially designed heatmap loss function, the model learned to utilize the surroundings to assist in judgments. Overall, our abandoned object segmentation framework was trained in an end-to-end manner through these two complementary subtasks. The network is optimized by a loss function with two components:

$$Loss = \alpha L_{dice} + \beta L_{heatmap} \tag{1}$$

where L_{dice} is soft dice loss, $L_{heatmap}$ is binary cross-entropy loss, by which we adjust the network by paying attention to the relevant region. α and β are the weighted paramaters.

1) HEATMAP LOSS

We denote the i-th target area mask as T^i, the i-th predicted heatmap as P^i, where N is the number of 2D blocks. Our proposed heatmap loss is defined as:

$$L_{heatmap} = \frac{1}{N} \sum_{i=1}^{N} |P^i, T^i| \tag{2}$$

2) AREA-AWARE GROUND-TRUTH GENERATION

We generate area-aware ground-truth from semantic segmentation ground-truth using the red dashed box in Figure 4.

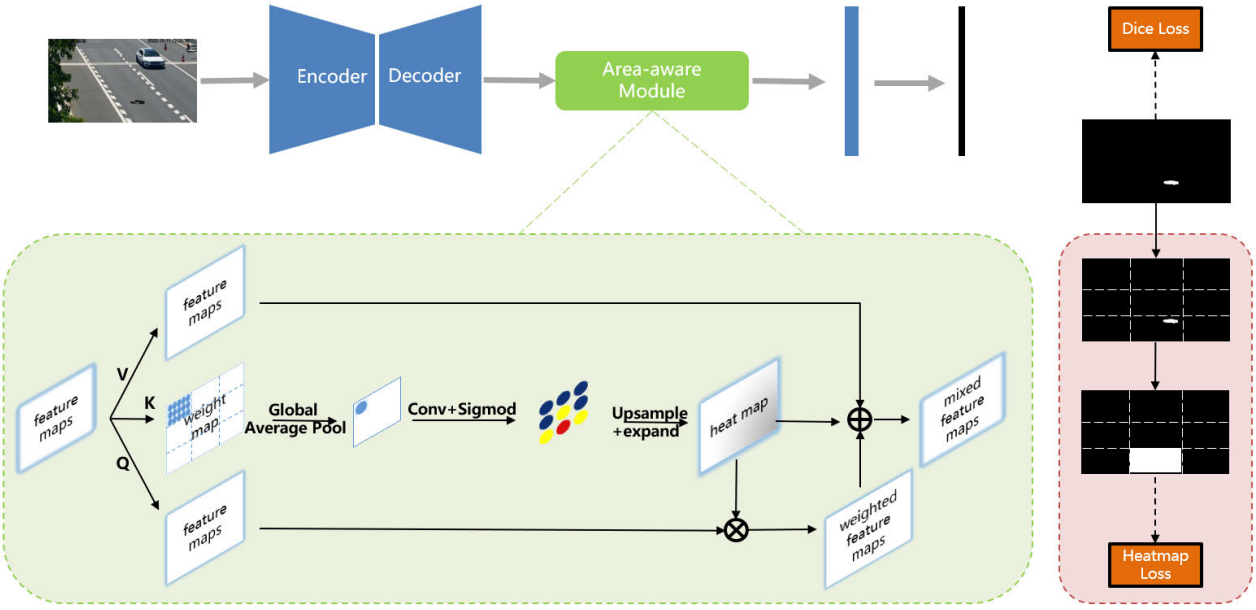

FIGURE 4. The pipeline of our proposed road abandoned object segmentation framework.

By finding the external rectangle of the annotated outline of an abandoned object, we obtain the initial regions of interest. Centered on these initial regions, we perform a two-fold box extended operation along the width and height to get a new mask. Then the mask is further split into many blocks with predefined numbers. The above binary masks are aggregated into small-size masks by performing a summation function in each block. Finally, we use the Softmax function to transform the reduced binary mask into the normalized target ground-truth.

V. EXPERIMENTS

In this section, we first select baseline models from the mainstream direction and demonstrate the result of the baseline model in our RAOD dataset. Subsequently, we exhibit exhaustive experimental details alongside our proposed segmentation framework, which is designed with area-aware attention mechanisms. Finally, we make ablation studies in each baseline separately, presenting comparison results utilizing uniform evaluation metrics.

