Vehicle-Road Collaboration: A Survey

Ziyang Zhang, *Student Member, IEEE,* Huan Li, *Senior Member, IEEE,* Jie Liu, *Fellow, IEEE*

***Abstract*—The abstract goes here.**

***Index Terms*—Vehicle-Road Collaboration, Object Detection, Deep Learning, Traffic Sign.**

I. Introduction

A. Vehicle-Road Collaboration

As we all know, Autonomous driving as an intersection of emerging technologies such as big data, 5G, artificial intelligence and IoT, which is considered one of the reality scenarios for the successful application of deep learning. To achieve autonomous driving, it is necessary to construct the vehicle-road synergy architecture in which "Smart Vehicle" and "Wise Road" interact with each other. Specifically, smart vehicle refer to autonomous driving cars, the smart road is a smart network with environmental perception, edge computing and information collaboration capabilities.

Smart vehicle are equipped with advanced on-board sensors, controllers, and actuators , realizing the exchange and sharing of information between cars and X (people, cars, roads and clouds) through modern communication and network technology. The vehicle has functions about complex environment perception, intelligent decision-making and collaborative control, which can realize the purpose of autonomous driving. First of all, on the smart vehicle side, lidar, millimeter wave radar, camera, etc. are important sensing components, which the environmental information they collected is one of the most essential sources of data generated by smart vehicle. Secondly, deploying edge devices based on digital and intelligent transportation infranstructure at roadside to perceive ,analyze and process the state of the smart vehicle. To make sure the low-lantency in communication, edge computing has become an important technical feature of the development of the autonomous driving in 5G. Edge computing nodes will be deployed on both the roadsides to communicate with the vehicle at any time, upload the data to the cloud server or perform real-time processing locally, and finally feedback the results to the vehicle.

B. Related Work

*1) **Model**:* LeCun et al.[1] apply multi-scale convolution networks to the task of traffic sign classification for GTSDB competition, the result show that it is accuracy of 98.97%, which above the human performance of 98.91%. Tabernik et al.[2] used Mask-RCNN to perform end-to-end traning for large-scale Traffic Sign detection and Recognition. In addition, some methods are proposed to improve the accuracy. Firstly, perform online hard-example mining for samples used by OHEM, and then with the help of a novel Geometric and appearance deformation distribution-based data enhancement technology, this model has achieved 2% to 3% average error rate on GTSDB. The biggest disadvantage of CNN is that it loses a lot of information in the pooling layer, which reduces the spatial resolution. To solve above problem, Kumar et al.[3] adopt novel deep learning model for traffic sign detection using capsule networks, and achieved the first place in detection accuracy on GTSDB that year. Due to CNN cannot detect small traffic sign, Tian et al.[4] proposed a multi-scale recurrent attention network for traffic sign detection to solve this problem. This method is used to select relevant features from the convolutional layer, and gradually modify the potential object area by repeatedly using the information of adjacent receptive fields. Specifically, based on the object detection algorithm DSSD, it adds and modifies some modules for the detection of traffic signs at different scales. The first is to add a multi-scale attention module. The second is the addition of a recurrent attention graph module. It can improve the utilization of image context, and compared with state of the art methods, which has improvement of 5.2% AUC on GTSDB. Simultaneously, there is a 1% accuracy and recall improvement on TT-100K. Zhu et al.[5] proposed a robust end-to-end CNN which can detect and classify traffic sign. Particularly, this neural network consists entirely of convolutional layer so that can be used for localization. In order to make full use of the contextual information in the image, Yuan et al.[6] proposed a noval end-to-end deep learning-based model. It includes a multi-resolution fusion network structure, which can learn more effective features from small objects. Meanwhile, the task of traffic sign detection is regared as a spatial sequence classification and regression task and a vertical spatial sequance attention model is proposed, which can improve accuracy furthermore.

