

RESEARCH

Open Access



# Vehicle logo detection using an IoAverage loss on dataset VLD100K-61

Xiaohui Shi<sup>1</sup>, Shengli Ma<sup>1\*</sup>, Yang Shen<sup>1</sup>, Yankun Yang<sup>1</sup> and Zexin Tan<sup>1</sup>

\*Correspondence:  
mashli@xauat.edu.cn

<sup>1</sup> Xi'an University of Architecture and Technology, No. 13, Middle Section of Yanta Road, Beilin District, Xi'an City, China

## Abstract

Vehicle Logo Detection (VLD) is of great significance to Intelligent Transportation Systems (ITS). Although many methods have been proposed for VLD, it remains a challenging problem. To improve the VLD accuracy, an Intersection over Average (IoAverage) loss is proposed for enhancing the bounding box regression. The IoAverage loss accelerates the convergence of bounding box regression than using the Intersection over Union (IoU) loss. In the experiments, IoAverage loss has been incorporated into the state-of-the-art object detection framework YOLOV5s, namely YOLOV5s-IoAv in this paper. The advantages of the IoAverage loss are verified on the PASCAL VOC2007 datasets. The results of using the IoAverage loss show performance gains of + 15.27% mAP0.5 and + 30.87% mAP0.5:0.95 higher than that of the Complete IoU (CIoU) loss. The application of YOLOV5s-IoAv is implemented to VLD on dataset VLD100K-61. VLD100K-61 is a self-collected dataset containing 100,041 images supplied by traffic surveillance cameras in the real world from 61 categories. YOLOV5s-IoAv achieves performance gains as + 15.27% mAP0.5:0.95 for VLD than YOLOV5s-CIoU. The proposed method yields the mAP0.5 value of up to 0.992 on the dataset VLD100K-61, providing a promising solution to vehicle logo recognition applications.

**Keywords:** Vehicle logo detection, IoAverage loss, YOLOV5s-IoAv, VLD100K-61 dataset

## 1 Introduction

Intelligent transportation systems (ITS) play an important role in intelligent cities. More and more scholars are devoted to the research of ITS [1–5]. Vehicle feature recognition based on computer vision technology is one of the research fields [6–9]. Vehicle feature recognition is helpful for vehicle tracking, suspect tracking, vehicle behavior analysis, and vehicle behavior understanding [10, 11]. Furthermore, the most important feature of the vehicle is the vehicle logo.

In recent years, vehicle log detection has received extensive research attention. The main reasons are as follows. Firstly, when the license plate is used as the only vehicle identification information, it may cause the problem of incorrect identification when the license plate is blocked, removed or tampered with. The vehicle logo contains critical information that may be used to improve the robustness and reliability of vehicle identification. Secondly, recognizing the car logo might be a crucial indication in locating an illegal vehicle or identifying suspicious automobiles. Thirdly, the vehicle

logo contains vital information about the vehicle owner's consumption-ability. A statistical study of a large number of car logos might be used to forecast client consumption levels at the retail center.

However, vehicle logo detection still faces many challenges so far. On the one hand, the traditional handcrafted descriptors for manually extracting features cannot detect the vehicle logo in real traffic with high accuracy due to the requirement of the strong prior knowledge that the design of handcrafted descriptors needed. Furthermore, the generalization ability of handcrafted descriptors is weak for VLD in urban environments because of the camera placement, background clutter, and vehicle pose or orientation variations, especially in the case of rainy or snowy days. On the other hand, there is no reliable large dataset for researchers to exploit the data-driven deep learning methods. The images in some existing datasets only show the vehicle logos without any traffic background. Some datasets are composed of clear images downloaded from the internet, rather than images collected from realistic traffic shots. Traffic surveillance cameras are usually installed on urban road sections, intersections, turns and tunnels. The diversity of parking postures, traffic jams, weather and lighting conditions are constantly changing. These all pose additional challenges to VLD. The captured vehicle logo images collected from realistic traffic surveillance cameras are often more motion blurred, diverse and authentic, but data-driven method appetites for these images amazingly. In addition, the robustness of the previously proposed deep learning methods is considered unsatisfactory. In addition, the sizes of vehicle logos are generally quite small, which creates further difficulties for VLD. Detection of small objects in large images is challenging. The detection of small objects is prone to the problems of overlapping bounding boxes, omissions, and incorrect markings, which leads to a decrease in the mAP value. As a result, bounding box regression is necessary to be further optimized for object detection.

Most of the loss functions face challenges in distinguishing regression cases during training for VLD. Suppressing redundant frames during training requires a huge cost for AI agents. In this paper, we proposed a method to solve this problem by enhancing bounding box regression into learning and inference. We put forward an IoAverage loss for improving the accuracy of the bounding box for vehicle logo detection. IoAverage loss has been applied to state-of-the-art object detection YOLOV5s. This combined method is called YOLOV5s-IoAv. The proposed approach YOLOV5s-IoAv is directly performed based on the dataset consisting of frontal or rear images of vehicles. The vehicle logo regions were detected without dependence on the existence of license plates. In addition, we construct a VLD100K-61 dataset baseline and statistically analyze the advantages of this dataset compared to other existing vehicle logo datasets. The contributions of this paper include: (1) IoAverage loss for improving the accuracy of the vehicle logo detection is proposed; (2) we construct a new multi-class dataset VLD100K-61 containing 100,041 images and 105,111 objects from 61 categories, respectively.

The remainder of this paper is organized as follows: Section II is a brief review of the related work about vehicle logo detection and the bounding box regression. In Section III, the VLD100K-61 dataset is introduced in detail. In Section IV, the details of the proposed IoAverage loss introduced into YOLOV5s for logo recognition are described. Section V reports the experimental results, and Section VI gives the summary.

## 2 Related work

Vehicle logo detection has been studied by scholars using different approaches. VLD methods can be divided into two categories: keypoint-based methods and deep-learning-based methods. In this section, firstly, we will briefly review these methods. Bounding box regression directly affects the accuracy of vehicle logo detection. An appropriate loss function for bounding box regression can lead to the result that the overlap and misclassification of bounding boxes are corrected after training. What's more, it can also improve the confidence of vehicle logo detection. Secondly, we will briefly review bounding box regression.