A. EVALUATED METHODS

1) IMAGE OBJECT DETECTION

For this direction, due to the excellent performance and speed of small object detection, we choose the YOLO series model with medium parameters as the baseline model. We assess its performance with various YOLO series backbone architectures, including YOLOv5m [28], YOLOv8m [29] and YOLOv9m [30], conducting a comprehensive analysis of both accuracy and speed. In addition, to taking advantage of the surroundings of the abandoned object, we extend the bounding boxes in the width and height directions before being fed into the detection network.

TABLE 3. The comparison results of YOLO models with different parameter settings in object detection.

Model	box extend	AP	AR	mAP@.5	mAP@50:95
YOLOv5m	N	0.918	0.555	0.674	0.437
YOLOv5m	Y	0.867	0.672	0.729	0.488
YOLOv8m	N	**0.921**	0.559	0.678	0.446
YOLOv8m	Y	0.869	**0.677**	**0.733**	**0.492**
YOLOv9m	N	0.917	0.552	0.673	0.439
YOLOv9m	Y	0.864	0.668	0.726	0.489

2) IMAGE SEMANTIC SEGMENTATION

In binary segmentation tasks, UNet has been proven to be a model of landmark significance. Given that we labeled all road abandoned objects as one category when annotating segmentation objects, UNet was perfect as our benchmark model to segment road abandoned objects in images. For the choice of loss function, as background pixels in the image far exceed the ones of abandoned objects belonging to the foreground in numbers, the standard binary cross-entropy loss is not suitable for our scenario. To alleviate the imbalance between positive (road abandoned objects) and negative (background) class instances, we adopted the Dice loss function that can better handle class imbalance issues.

3) ROAD ANOMALY DETECTION

Following the mainstream works in road anomaly detection, we adopt at least one method per group mentioned in Section II-B3. For the category-based group, the maximum softmax probability (MSP) method [8] is evaluated as the