*2) **Sample**:* For the object detector, what sample is important to it? In response to this problem, Cao et al.[7] proposed the concept of prime sample and research the importance of the sample by using mAP as an indicator. To find the main sample, a hierarchical local ranking strategy is designed and proposed an efficient sampling strategy PISA. There are two components in PISA: Importance-based Sample Reweighting(ISR) and Classification-aware Regression Loss(CARL). ISR makes the training process more inclined to the prime samples, instead of treating all samples uniformly. It assigning different loss weights to the samples according to their importance. CARL adopts a common goal to learn classifiers and regressors to suppress unimportant samples and highlight main samples. The results show that PISA achieved a mAP improvement of about 2% on COCO dataset and PISA is also better than the benchmark on VOC dataset. Apart from this, Ma et al.[8] proposed method which can use a few samples to detect object. It is consist of attention-RPN, Multi-Relation Detector and Contrastive Training strategy. Finally, it is achieved state of the

art on the dataset contributed by the author and other datasets.

*3) **Domain migration**:* Domain adaption will emerge when a model trained on large dataset is deployed to edge devices. Wong et al.[9] proposed MicorNet for edge device to the detection of traffic sign. It is a highly compact DNN-based architecture, which is based on macro architecture design priciples (e.g., parameter accuracy optimization, spectral macroarchitecture augmentation, etc.) as well as numerical macro architecture optimization strategies. The results on GTSDB shows that the model requires only a few of parameters and more less calculation to maintain high recognition performance (98.9% accuracy). In addition, the inference time running on the high performance processor in Cortex-A53 only need 32.19ms. Due to the trade-off between accuracy and inference speed, most of the existing weight clipping techniques cause a decrease in inference speed and limit the acceleration performance on mobile terminal devices. In order to solve this problem, Ma et al.[10] proposes a sparse pattern-based inference framework, which is an highly efficient algorithm level pruning framework and compiler level inference framework. It has high accuracy as well as friendly to hardware devices. The test on mobile platforms shows it can use large-scale DNN-based models for real-time inference.

II. Environmental perception

Most related work recently are foucs on muti-sensor collaborative perception and data fusion technologies at present, which include traffic sign detection, plate recognition, lane detection and anomaly detection in traffic videos. In this paper, we described in detail the latest research situation in abovementioned directions. In addition, some unresolved challenges and open issues are also proposed.

A. *Traffic Sign Detection*

Traffic sign detection essentially belongs to the catogory of object detection, which plays an import role in autonomous driving. In the past 20 years, although computer vision has made a serires of breakthroughs in the detection of general opjects, there are still many difficults and challenges in the detection and recognition of traffic sign in autonomous driving, such as:

- **Low pixels**
  Even if the pixels of the vehicle-mounted camera are very high, the traffic sign occupies only a small area in the entire pictrue captured by the camera. When the region of the traffic sign is cut out from the picture, its pixels are not high enough.
- **Light changes**
  As shown in Figure 1, it's particulary difficult to detect traffic signs when the vehicle is driving under strong light or at night.
- **Obstacles**
  It is common for traffic signs to be obstacled by objects, such as occluded by leaves, muddy water pollution, and obstruction of signs after snowfall. Figure 2 shows a situation where the traffic sign is hidden by leaves. It is difficult to identify the correct sign at this time.

Fig. 1: Light changes

Specifically, there are mainly two types of occlusion in object detection: (1) The object to be detected is obstacled by interference objects. (2) The mutual occlusion between the objects to be detected. For the former type of occlusion, it is difficult to solve it with efficient solutions. The existing work is to use more data and richer features. As to the latter, the existing methods don't make full use of the contextual information of the image to fuse more features. To slove the above problem, Wang et al.[11] first analyzed the data set and quantitatively described the impact of occlusion on object detection. Specifically, they inspired by the priciple of same-sex repulsion and opposite-sex attraction of mangnets, a novel loss function called Repulsion Loss is proposed to make the prediction bounding box as close to the ground truth as possible, while repelling the same kind to avoid misdetection. This method improves the performance of the Faster-RCNN based object detector and reduce the sensitivity of NMS to the IoU threshold. The former is only considered from the perspective of the optimization goal. Zhang et al.[12] mitigated the impact of the occlusion problem on object detection results from the perspective of optimization goal and network architecture jointly. Specifically, starting from the loss function and the key pooling operation in the object detector respectively, the problem of detecting occluded objects is improved. In terms of the loss function, the idea is similar to [3], but a simple and easier to implement form is adapted, which is the candidate proposals assigned to the same ground truth are close to the average of the candidate proposals themselves. Since the optimization goal is only near problem, rather than far away, there is no need to IoU, which simplifies the implementation. Zamir et al.[13] proposed a novel network architecture called MIRNet, which is consisted of parallel multi-resolution convolution streams, spatial and channel attention mechanisms, multi-sccle feature aggregation and Information exchange across multi-resolution streams, which can maintain spatially accurate high-resolution representayions through neural networks and obtain sufficient information from low-resolution representayions. In short, MIRNet has achieved excellent results in image enhancement task by learning rich features.