### 2.1 Vehicle logo detection

#### 2.1.1 Keypoint-based methods

There are many keypoint-based methods for VLD including Local Binary Patterns (LBP) [12], Scale-Invariant Feature Transform (SIFT) [13–16], Speeded Robust Features (SURF) [17], Binary descriptor based on BRIEF (ORB) and Histograms of Oriented Gradients (HOG) method [18, 19]. Those methods have been studied as features to represent the vehicle logo. A multi-class Support Vector Machine (SVM) was then used to classify the regions. Even after image transformation, the feature points will remain stable, so that the image can still be correctly recognized. Keypoint-based methods were best for well-defined shapes and affine transformations. Psyllos et al. [13] presented an enhanced SIFT-based feature-matching scheme. The scheme demonstrated good performance, yielding a 94% logo recognition rate on a dataset of 1200 images with 10 classes. Haoyu Peng et al. [20] proposed a new Vehicle Logo Recognition (VLR) method based on statistical random sparse distribution (SRSD) features and multi-scale scanning for low-resolution and inferior-quality images. The results show the recognition rate of 97.21% on a dataset consisting of 3370 images with 56 classes. Llorca et al. [18] proposed a HOG + SVM framework for vehicle logo recognition. The proposed method is evaluated on a collection of 3579 vehicle logo images belonging to 27 different car manufacturers. The results indicated the recognition rate of 92.59%. Quan Sun et al. [21] proposed an improved vision-based scheme based on HOG + SVM methods. Ruilong Chen et al. [19, 22] introduced a framework based on spatial SIFT + LR (Logistic Regression) methods. The result shows a classifying precision of 99.93%. He also proposed an online image recognition framework using Cauchy prior logistic regression. As a result, accuracy reached as high as 98.80%. Sotheeswaran et al. [23] have proposed an approach for VLD using a coarse-to-fine strategy. The VLR accuracy is reported to be 86.3% on the dataset consisting of 250 images with 25 elliptical shapes of vehicle logos. Jiandong Zhao et al. [24] extracted the vehicle logo features from the HU invariant and identifies the logo using the SVM. Cross-validation (CV) methods have been introduced for optimizing SVM parameters. Grey Wolf Optimize (GWO) is used for further optimization of the kernel function. The average recognition rate is 92%. Kittikhun Meethongjan et al. [25] have provided a method based on the HOG descriptor and feature selection through two sparse scores. The method achieved a precision of 75.25%.

### 2.1.2 Deep-learning-based methods

Shuo Yang et al. [26] established a dataset known as VLD-45 (Vehicle Logo Dataset), which contained 45,000 images and 50,359 objects of 45 categories from the website. The dataset was evaluated by 6 detectors, and YOLOV4 achieved the highest mAP value of 0.847. Ye Yu et al. [27] introduced a new learning-based approach called Multilayer Pyramid Network Based on Learning (MLPNL) to extract valuable features. The scheme has been tested on the HFUT-VL and XMU datasets. The results demonstrated that the MLPNL was faster than most deep-learning-based learning methods. MLPNL requires only 511.03 s to achieve a 98.92% recognition rate on 100 training images of the HFUT-VL dataset. Under the same recognition rate, MLPNL was six times faster than the Darknet53 framework. Yongtao Yu et al. [28] proposed two networks to detect a vehicle logo with a detection rate of 0.987 and a recognition rate of 0.994. The convolutional neural network for VLR was explored. As a result of significant computing costs, the pretraining method was implemented [29]. An accuracy of 95.18% was observed on the XMU dataset of 11,500 logo pictures from 10 manufacturers. Linghua Zhou et al. [30] coupled Filter-DeblurGAN with the VL-YOLO algorithm to recognize blurred car logos. This method yielded a final value of 0.981 mAP on the dataset LOGO-17. Chun Pan et al. [31] developed a VLR technique using a Convolutional Neural Network (CNN). A comparison was made between the CNN and SIFT methods. Comparative results show that VLR based on CNN had an average accuracy rate of 8.61% greater than SIFT. CNN and Multi-Task Learning (MTL) were integrated [32]. The expanded Xiamen University VLR dataset indicated that the approach performed well, with an accuracy of 98.14%. Li Huan et al. [33] created an algorithm using the Hough transform and Deep Learning. This algorithm was built in three stages. The logo region is first located. The shapes of the logo are then detected. Finally, Deep Belief Networks were used to classify the vehicle logo (DBNs). This algorithm achieved a recognition rate of 92%. On publicly accessible vehicle logo datasets, Ruilong Chen et al. [34] suggested a capsule network is suitable for rotating and noisy pictures, and obtained the maximum accuracy of 100%. Foo Chong Soon et al. [35, 36] proposed a method for automatically searching for and optimizing the CNN architecture. The experimental outcomes reach an accuracy of 99.1% on a dataset including 13 vehicle manufacturers. This VLR approach relied on deep CNN and the whitening transformation technique was also proposed. This methodology was claimed to have a classification accuracy of 99.13%. Shuo Yang et al. [37] constructed a new dataset known as VLD-30. For VLD, the original YOLOv3 model was modified. On the dataset VLD-30, the improved YOLOV3 produced a result of 0.899 mAP. Zhongjie Huang et al. [38] combined the Faster-RCNN model with VGG-16 images on a dataset including 4000 photos of 8 distinct car logos. The mAP scored 94.33%. Ruikang Liu et al. [39] presented a VLR approach based on improved matching, restricted region extraction, and the SSFPD network, a single deep neural network based on a modified ResNeXt model and Feature Pyramid Networks. The mAP was reported to be 99.52%. Hoanh Nguyen et al. [40] exhibited a deep learning-based car logo identification system that is built on a single-shot framework with multi-scale feature fusion. The vehicle region detection network was based on ResNet-50, while the logo recognition network used the

Darknet-53 method. The mAP level was reported to be 90.5%. Junxing Zhang et al. [41] developed SS-VLD, a much enhanced single-stage approach that employed multi-scale prediction with feature fusion on the stage of up-sampling. 0.881 mAP

**Table 1** The experimental results of some state-of-the-art handcrafted descriptor-based and deep-learning-based methods on different datasets from the year 2010–2020