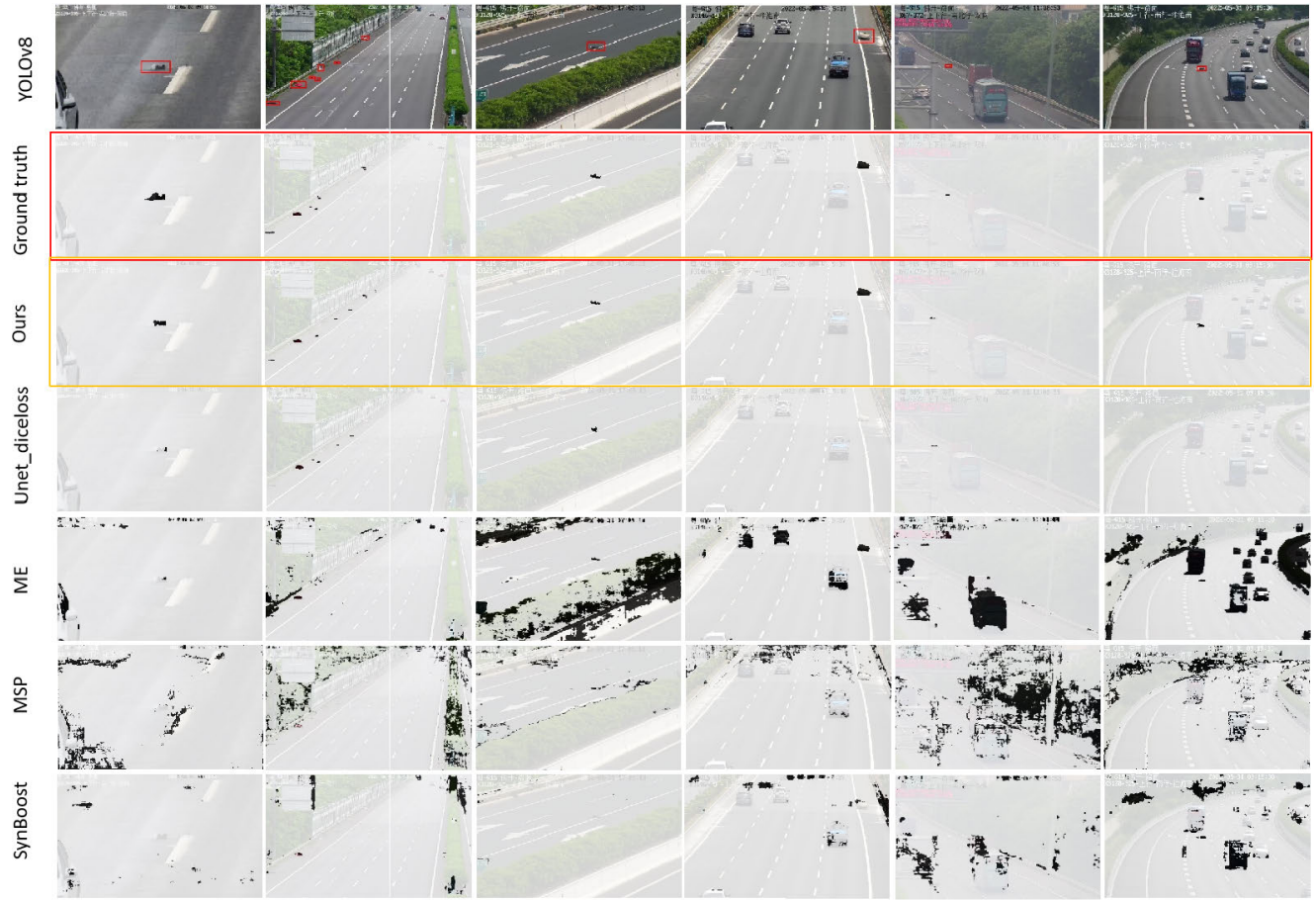

FIGURE 5. The qualitative examples of our proposed and mainstream methods. The solid red line represents the ground-truth of the semantic segmentation task and the yellow line represents the results from our proposed segmentation method.

TABLE 4. The dice score of our method with different parameters setting in semantic segmentation.

Loss type	block number	α	dice score
L1	4	0.3	0.5545
L1	4	0.5	0.6437
L2	4	0.3	0.5917
L2	8	0.3	0.5532

TABLE 5. The dice score of three mainstream methods in road anomaly detection.

Model	dice score
MSP	0.0122
ME	0.0276
SynBoost	0.0232

common baseline, of which the results correspond the output of a DNNs. For the uncertainty prediction-based group, we use the maximized entropy (ME) method [46], which is introduced to improve the prediction probability of the output of DNN. For the generation-based group, we select the SynbBoost method [12] to perform pixel-level anomaly detection in our road abandoned object datasets. Without

TABLE 6. The dice score of our method (Baseline+heatmap_loss) and other semantic segmentation methods with different loss designs in our dataset.

Model	dice score
Baseline	0.5526
Baseline+bce_loss	0.4032
Baseline+focal_loss	0.4847
Baseline+heatmap_loss	**0.6437**

training procedures in the above methods, we can obtain the prediction results by direct inference process.

B. IMPLEMENTATION DETAILS

To facilitate the analysis of different algorithms, we randomly select 5,000 images of the RAOD dataset and split them into train/test sets with 4500/500, respectively. All the experiments are conducted on Ubuntu OS with NVIDIA Tesla V100 GPU.

For the training process of YOLO models, we adopt mosaic image preprocessing technology, which joins four images performed random cropping, scaling, and rotation operations together to generate the final synthesized image, and multi-scale strategy ranging from 0.6 to 1.5. The input

resolution is 640 × 640, the mini-batch size is 4 on 2 GPUs and 150 epochs have been trained. We follow the general evaluation metrics for object detection with average precision (AP), average recall (AP) and mAP(0.5).

For the training process of UNet, the soft dice loss is applied in training work. Inspired by the extended preprocess in image object detection experiments, we design the area-aware module with heatmap loss based on the UNet network. The input image is resized to 800 × 800 for the convenience of block division operation in the area-aware module. Given the small size of road abandoned objects, we do not adopt any augmentation techniques to evaluate the segmentation performance of the model for small objects. The experimental setup: the mini-batch size is set to 2 on 1 GPU and the training duration is 250 epochs. In particular, the supervision of the area-aware module is a kind of heatmap mask, generated by the original annotation mask. The summation function is applied in each block division and we perform normalization by softmax function alongside 2D image space. We use dice scores as a metric to balance precision and sensitivity scores.

For the inferring process of road anomaly detection, we set the different threshold values to get binary images on the output of different methods by observing performance. The set values are 0.7, 0.9, and 0.8 separately for MSP, ME, and SynBoost. To make fair comparisons with the results of image semantic segmentation, the dice score is chosen as the evaluation metric.