In addition to improving the existing object detection models, data enhancement techniques can also be used to deal with occlusion problems. Nvidia researchers pro-

posed a panoramic-aware image synthesis method in [14], which greatly improves the quality of image generation. It can generate clear and separable instance images even in scenes where multiple object instances obscure each other.

Fig. 2: Obstacles

- **Motion blur**
  As shows in Figure3, due to the shake of car during driving, the image of traffic sign by the vehicle-mounted camera will be blurred. It is difficult to meet the situation that the camera is facing the sign simultaneously, which will cause the traffic sign in image to be deformation.

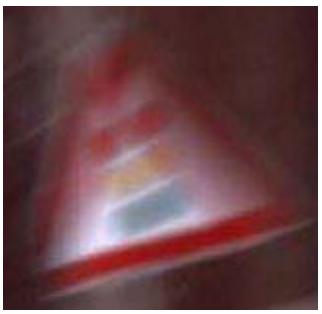

Fig. 3: Motion blur

- **Real-time detection**
  The real-time is critical for autonomous driving, it is obviously that the vehicle must detect the traffic sign passing by the roadside in realtime and make corresponding decisions to avoid accidents during high-speed driving.

Some mature traffic sign detection technologies have been commercialized at present. For example, Autopilot-Tesla's autonomous driving assistance system, which have supported the recognition and detection of traffic lights, stop sign and roundabouts. It is also mean that the system will gradually have the ability to fully autonomous driving.

B. Plate Recognition

Subsubsection text here.

C. Lane Detection

Subsubsection text here.

D. Anomaly Detection in Traffic Videos

Subsubsection text here.

III. Overview of Related Knowledge

In this section, we provided a review of deep learning and object detection from different perspective.

A. Deep Learning

Since some of the techniques discussed in this paper rely on the specific forms of deep learning, therefore, we first provide a brief background on deep learning.

As a branch of machine learning, deep learning is a feature learning approaches[15] that uses multiple layers of non-linear to learning the abstract representayions of high-dimensional data[16] and build computational models.

*1) **Features**:*

- **Data size**
  Figure 4 and Figure 5 list the changing trend of number of samples and the size of some data sets over time respectively. Early machine learning algorithms are relatively simple, easy to train quickly, and require relatively small data sets. With the continuous improvement of computing power, the designed algorithm is becoming more and more complex, and the demand for the amount of data also increases. Especially in deep learning, the number of network layers is generally deep, and the number of parameters of the model can reach tens of millions or even billions, and the scale of the data set required is usually huge. The research progress of deep learning is essentially based on the paradigm of "driving small tasks with big data", relying on large amounts of data to train classifiers to solve a single task. In recent years, it is a very valuable direction to study algorithm models that require less data to drive large tasks with small data.

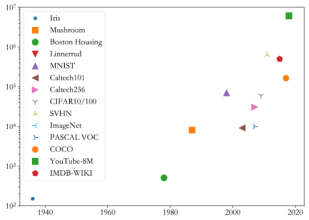

Fig. 4: Sample number trend

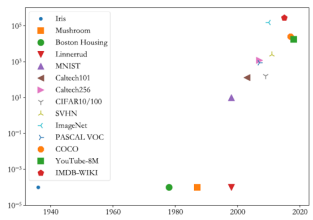

Fig. 5: Size trend

- **Computing ability**
  The improvement of computing power is an important factor in the revival of artificial intelligence. Traditional machine learning algorithms do not have stringent requirements on amount of data and computing power like neural networks. Usually, satisfactory results can be obtained by serial training on the CPU. However, deep learning relies heavily on parallel accelerated computing devices. Most current neural networks use deep learning accelerated hardware devices such as NVIDIA GPU and Google TPU to train model parameters. Figure 6 illustrates the indicator transformation curve of 1 billion floating point operations per second (GFLOPS) of