Author	Images	Image size	Classes	Scheme/ platform	mAP	Precision	Years	Datasets
Apostolos [13]	1200	100 × 100	10	SIFT + EMS	/	94%	2010	/
Chun Pan [31]	200	140 × 100	26	CNN	/	99.23%	2013	/
Llorca [18]	3579	640 × 480	27	HOG + SVM	/	92.59%	2013	/
Haoyu Peng [20]	3370	50 × 50	16	SRSD + MS	/	97.21%	2015	/
Yue Huang [29]	11,500	70 × 70	10	CNN + PS	/	99.07%	2015	XMU
Yizhang Xia [32]	19,780	64 × 64	15	CNN + MTL	/	98.14%	2016	Xiamen University
Ruilong Chen [22]	11,500	70 × 70	10	Spatial-SIFT + LR	/	99.93%	2016	/
Ruilong Chen [19]	10,000	70 × 70	10	CPLR	/	98.80%	2016	/
Li Huan [33]	80,000	1360 × 1024	10	DBNs	/	92%	2017	/
Sotheeswaran [23]	500	800 × 600	25	CtFS	/	86.30%	2018	/
Foo Chong [35]	1500	70 × 70	13	HPO-DCNN	/	99.10%	2018	/
Ruilong Chen [34]	1500	70 × 70	10	CN	/	100%	2018	/
Yu Ye [54]	16,000	64 × 96	80	OE-POEM	/	99.10%	2018	HFUT-VL
Foo Chong [36]	11,500	70 × 70	10	CNN + ZCA	/	99.13%	2019	XMU
Shuo Yang [37]	2010	320 × 240 – 1500 × 938	30	YOLOv3	0.899	/	2019	VLD-30
Ruikang Liu [39]	14,950	10 × 10–150 × 150	13	SSFPD + EMS	/	99.52%	2019	CVLD
Zhongjie [38]	4000	1000 × 600	8	Faster-RCNN + VGG-16	0.943	/	2019	/
Yongtao Yu [28]	75,000	/	15	CDCN	/	98.10%	2019	/
Junxing [41]	3015	938 × 668	45	Ss-VLD	0.881	/	2020	/
Shuo Yang [42]	2010	320 × 240 – 4608 × 3450	/	Faster-RCNN	0.875	/	2020	/
Nguyen [40]	8000	720 × 1160	13	Darknet-53	0.905	/	2020	/
Linghua Zhou [30]	18,089	576 × 576	17	Filter-Deblur-GAN + VL-YOLO	<b>0.981</b>	/	2020	LOGO-17
Kittikhun [25]	4000	/	40	HOG + SS	/	75.25%	2020	VLR-40
Ye Yu [27]	1500	70 × 70	10	MLPNL	/	<b>99.98%</b>	2021	
Junxing [53]	2010	320 × 240 – 1500 × 938	30	F-RCNN + VLD-C	0.911	/	2021	VLD-45-S
Shuo Yang [26]	45,000	610 × 378 – 7359 × 4422	45	YOLOV4	0.847	/	2021	VLD-45

Bold indicates the maximum value of this column

EMS Enhanced Matching Scheme, PS Pretraining Strategy, LR logistic regression, CN Capsule Network, CtFS Coarse-to-Fine Strategy, HPO-DCNN hyper-parameters optimization of deep CNN, CPLR Cauchy Prior Logistic Regression, SS Sparsity Score, CDCN Cascaded Deep Convolutional Network; Ss single state

was reported as the finding. The VLD-A, B, and C models of deep convolutional networks were proposed. On the dataset VLD-45, the combination of F-RCNN with VLD-C achieved the greatest mAP of 0.874. The summary of the literature review is shown in Table 1.

## 2.2 2.2 Bounding box regression

Although deep learning architectures have been extensively explored, the loss function for bounding box regression also plays an essential role in vehicle logo identification. The mAP value of VLD will be directly affected by bounding box regression. Inspired by the success of the data-driven technique, deep learning models based on bounding box regression have been implemented for car logo identification. Detectors are classified into three kinds. Those are one-stage [43]-[48], two-stage [49, 50], and multi-stage detectors [51, 52]. The  $\ell_n$ -norm loss functions used in bounding box regression are sensitive to scale variation. IoU loss, which is scale-invariant, was also used. To solve the problem of gradient disappearance produced by IoU loss in non-overlapping situations, generalized IoU (GIoU) loss [55] was developed. However, GIoU has drawbacks such as delayed convergence and incorrect regression. Distance-IoU (DIoU) loss was suggested for directly reducing the normalized distance between two central points of the bounding boxes to obtain accelerated convergence. More geometric parameters, such as overlap area, center point distance, and aspect ratio, were taken into account by CIoU [56]. As a result, CIoU achieves faster convergence and superior performance than DIoU. However, the CIoU did not include appropriate punishment terms. CIoU also has a fatal flaw in that when the value of IoU is less than 0.5, it degenerates into DIoU.

DIoU loss only incorporates distance[56], and can improve the performance with gains of 3.29% AP and 6.02% AP75 when IoU is used as an assessment metric. CIoU loss considers three essential geometric variables, resulting in higher performance increases of +5.67%AP and +8.95% AP75. Figure 1 depicts the different bounding box regression findings, with the green box being the best result for VLD.

Despite the fact that numerous researchers have performed studies on the identification of car logos, the following flaws remain: (1) there is no massive dataset made up of images from real-world traffic cameras. Furthermore, this collection includes a wide variety of vehicle logotypes. (2) No effective approach exists that can achieve high accuracy for vehicle logo detection on such a large real-traffic dataset.



**Fig. 1** Diversity of bounding box regression, where green box is the ground-truth box

**Table 2** The detail statistics of vld100k-61

Parameters	VLD-100 K-61
Total number/target number	100,041/105339
Average size of image (w × h, in pixels)	1262 × 725
Maximum size of image	5184 × 3456
Minimum size of image	193 × 143
Proportion of target	0.408% (+ 12.197%, -0.403%)
Average seize of target (w × h, in pixels)	64 × 47

Proportion of target refers to the ratio of the area of the vehicle logo to that of the image

### 3 VLD100K-61 dataset

VLD100K-61 dataset consists of the images provided by the Institute of Static Transportation Research of Xi’an University of Architecture and Technology. The dataset is 36.78 GB in size. These photographs were taken in 2021 along roadside parking lots, underground parking lots, and surface parking lots, mostly in the cities of Lanzhou, Longnan, and Baiyin in Gansu Province, China. The surveillance camera takes these photographs primarily between 05:00 a.m. and 02:30 a.m. the following day. Most of the photos were taken in March. The specific details of the database are shown in Table 2.

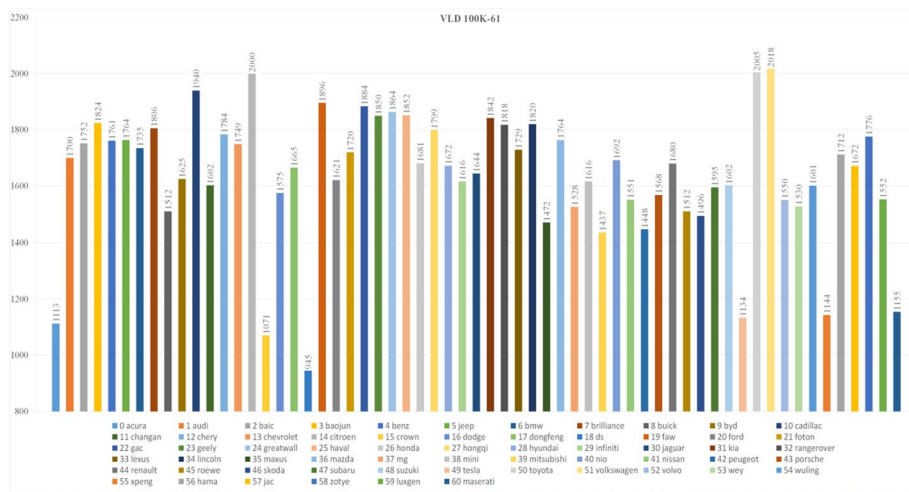
The dataset contains a total of 100,041 RGB images from 61 manufacturers with bounding box annotations. The detection of these 61 types of vehicle logos can identify more than 99% of vehicles in China. The average size of the images in this dataset is 1262 × 725 pixels. According to the total number of images and the vehicle logo classes, our dataset is named VLD100K-61. Figure 2 shows the vehicle logos included in this dataset.