C. COMPARISON RESULTS AND DISCUSSIONS

We undertake comprehensive evaluations across various dimensions, with quantitative results presented separately in Tables 3 through 6, corresponding to Image Object Detection, Image Semantic Segmentation, and Road Anomaly Detection, respectively. Qualitative illustrations comparing our proposed works with established methodologies are depicted in Figure 5.

1) IMAGE OBJECT DETECTION

As illustrated in Table 3, although YOLOv9 is the latest work in the YOLO series, its performance on our dataset is not as good as YOLOv5 and YOLOv8. Furthermore, no matter whether to use box extended operation, YOLOv8 is slightly higher than other methods in mAP metric. In addition, for each backbone, we observe that performing the extended operation on the bounding box can improve the mAP by nearly 4 points.

2) IMAGE SEMANTIC SEGMENTATION

In Table 4, we show different parameter configurations of our proposed model. A UNet model employing dice loss serves as the baseline, against which we compare our area-aware module trained using two heatmap loss formulations: L1 (Mean Absolute Error, MAE) and L2 (Mean Squared Error, MSE). The feature map and heatmap mask division are varied with block counts of 4 or 8, and the heatmap loss weighting during training is also manipulated. Analysis of the loss functions indicates that L2 loss generally yields superior dice scores compared to setups detailed in the initial and third rows. When the loss function and block number are held constant, equating the weights of dice and heatmap losses optimizes segmentation performance. Additionally, comparing rows three and four suggests that a block count of 4 is optimal, implying an increase in blocks may divert model attention from intricate surroundings.

To validate our heatmap loss innovation, we contrast its performance with other UNet-based models using different loss criteria (Table 6), all assessed via dice score on our proprietary dataset. The heatmap loss emerges as the superior option, outperforming alternative loss functions.

3) ROAD ANOMALY DETECTION

Regrettably, the chosen methods are not effective in our dataset, low dice scores in Table 5 demonstrate that. The major reason is that the backbone module of these evaluated methods is trained on urban street scene datasets. This indicates that our dataset is more suitable for road anomaly detection than the commonly used benchmark. More visualization results can be obtained in the last three rows in Figure 5.

VI. CONCLUSION

In this paper, we have proposed a benchmark for road abandoned object detection from video surveillance. First, we construct a large-scale abandoned object dataset including 557 video sequences from our commercial ITS and 18891 images with pixel-level mutual annotations. Second, we have successfully evaluated several mainstream methods including image object detection, image semantic segmentation, and road anomaly detection and provided essential pre-trained models on the RAOD dataset. Finally, we have proposed a novel segmentation framework with our elaborately designed area-aware module based on UNet to improve segmentation accuracy. In particular, we present the currently largest dataset for road abandoned object detection. Our dataset is publicly available at https://github.com/yajunbaby/A-Benchmark-for-Road-Abandoned-Object-Detection-from-Video-Surveillance.

REFERENCES

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.

[2] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The ApolloScape dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1–12.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[4] G. Neuhold, T. Ollmann, S. R. Buló, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5000–5009.

[5] P. Pinggera, S. Ramos, S. Gehrig, U. Franke, C. Rother, and R. Mester, "Lost and found: Detecting small road hazards for self-driving vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 1099–1106.

[6] R. Chan, K. Lis, S. Uhlemeyer, H. Blum, S. Honari, R. Siegwart, P. Fua, M. Salzmann, and M. Rottmann, "SegmentMeIfYouCan: A benchmark for anomaly segmentation," 2021, *arXiv:2104.14812*.

[7] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Advances in Neural Information Processing Systems*, vol. 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2018.

[8] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2017. [Online]. Available: https://openreview.net/forum?id=Hkg4TI9xl

[9] M. Angus, K. Czarnecki, and R. Salay, "Efficacy of pixel-level OOD detection for semantic segmentation," 2019, *arXiv:1911.02897*.

[10] A. Atanov, A. Ashukha, D. Molchanov, K. Neklyudov, and D. Vetrov, "Uncertainty estimation via stochastic batch normalization," in *Advances in Neural Networks—ISNN 2019*, H. Lu, H. Tang, and Z. Wang, Eds. Cham, Switzerland: Springer, 1007, pp. 261–269.

[11] F. K. Gustafsson, M. Danelljan, and T. B. Schon, "Evaluating scalable Bayesian deep learning methods for robust computer vision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1289–1298.

[12] G. Di Biase, H. Blum, R. Siegwart, and C. Cadena, "Pixel-wise anomaly detection in complex driving scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16913–16922.