NVIDIA GPUs and x86-based CPUs from 2008 to 2017. It can be seen that the x86-based CPU curve changes relatively slowly, while the floating-point computing power of NVIDIA's GPU has increased exponentially, which is mainly driven by the increasing amount of game calculations and deep learning calculations.

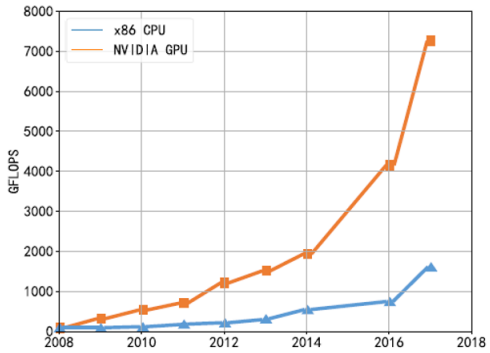

Fig. 6: NVIDIA GPU FLOPS trend (Data from NVIDIA)

- **Network size**
  Figure 7 shows the changing trend of the number of network model layers for classification tasks on the ImageNet dataset. It can be seen that the early multi-layer neural network has several layers. With the rise of deep learning and the increase in computing power, neural network models with dozens of layers or even hundreds of layers have been proposed, and the size of the input pictures has gradually increased. These changes have made the total amount of parameters of the network reach tens of millions.

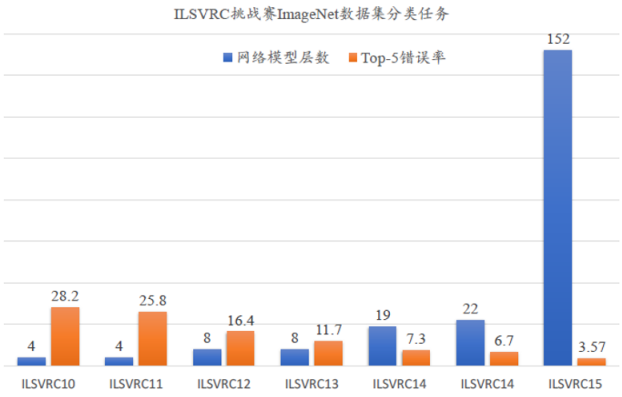

Fig. 7: Trend of network layer changes

- **General intelligence**
  General intelligence is the kernel of deep learning. In other words, the features of each layer of the neural network are not manually designed using prior knowledge, but are learned from data using a common learning process. It has always been the common desire of human to design a universal intelligent mechanism that can learn and adjust itself like the human brain. From the current point of view, deep learning is one of the algorithms closest to general intelligence.

*2) **Class models in deep learning**:*

- **CNN**
  CNN consists of a series of convolutional layers, pooling layers and fully connected layers. Among them, the convolutional layer and the pooling layer are the core of CNN. Specifically, the inner product operation of the sliding window and filter of the image is the so-called convolution operation. The purpose is to extract different features of the input[17].
- **RNN**
  RNN is a type of neural network used to process sequence data[18]. In other words, the following data is related to the previous data. The basic neural network only establishes weight connections between layers. The biggest difference of RNN is that weight connections are also established between neurons in the same layer, which can connect previous information to the current task.
- **LSTM**
  Schmidhuber et al[19].proposed LSTM to solve the short-term memory problem of RNN, which is a special tyoe of RNN that can learn long-term dependent information.LSTM has three types of gate structures: forget gate, input gate and output gate. Firstly the forget gate determine what information to discard from the cell state. Secondly, the input gate determine what new information is stored in the cell state. Finally, the output gate determine what value to output based on the cell state. The key to LSTM is the cell state. The "memory" unit of LSTM is called a cell. The input of the cell is the previous state and the current input. These "cells" determine which previous information and state need to be retained/remembered, and which should be erased.
- **RL**
  Reinforcement learning is the process of continuous interaction between the agent and the environment[20]. Among them, the agent perceives the state of the environment so that it can selects and executes corresponding actions according to the strategy to maximize the reward obtained. Reinforcement learning includes five components:
  - **State:** reflects the state characteristics of the environment.
  - **Action:** represents the behavior of the agent.
  - **Policy:** represents the decision model of the agent, it is a probability distribution which is expressed in the probability of performing the corresponding action in a certain state.
  - **Reward:** represents the feedback given to the agent after receiving a certain action in a certain state, reflecting the quality of the action.
  - **State transition probability:** represents the probability distribution of the state transitioning to another after the current environment accepts an action.
  Reinforcement learning problems are generally modeled by Markov decision process(MDP).