Figure 3 shows the number of images that belong to different types of car logos. Among them, the least number of images is 945, and this image collection is the DS logo. The car with the VS logo has the most images. The number of images is 2018. The average number of images of each vehicle logo is 1640.



**Fig. 2** Sixty-one classes of vehicle logo

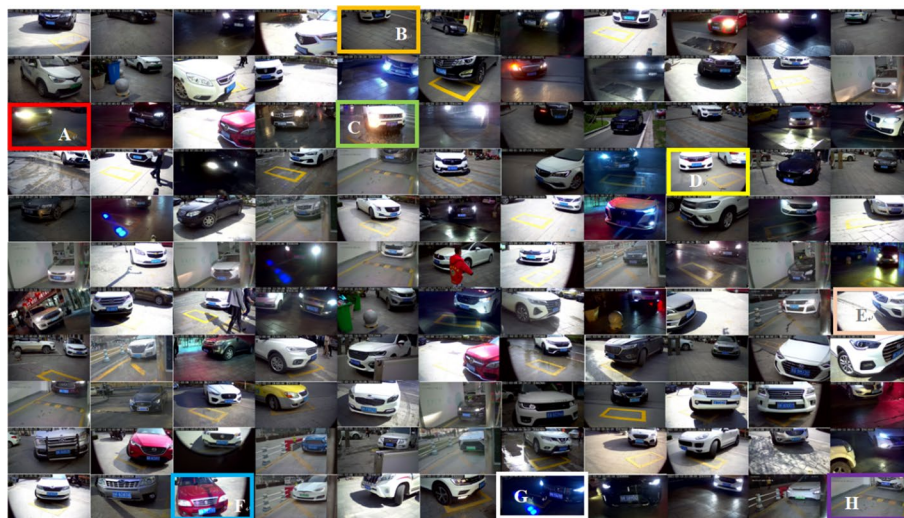




**Fig. 3** Distribution of the number of images of different vehicle logo

The photos are taken in different lighting environments and different weather conditions. There are clear photos during the day, and there are also blurred photos under the condition of car lights or street lights. The car colors in the dataset are more than 20 colors. The logo photos were taken from various angles in the dataset. Some car logos are even half-observed. Some car logos are particularly blurred at night. Some car logos are very reflective under high light, so that only half of the car logo can be captured by the surveillance camera. The dataset is very valuable for the data-driven training method.

We take some photos as samples in the data set in terms of illumination conditions, weather conditions, perspective, distortions, occlusions, image qualities, and vehicle-camera distances as shown in Fig. 4.



**Fig. 4** The examples for the dataset of VLD100K-61. **A** High beam environment, **B** half-observed, **C** rainy day, **D** multi-objects, **E** strong sunlight, **F** clear logo, **G** low illumination condition, **H** indoor parking lot



VID100K-61 encourages researchers to develop a data-driven training method for their purpose. VLD100K-61 even provides a better dataset benchmark for small target detection. Our new dataset provides several research challenges involving small-sized objects, shape deformation, and low contrast. According to the result, our dataset has very significant research value for the task of vehicle logo detection, even being valuable for small-scale object detection.

#### 4 IoAverage

IoAverage is defined as the intersection area divided by the average area of the predicted and ground-truth bounding box. IoU loss converges to poor solutions for non-overlapping cases, while GIoU loss leads to a slow convergence, especially for the boxes at horizontal and vertical orientations. DIoU loss is calculated by simultaneously considering the overlap area and the center point distance of bounding boxes. However, DIoU ignored the consistency of aspect ratios for bounding boxes, which is also an important geometric factor. CIoU loss considered the effect of consistency on aspect ratios. However, its penalty term is too mild, and the convergence speed of the loss function is too slow. Generally, the IoU-based loss can be defined as:

$$\mathcal{L} = 1 - IoU + P(\mathbf{B}, \mathbf{B}_{gt}), \quad (1)$$

where  $P(\mathbf{B}, \mathbf{B}_{gt})$  is the penalty term for predicted box  $\mathbf{B}$  and target box  $\mathbf{B}_{gt}$  by designing proper penalty terms, the CIoU loss has the ability to enhance the IoU loss. In the training phase, a bounding box  $\mathbf{B} = [x, y, w, h]^T$  is forced to approach its ground-truth box  $\mathbf{B}_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]^T$  by minimizing the loss function:

$$\min_{\theta} \sum_{\mathbf{B}_{gt}} \mathcal{L}(\mathbf{B}, \mathbf{B}_{gt} | \theta). \quad (2)$$

Considering the geometric factors for modeling regression relationships in the experiment, a loss function could take three geometric factors into account, i.e., overlap area, distance, and aspect ratio [57]. Generally, a complete loss can be defined as:

$$\mathcal{L} = S(\mathbf{B}, \mathbf{B}_{gt}) + D(\mathbf{B}, \mathbf{B}_{gt}) + V(\mathbf{B}, \mathbf{B}_{gt}), \quad (3)$$

$$S = 1 - IoU, \quad (4)$$

$$D = \frac{\rho^2(\mathbf{p}, \mathbf{p}_{gt})}{c^2}. \quad (5)$$

The CIoU [57] loss is proposed by imposing the consistency of the aspect ratio:

$$V = \frac{4}{\pi^2} \left( \arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w}{h} \right)^2, \quad (6)$$

where  $V$  measures the consistency of aspect ratio. The loss function is then defined as:

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(\mathbf{p}, \mathbf{p}_{gt})}{c^2} + \alpha V. \tag{7}$$

And the trade-off parameter  $\alpha$

$$\alpha = \begin{cases} 0 & \text{if } IoU < 0.5 \\ \frac{V}{(1-IoU)+V} & \text{if } IoU \geq 0.5 \end{cases}. \tag{8}$$

However, we found that CIoU does not have such an absolute advantage in bounding box regression. Therefore, we accelerated the IoU scheme directly and proposed the IoAverage scheme. The definition of IoAverage is as follows:

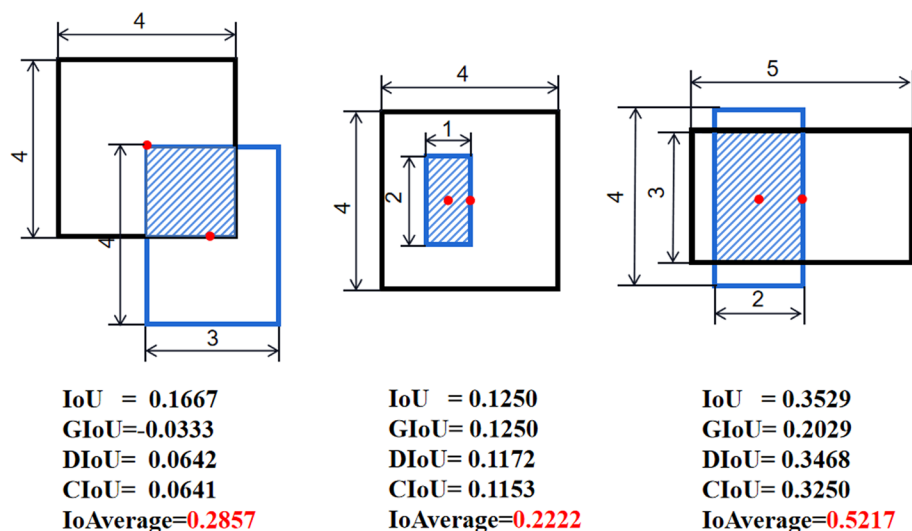
$$IoAverage = \frac{A_{intersection}}{(A_{box1} + A_{box2})/2}, \tag{9}$$

where the  $A_{intersection}$  refers to the area of the intersection. The denominator term refers to the average area of the two boxes. Although IoU is used as an evaluation index to measure whether two bounding boxes are completely overlapped, it brings spatial information redundancy. But, IoAverage can eliminate this redundant spatial information. As a result, the IoAverage scheme can anchor the bounding box more accurately. The denominator term refers to the average area of the two boxes.  $0 \leq IoAverage \leq 1$ , as shown in Fig. 5, IoAverage has the smallest loss value. When the two boxes completely overlap, the value of IoAverage is 1. When the two boxes have no intersection, the value of IoAverage is 0.

The IoAverage loss function can be defined as:

$$\mathcal{L}_{IoAv} = 1 - IoAverage. \tag{10}$$

We incorporated the IoAverage loss into YOLOV5s to evaluate the performance of IoAverage on the dataset PASCAL VOC2007. The performance of IoU, GIoU, DIoU and CIoU loss have been compared with that of the IoAverage loss in the next section. Finally, we apply the proposed scheme to vehicle logo recognition.



**Fig. 5** The results of five schemes in three cases

## 5 Experiment

### 5.1 5.1 Experimental setup

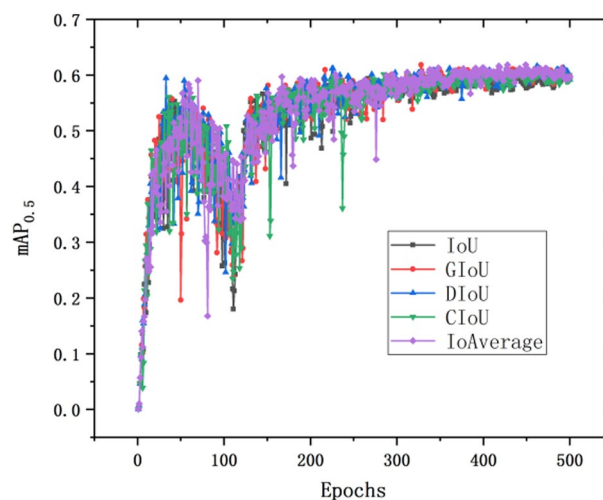
The model is trained on the training dataset with 500 epochs with a batch size of 312 examples and a learning rate of 0.001. We take mean Average Precision (mAP) as the evaluation index, which is the most commonly used index in image evaluation. Two GPUs with model Tesla V100-SXM2-32 GB are used for calculation. For VOC2007, there are 2501 images in the training set, 2510 images in the validation set, and 4952 images in the test set. The ratio of the number of images in the training set, validation set and test set is 6:2:2 for VLD100K-61. Split the 100,041 images into 60,049 images in the training set, 19,988 images in the validation set, and 20,004 images in the test set.

### 5.2 Experimental results on VOC2007

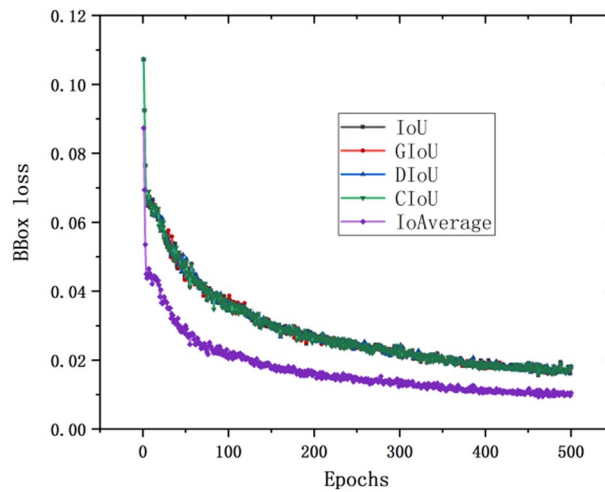
The research results show that the IoAverage loss has greater advantages in object detection than other IoU-based losses, especially in bounding box regression. Figure 6 shows that the result of the IoAverage loss has a slight improvement for mAP0.5, but it has greatly reduced the bounding box loss as shown in Fig. 7.

Table 3 analyzes the minimum bounding box loss obtained by using different strategies after training for 500 epochs. The results show that the proposed IoAverage scheme can reduce the bounding box loss by about 40% compared to the IoU-based schemes. Compared to the IoU scheme, the value of the bounding box is reduced by 39.86%, and compared to the DIoU scheme, the value of the bounding box is reduced by 41.67% as shown in Table 3.

Table 4 shows the results obtained by testing different schemes on the VOC test dataset. The research results show that compared to the IoU loss, the use of the IoAverage loss increases the value of mAP0.5 by 7.191%, which is much greater than that of the other IoU-based loss. The advantage is particularly shown in the improvement of the value of mAP0.5<sub>0.95</sub>. Using IoAverage loss improves the value of mAP0.5<sub>0.95</sub> by 30.87% compared to the value obtained by using IoU loss. The increase in the value of mAP0.5<sub>0.95</sub> by less than 0.792% using IoU-based loss.