[13] K. Lis, K. K. Nakka, P. Fua, and M. Salzmann, "Detecting the unexpected via image resynthesis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2152–2161.

[14] K. Lis, S. Honari, P. Fua, and M. Salzmann, "Detecting road obstacles by erasing them," 2020, *arXiv:2012.13633*.

[15] S. Qi and D. Yu, "Railway obstacle detection based on radar and image data fusion," *J. Phys., Conf. Ser.*, vol. 1965, no. 1, Jul. 2021, Art. no. 012141.

[16] T. Xiao, Y. Xu, and H. Yu, "Research on obstacle detection method of urban rail transit based on multisensor technology," *J. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 61–67, Jan. 2021.

[17] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[18] H. Fu, M. Xiang, H. Ma, A. Ming, and L. Liu, "Abandoned object detection in highway scene," in *Proc. 6th Int. Conf. Pervasive Comput. Appl.*, Oct. 2011, pp. 117–121.

[19] O. Zendel, K. Honauer, M. Murschitz, D. Steininger, and G. F. Domínguez, "WildDash—Creating hazard-aware benchmarks," in *Proc. ECCV*, 2018.

[20] H. Blum, P.-E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena, "FishyScapes: A benchmark for safe semantic segmentation in autonomous driving," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2403–2412.

[21] D. Hendrycks, S. Basart, M. Mazeika, M. Mostajabi, J. Steinhardt, and D. X. Song, "Scaling out-of-distribution detection for real-world settings," in *Proc. ICML*, 2022.

[22] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.

[23] Z. Liang, J. Shao, D. Zhang, and L. Gao, "Small object detection using deep feature pyramid networks," in *Advances in Multimedia Information Processing—PCM 2018*, W.-H. Cheng, T. Yamasaki, M. Wang, and C.-W. Ngo, Eds. Cham, Switzerland: Springer, 1007, pp. 554–564.

[24] C. Eggert, S. Brehm, A. Winschel, D. Zecha, and R. Lienhart, "A closer look: Small object detection in faster R-CNN," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 421–426.

[25] R. Zhang, J. Zhang, J. Gui, C. Gao, and X. Bao, "Abandoned object detection algorithm based on improved of YOLOv2 network," *J. Zhejiang Sci-Tech Univ., Natural Sci. Ed.*, 2018.

[26] A. Benjumea, I. Teeti, F. Cuzzolin, and A. Bradley, "YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles," 2021, *arXiv:2112.11798*.

[27] M. Liu, X. Wang, A. Zhou, X. Fu, Y. Ma, and C. Piao, "UAV-YOLO: Small object detection on unmanned aerial vehicle perspective," *Sensors*, vol. 20, no. 8, p. 2238, Apr. 2020, doi: 10.3390/s20082238.

[28] G. Jocher. (2020). *YOLOv5 by Ultralytics*. [Online]. Available: https://github.com/ultralytics/yolov5

[29] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLO*. [Online]. Available: https://github.com/ultralytics/ultralytics

[30] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, Eds. Cham, Switzerland: Springer, 1007, pp. 234–241.

[32] A. A. Cervera-Uribe and P. E. Méndez-Monroy, "U19-Net: A deep learning approach for obstacle detection in self-driving cars," *Soft Comput.*, vol. 26, no. 11, pp. 5195–5207, Jun. 2022, doi: 10.1007/s00500-022-06980-6.

[33] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608019302503

[34] Z. Zhou, M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), L. Maier-Hein, T. Syeda-Mahmood, Z. Taylor, Z. Lu, D. Stoyanov, A. Madabhushi, J. Tavares, J. Nascimento, M. Moradi, A. Martel, J. Papa, S. Conjeti, V. Belagiannis, H. Greenspan, G. Carneiro, and A. Bradley, Eds. Cham, Switzerland: Springer, 2018, pp. 3–11.

[35] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[36] T. Yang, Y. Zhou, L. Li, and C. Zhu, "DCU-Net: Multi-scale U-Net for brain tumor segmentation," *J. X-Ray Sci. Technol.*, vol. 28, no. 4, pp. 709–726, Aug. 2020.

[37] Y. Tian, R. S. Feris, H. Liu, A. Hampapur, and M.-T. Sun, "Robust detection of abandoned and removed objects in complex surveillance videos," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 41, no. 5, pp. 565–576, Sep. 2011.