B. Object Detection

The existing object detection methods can be divided into two categories: traditional-based approaches and deep

learning-based approaches.

*1) **Traditional-Based Approches**:* The research on vision-based traffic sign detection can be traced back to 20 years ago. Due to traffic signs have special shapes and colors, traditional detection approaches are usually based on color threshold, visual saliency detection, morphological filtering and edge/contour analysis.

In 2005, the researchers from Siemens published a paper using traditional machine learning algorithms to identify traffic signs[21]. This method completes the recognition of traffic signs in 6 steps. First, it is need to extract each small area by using the sliding window method, then use wavelet filtering and AdaBoost to find all the areas in the sliding window that may contain traffic sign information, furthermore, execute LDA transformation for dimensionality reduction. Finally, with the help of Native Bayes Classifier to predict Types of traffic signs. Actually treat traffic signs as a classification problem.

The method of using wavelet filtering to identify traffic signs is not the original one in the above paper.Before deep learning, there was a classic paper in the field of image localization[22]. This paper was published in 2001, which is a classic article about object detection. The highlight is the use of "Integral Image" to calculate Haar-like features to achieve accelerated calculations; AdaBoost is used to train weak classifiers to form strong classifiers, which has higher classification accuracy and speed.

*2) **Deep Learning-Based Approches**:* As shown in Table 2, object detection models based on deep learning are mainly divided into two categories: two-stage object detection models and one-stage object detection models. The former is to first generate a series of candidate frames as samples by specific algorithm, and then classify the samples through CNN; the latter directly converts the problem of object frame positioning into regression problems without generating candidate frames. Due to the difference between the two methods that there are differences in performance obviously. The former is superior in detection accuracy and positioning accuracy, while the latter is superior in algorithm speed.

Specifically, YOLOv3 predicts an objectness score for each bounding box using logistic regression and use binary cross-entropy loss[23] for the class predictions. As an improved version of YOLOv3, YOLOv4[24] adopts a series of the latest computer vision technologies, which modify state-of-the-art methods and make them more effecient and suitable,including CBN[25], PAN[26], SAM[27], etc. It improves YOLOv3's AP and FPS by 10% and 12%, respectively. Faster R-CNN[28] introduce the RPN and down-stream detection network to share convolutional features of the entire image, which has a high accuracy.

TABLE I: Deep learning-based models

Model	Type	Backbone	mAP-50	FPS	Year
Faster R-CNN	Two-stage	ResNet-101	78.8%	5	2016
YOLOv3	One-stage	Darknet-53	55.3%	20	2018
YOLOv4	One-stage	CSPDarknet53	64.9%	23	2020

C. *Plate Recognition*

Subsection text here.

D. *Lane Detection*

Subsection text here.

E. *Anomaly Detection in Traffic Videos*

Subsection text here.

F. *Pedestrian & Obstacle Recognition*

Subsection text here.

IV. Traffic Sign Dataset

Table 1 is a summary of currently known traffic sign data sets.