**Fig. 6** The variation of mAP0.5 value with epochs using different schemes on the training dataset



**Fig. 7** The variation of bounding box loss value with epochs using different schemes on the training dataset

**Table 3** The value of bbox loss obtained by different schemes after training 500 epochs

Value/scheme	$\mathcal{L}_{IoU}$	$\mathcal{L}_{GIoU}$	$\mathcal{L}_{DIoU}$	$\mathcal{L}_{CIoU}$	$\mathcal{L}_{IoAverage}$
BBox loss	0.01556	0.01602	0.01607	0.01577	<b>0.009357</b>
Relative reduction %	<b>39.86%</b>	<b>41.59%</b>	<b>41.77%</b>	<b>41.67%</b>	–

Bold indicates the results obtained by calculating the BBox loss using the method proposed in this paper

Relative reduction is calculated based on the bbox loss calculated by the IoAverage scheme

**Table 4** Comparison between the performance of yolov5s trained using its own loss ( $\mathcal{L}_{IoU}$ ) as well as  $\mathcal{L}_{GIoU}$ ,  $\mathcal{L}_{DIoU}$ ,  $\mathcal{L}_{CIoU}$  and proposed scheme  $\mathcal{L}_{IoAverage}$  losses. The results are reported on the test dataset of pascal voc 2007

Loss/evaluation	mAP0.5	mAP0.5 <sub>.95</sub>
$\mathcal{L}_{IoU}$	0.598	0.379
$\mathcal{L}_{GIoU}$	0.601	0.382
Relative improv. %	0.502%	0.792%
$\mathcal{L}_{DIoU}$	0.599	0.379
Relative improv. %	0.167%	0.0%
$\mathcal{L}_{CIoU}$	0.596	0.38
Relative improv. %	–0.334%	0.264%
$\mathcal{L}_{IoAverage}$	<b>0.641</b>	<b>0.496</b>
Relative improv. %	<b>7.191%</b>	<b>30.87%</b>

Bold indicates the training results obtained using the method proposed in this paper

Considering that CIoU is the state-of-the-art deep models of object detection in YOLOV5, in order to verify the advantages of IoAverage loss, object detection is performed on the test dataset. We compared the results of object detection using CIoU and IoAverage loss. The results show that the IoAverage can achieve a better bounding box effect and high confidence value. As shown in Fig. 8, the main advantages of using IoAverage loss performance are that (1) the bounding box is more accurate, (2) obtaining higher confidence, (3) correcting the wrong classification, (4) adding the missing bounding box, (5) removing the overlapping bounding boxes.

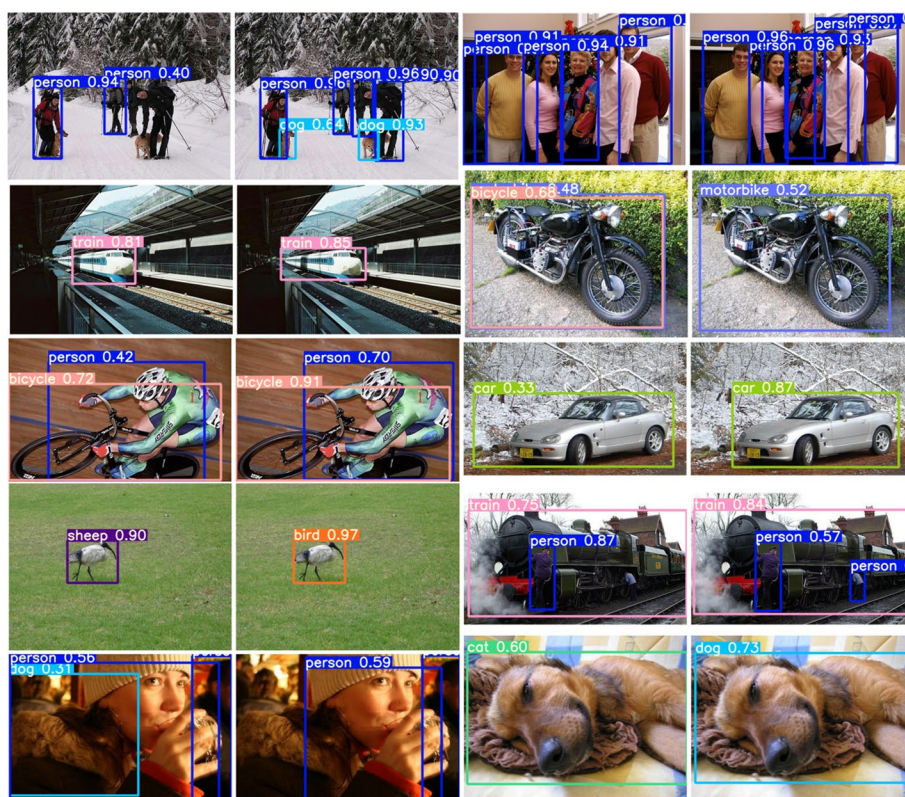


Fig. 8 Detection results using CloU loss and IoAverage loss on the VOC2007 dataset

### 5.3 Experimental results on VLD100K-61

We applied the proposed IoAverage loss to the vehicle logos detection on the VDL100k-61 dataset. The vehicle logo detection training finally took 53 h when two GPUs were used. As shown in Table 5, compared with the IoU loss, the value of the mAP0.5 calculated by the IoAverage loss has a slight increase of 0.1%. The main reason is that VLD100K-61 is a large and diverse dataset. The IoAverage loss has already received a high score of 0.992 of the mAP0.5.

The value of mAP0.5<sub>.0.95</sub> has greater potential for improvement. Compared with the CloU loss, the IoAverage loss increased the value of mAP0.5<sub>.0.95</sub> by 15.27%, and finally achieved a high score of 0.868 mAP0.5<sub>.0.95</sub>. IoAverage loss has significantly improved the value of mAP0.5<sub>.0.95</sub>, which is consistent with the performance on the VOC2007 dataset.

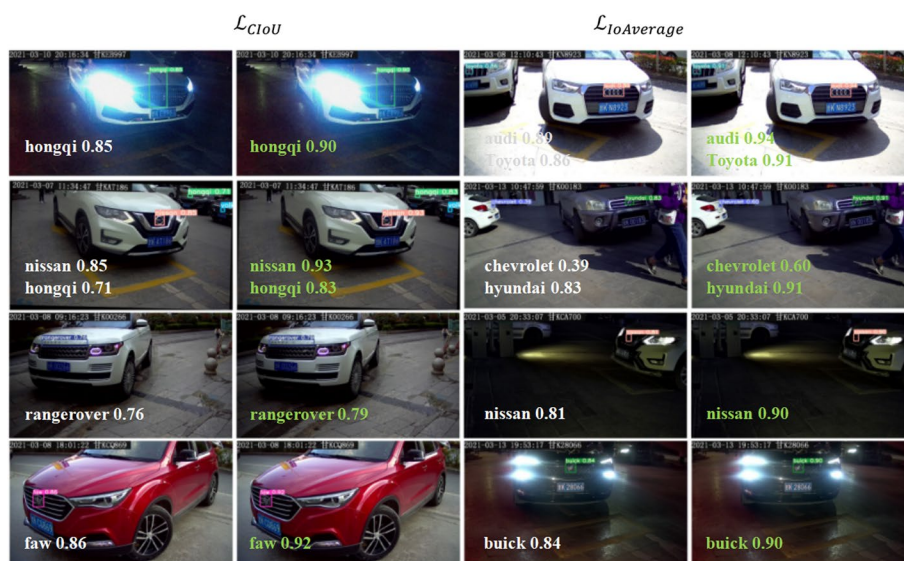
Figure 9 shows the comparison of the results of vehicle logo detection in different conditions, such as parking in the daytime, parking at midnight, multi-targets image, reflection of car headlights at night, and cars parked diagonally. The results show that the