[38] K. Lin, S.-C. Chen, C.-S. Chen, D.-T. Lin, and Y.-P. Hung, "Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1359–1370, Jul. 2015.

[39] M. Hein, M. Andriushchenko, and J. Bitterwolf, "Why ReLU networks yield high-confidence predictions far away from the training data and how to mitigate the problem," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 41–50.

[40] D. J. C. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Comput.*, vol. 4, no. 3, pp. 448–472, May 1992.

[41] R. M. Neal, *Bayesian Learning for Neural Networks*. Berlin, Germany: Springer, 1996.

[42] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn.*, Jun. 2016, pp. 1050–1059.

[43] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian SegNet: Model uncertainty in deep convolutional encoder–decoder architectures for scene understanding," 2015, *arXiv:1511.02680*.

[44] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 1–12.

[45] C. Creusot and A. Munawar, "Real-time small obstacle detection on highways using compressive RBM road reconstruction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 162–167.

[46] R. Chan, M. Rottmann, and H. Gottschalk, "Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5108–5117.

YAJUN XU received the M.S. degree from the Institute of Information Engineering, School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China.

She is currently a Software Development Engineer with the AI Innovation Center, China Unicom. Her research interests include computer vision and artificial intelligence.

HUAN HU received the B.S. degree from the School of Science, Hubei University of Technology, Hubei, China, in 2013, and the M.S. degree from the School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Sichuan, China, in 2017.

In 2017, he joined CloudMinds Technologies Inc., as an Algorithm Engineer. Since 2020, he has been a Deep Learning Algorithm Engineer with the AI Innovation Center, China Unicom Digital Technology Company Ltd., Beijing, China. His research interests include machine vision, computer vision, and deep learning.

XIAOYA ZHU received the M.S. degree from the University of Electronic Science and Technology of China.

From 2018 to 2020, she was a Deep Learning Algorithm Engineer at CloudMinds Technologies Inc. Since 2020, she has been a Software Engineer with the AI Innovation Center, China Unicom. Her research interests include deep learning, computer vision, and intelligent transportation.

YIBING NAN received the Ph.D. degree from Beijing Institute of Technology, China.

From 2016 to 2019, he was a Senior Engineer with CloudMinds Technologies Inc. Since 2019, he has been a Senior Algorithm Expert with the AI Innovation Center, China Unicom. His research interests include artificial intelligence, deep learning, computer vision, and intelligent video analysis. He is the author of some 20 refereed international papers, and held more than 50 patents.

KAI WANG received the Ph.D. degree from Nanyang Technological University, Singapore, in 2013.

Since 2019, he has been the AI Director with the AI Innovation Center, China Unicom. Before that, he was with CloudMinds Technologies Inc. and the Central Research Institute, Huawei Technologies. He has published more than 20 refereed papers on international journals and conferences, and granted more than 40 patents. His research interests include generative AI, computer vision, computer graphics, and human–computer interaction.

ZHAOXIANG LIU received the B.S. and Ph.D. degrees from the College of Information and Electrical Engineering, China Agricultural University, Beijing, China, in 2006 and 2011, respectively.

In 2011, he joined VIA Technologies Inc., Beijing. From 2012 to 2016, he was a Senior Researcher with the Central Research Institute, Huawei Technologies, Beijing. From 2016 to 2019, he was the Senior Manager of CloudMinds Technologies Inc., Beijing. Since 2019, he has been the Director of AI Research with the AI Innovation Center, China Unicom. He has published over 20 refereed papers on international journals and conferences, and hold more than 40 patents. His current research interests include artificial intelligence, computer vision, deep learning, robotics, and human–computer interaction.

SHIGUO LIAN (Member, IEEE) received the Ph.D. degree from Nanjing University of Science and Technology, China.

He was a Research Assistant with the City University of Hong Kong, in 2004. From 2005 to 2010, he was a Research Scientist with France Telecom Research and Development Beijing. From 2010 to 2016, he was a Senior Research Scientist and the Technical Director of the Huawei Central Research Institute. From 2016 to 2019, he was the Senior Director of CloudMinds Technologies Inc. Since 2019, he has been the Chief AI Scientist of China Unicom. He has authored over 100 refereed international journal articles covering topics of artificial intelligence, robotics, human–computer interface, and multimedia communication. He has authored or coedited over ten books and hold over 200 patents.

Dr. Lian is on the editor board of several refereed international journals.