**Table 5** Comparison between the performance of yolov5s trained using  $\mathcal{L}_{CloU}$  loss and proposed loss. The results are reported  $\mathcal{L}_{IoAverage}$  on the test dataset of vld100k-61

Loss/evaluation	P	R	mAP0.5	mAP0.5 <sub>.0.95</sub>
$\mathcal{L}_{CloU}$	0.992	0.985	0.991	0.753
$\mathcal{L}_{IoAverage}$	<b>0.993</b>	<b>0.983</b>	<b>0.992</b>	<b>0.868</b>
Relative improv. %	<b>0.1%</b>	<b>-0.2%</b>	<b>0.1%</b>	<b>15.27%</b>

Bold indicates the experimental comparison results on yolov5s using the method proposed in this paper





**Fig. 9** Detection results using CloU loss and IoAverage loss on the VLD100k-61 dataset

IoAverage loss incorporated into YOLOV5s can be used to identify vehicles’ logos with higher accuracy and confidence.

### 6 Summary

Smart cities are the future trend of urban development. The detection of vehicle logos is the basic requirement of a smart city. Statistics on vehicle logos can be used for many purposes, such as analyzing the vehicle market and guiding parking lot services. Firstly, we constructed a comprehensive vehicle logo dataset, 100,041 images with 61 classes, namely the VLD100k-61 dataset, which consists of images taken by surveillance cameras in real traffic. The average image size is 1262\*725 pixels. The images are obtained under various environments, which improves the robustness of the dataset. We also release the benchmark vehicle logo image VLD100K-61 dataset for the research community. Secondly, to accelerate the convergence of bounding box regression, we proposed an IoAverage loss. The IoAverage loss incorporated into YOLOV5s, named YOLOV5s-IoAv, can improve the accuracy of VLD. The advantages of the IoAverage loss are verified on the VOC2007 dataset. The mAP0.5 and mAP0.5,0.95 are increased by 15.27% and + 30.87%, respectively, using YOLOV5s-IoAv, which is higher than the original YOLOV5s combined with CIoU loss. Finally, we apply the YOLOV5s-IoAv to VLD based on the VID100k-61 dataset. The mAP0.5 value of the VLD is increased to 0.992. In addition, the mAP0.5,0.95 value of the VLD increased by 15.27% compared to the CIoU loss. The improvement of vehicle logo recognition accuracy lays the foundation for the construction of smart cities. For future work, we will continue to research vehicle feature recognition and detection contributing to intelligent transportation systems.

#### Acknowledgements

We would like to thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

#### Author contributions

All authors participated in the design of the analytics, performance measures, experiments, and writing of the manuscript. All authors read and approved the final manuscript.



**Funding**

This work is supported by the Special Scientific Research Project of Education Department of Shaanxi Province, under Grant 19JK0472.

**Availability of data and materials**

The dataset source codes are available at <https://github.com/mashli/datasets/VLD100K-61>

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 17 June 2022 Accepted: 27 March 2023

Published online: 05 April 2023

**References**

1. S. Wan, X. Xu, T. Wang, Z. Gu, An intelligent video analysis method for abnormal event detection in intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* **22**(7), 4487–4495 (2021)
2. R. Zhang, A. Ishikawa, W. Wang, B. Striner, O.K. Tonguz, Using Reinforcement learning with partial vehicle detection for intelligent traffic signal control. *IEEE Trans. Intell. Transp. Syst.* **22**(1), 404–415 (2021)
3. F. Zhu, Y. Lv, Y. Chen, X. Wang, G. Xiong, F.-Y. Wang, Parallel transportation systems: toward iot-enabled smart urban traffic control and management. *IEEE Trans. Intell. Transp. Syst.* **21**(10), 4063–4071 (2020)
4. L. Zhu, F.R. Yu, Y. Wang, B. Ning, T. Tang, Big data analytics in intelligent transportation systems: a survey. *IEEE Trans. Intell. Transp. Syst.* **20**(1), 383–398 (2019)
5. Z. Lu, G. Qu, Z. Liu, A survey on recent advances in vehicular network security, trust, and privacy. *IEEE Trans. Intell. Transp. Syst.* **20**(2), 760–776 (2019)
6. R. Theagarajan, N.S. Thakoor, B. Bhanu, Physical features and deep learning-based appearance features for vehicle classification from rear view videos. *IEEE Trans. Intell. Transp. Syst.* **21**(3), 1096–1108 (2020)
7. F. Chen, W. Hu, A video-based method with strong-robustness for vehicle detection and classification based on static appearance features and motion features. *IEEE Access* **9**, 13083–13098 (2021)
8. X. Shao, C. Wei, Y. Shen, Z. Wang, Feature enhancement based on cyclegan for nighttime vehicle detection. *IEEE Access* **9**, 849–859 (2021)
9. Z. Wang, J. Fang, X. Dai, H. Zhang, L. Vlacic, Intelligent vehicle self-localization based on double-layer features and multilayer LIDAR. *IEEE Trans. Intell. Veh.* **5**(4), 616–625 (2020)
10. L. Fridman et al., MIT advanced vehicle technology study: large-scale naturalistic driving study of driver behavior and interaction with automation. *IEEE Access* **7**, 102021–102038 (2019)
11. W. Wang, A. Ramesh, J. Zhu, J. Li, D. Zhao, Clustering of driving encounter scenarios using connected vehicle trajectories. *IEEE Trans. Intell. Veh.* **5**(3), 485–496 (2020)
12. A. Satpathy, X. Jiang, H. Eng, LBP-based edge-texture features for object recognition. *IEEE Trans. Image Process.* **23**(5), 1953–1964 (2014)
13. A.P. Psyllos, C.E. Anagnostopoulos, E. Kayafas, Vehicle logo recognition using a SIFT-Based enhanced matching scheme. *IEEE Trans. Intell. Transp. Syst.* **11**(2), 322–328 (2010)
14. A. Psyllos, C. Anagnostopoulos and E. Kayafas, M-SIFT: a new method for Vehicle Logo Recognition, 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012), 2012, pp. 261–266.
15. Q. Gu, J. Yang, G. Cui, L. Kong, H. Zheng, R. Klette, Multi-scale vehicle logo recognition by directional dense SIFT flow parsing. *ICIP* **2016**, 3827–3831 (2016)
16. D. G. Lowe, Object recognition from local scale-invariant features, Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, pp. 1150–1157 vol.2.
17. T. Senthilkumar, S.N. Sivanandam, Logo classification of vehicles using SURF based on low detailed feature recognition. *Int. J. Comput. Appl* **3**, 5–7 (2013)
18. D. F. Llorca, R. Arroyo and M. A. Sotelo, Vehicle logo recognition in traffic images using HOG features and SVM, 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), 2013, pp. 2229–2234.
19. R. Chen, M. Hawes, O. Isupova, L. Mihaylova and H. Zhu, Online vehicle logo recognition using Cauchy prior logistic regression, 2017 20th International Conference on Information Fusion (Fusion), 2017, pp. 1–8.
20. H. Peng, X. Wang, H. Wang, W. Yang, Recognition of low-resolution logos in vehicle images based on statistical random sparse distribution. *IEEE Trans. Intell. Transp. Syst.* **16**(2), 681–691 (2015)
21. Q. Sun, X. Lu, L. Chen, H. Hu, An Improved vehicle logo recognition method for road surveillance images. Seventh Int Symp Comput Intell Design **2014**, 373–376 (2014)
22. R. Chen, et al. Vehicle logo recognition by spatial-SIFT combined with logistic regression. *Fusion* **2016** 2016.
23. S. Sotheeswaran, A. Ramanan, A coarse-to-fine strategy for vehicle logo recognition from frontal-view car images. *Pattern Recognit Image Anal.* **28**(1), 142–154 (2018)

24. J. Zhao, X. Wang, Vehicle-logo recognition based on modified HU invariant moments and SVM. *Multimed. Tools Appl.* **78**(1), 75–97 (2019)
25. K. Meethongjan, T. Surinwarangkoon, V.T. Hoang, Vehicle logo recognition using histograms of oriented gradient descriptor and sparsity score. *Telkomnika* **18**(6), 3019–3025 (2020)
26. S. Yang, et al. VLD-45: A big dataset for vehicle logo recognition and detection. *IEEE Transactions on Intelligent Transportation Systems* PP:99(2021):1–7.
27. Y. Yu, et al. A multilayer pyramid network based on learning for vehicle logo recognition. *IEEE Transactions on Intelligent Transportation Systems* PP:99(2020):1–12.
28. Y. Yu, et al. A cascaded deep convolutional network for vehicle logo recognition from frontal and rear images of vehicles. *IEEE Transactions on Intelligent Transportation Systems* PP:99(2019):1–14.
29. Y. Huang et al., Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy. *IEEE Trans. Intell. Transp. Syst.* **16**(4), 1951–1960 (2015)
30. L. Zhou et al., Detecting motion blurred vehicle logo in iov using filter-DeblurGAN and VL-YOLO. *IEEE Trans. Veh. Technol.* **69**(4), 3604–3614 (2020)
31. C. Pan et al. Vehicle logo recognition based on deep learning architecture in video surveillance for intelligent traffic system. *let International Conference on Smart & Sustainable City IET*, 2014.
32. Y. Xia, F. Jing, B. Zhang. Vehicle Logo Recognition and attributes prediction by multi-task learning with CNN. 2016 12th International Conference on Natural Computation and 13th Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) IEEE, 2016.
33. L. Huan, W. Li, Q. Yujian, Vehicle logo retrieval based on Hough transform and deep learning. *ICCVW 2017*, 967–973 (2017)
34. R. Chen, M. A. Jalal, L. Mihaylova, R. K. Moore, Learning Capsules for Vehicle Logo Recognition, 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 565–572.
35. F.C. Soon et al., Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition. *IET Intel. Transport Syst.* **12**(8), 939–946 (2018)
36. S. F. Chong, et al. Vehicle logo recognition using whitening transformation and deep learning. *Signal Image and Video Processing* (2018).
37. S. Yang et al., Fast vehicle logo detection in complex scenes. *Optics Laser Technol.* **110**, 196–201 (2019)
38. H. Sun, et al. Recognition of vehicle-logo based on faster-RCNN. 2019.
39. R. Liu et al., Vehicle logo recognition based on enhanced matching for small objects constrained region and SSFPD network. *Sensors* **19**(20), 4528 (2019)
40. H. Nguyen. Vehicle logo recognition based on vehicle region and multi-scale feature fusion. *J. Theor. Appl. Inf. Technol* 98.16 (2020).
41. Z. Junxing, et al. Single stage vehicle logo detector based on multi-scale prediction, 2020 *IEICE Trans. Inf. Syst.* E103-D, 10, 2020.
42. S. Yang, et al. A new dataset for vehicle logo detection. *International Symposium on Artificial Intelligence and Robotics*. Springer, Cham, 2018.
43. J. Redmon, A. Farhadi. YOLO9000: better, faster, stronger. *IEEE Conference on Computer Vision & Pattern Recognition* (2017):6517–6525.
44. Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
45. W. Liu, et al. Ssd: Single shot multibox detector. *European conference on computer vision*. Springer, Cham, 2016.
46. C.-Y. Fu, et al. Dssd: Deconvolutional single shot detector. *arXiv preprint arXiv:1701.06659* (2017).
47. T.-Y. Lin, et al. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*. 2017.
48. Z. Tian, et al. Fcos: Fully convolutional one-stage object detection. *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
49. R. Girshick. Fast r-cnn. *Proceedings of the IEEE international conference on computer vision*. 2015.
50. S. Ren et al., Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inform. Process. Syst.* **28**, 91–99 (2015)
51. S. Gidaris, N. Komodakis. Object detection via a multi-region and semantic segmentation-aware cnn model. *Proceedings of the IEEE international conference on computer vision*. 2015.
52. Z. Cai, N. Vasconcelos. Cascade R-CNN: Delving into high quality object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
53. J. Zhang et al., Vehicle logo detection based on deep convolutional networks. *Comput. Electr. Eng.* **90**, 107004 (2021)
54. Y. Yu et al., Vehicle logo recognition based on overlapping enhanced patterns of oriented edge magnitudes. *Comput. Electr. Eng.* **71**, 273–283 (2018)
55. H. Rezaatofghi, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* IEEE, 2019.
56. Zheng, Z., et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. (2020).
57. Z. Zheng, P. Wang, W. Liu et al., Distance-IoU loss: faster and better learning for bounding box regression. *Proc. Conf. AAAI Artif. Intell.* **34**(07), 12993–13000 (2020)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.