TABLE II: Traffic sign data sets

Name	Area	Number	Size	Format	Category
CTSDB	China	6K	190M	PNG	58
CCTSDB	China	15K	7.68G	PNG	3
TT-100K	China	100K	105G	PPM	3
Mapillary	6 continents	6K	100G	JPG	300
GTSDB	German	50K	1.2G	PPM	40
LISA	US	6K	7.7G	VOC	47
BelgiumTS	Belgium	3.6K	250M	PPM	62
DFG	Slovenian	7K	7G	JPEG	200
KUL Belgium	Belgium	16K	50G	PPM	62
SwedishTS	Swedish	20K	15G	JPG	-

A. *CTSDB*

The CTSDB include traffic sign recognition database(TSRD), traffic sign detection database(TSDD) and traffic panel database(TPD). All databases are collected by camera under nature scenes or from BAIDU Street View. Among them, the TSRD includes 6164 traffic sign images containing 58 sign categories. The images are divided into two sub-database as training database and testing database. The training database includes 4170 images while the testing one contains 1994 images. All images are annotated the four coordinates of the sign and the category.

B. *CCTSDB*

The CCTSDB is obtained after expansion on the basic of CTSDB, which include more than 15,000 traffic sign images containing three major categories with no subdivision.

C. *TT-100K*

TT-100K provides 100,000 images containing 30,000 traffic-sign instances. Each traffic-sign in the benchmark is annotated with a class label, its bounding box and pixel mask. It contains three major categories: warning, prohibitory and mandatory.

D. *Mapillary*

The Mapillary Traffic Sign Dataset is the world's largest and most diverse publicly available traffic sign dataset to detect and recognize traffic signs. The dataset consists of 100,000 images containing more than 300 traffic sign classes with bounding box annotations.

E. *GTSDB*

The GTSDB are divided into two sub-dataset as training dataset and test dataset. The training dataset include 50,000 images containing 40 categories and the testing dataset include 900 images which consist of 600 training images and 300 testing images.

F. *LISA*

The LISA Traffic Sign Dataset is a set of videos and 7855 annotated on 6610 frames containing 47 US traffic signs. The dataset include some images in color and some in grayscale particularly.

G. *BelgiumTS*

The Belgium Traffic Sign Dataset includes 3600 images containing 62 Belgium traffic sign categories which consist of 2400 training images and 1200 testing images.

H. *DFG*

The DFG Traffic Sign Dataset consists of 200 traffic sign categories captured in Slovenian roads spanning in around 7000 high-resolution images. The dataset divided into 5254 training images and 1703 testing images.

I. *KUL Belgium*

The KUL Belgium Traffic Sign Dataset includes 16,000 images containing 62 Belgium traffic sign categories which consist of 11,000 training images and 5000 testing images.

J. *SwedishTS*

The Swedish Traffic Sign Dataset includes more than 20,000 images with 20 ratio labeled containing 8 traffic sign categories. The dataset does not define the proportion of training dataset and testing dataset.

V. Challenges and Open Issues

Although the autonomous driving has been achieved a series of breakthrough results at present, there are many challenges remain in deploying deep learning on the autonomous driving. We next discuss some of these problems.

A. *The task offloading and resource allocation*

Due to in-vehicle applications requires more computing and communicatuion capabilities, it is necessary to design reasonable strategy for task offloading and resource allocation in autonomous driving to performs well in computationally intensive and delay-sensitive tasks. Qi et al.[29] and Liu et al.[30] treat the offloading decision of Multi-tasks as a long-term planning problem. Different service offloading decision frameworks are proposed, which can provide the best strategy through Deep Reinforcement Learning(DRL).

B. *Model Compression*

Although deep learning-based models perform well on large data sets, the domain adaption is emerged when the model is migrated to edge devices with limited resources and computing power. In recent years, some approaches about model compression and pruning are proposed to solve above issues.

C. *Communication Delay*

The existing object detection algorithms still have high inference latency when process the tigh-dimensional data[31], such as image. In addition, there is also high-lantency during upload data collected by sensors to cloud-server and return the result to the vehicle, which brings great challenges to the real-time of autonomous driving.

VI. Conclusion

The conclusion goes here.

References

[1] Sermanet P, LeCun Y. Traffic sign recognition with multi-scale convolutional networks[C]//The 2011 International Joint Conference on Neural Networks. IEEE, 2011: 2809-2813.

[2] Tabernik D, Skočaj D. Deep learning for large-scale traffic-sign detection and recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(4): 1427-1440.

[3] Kumar A D. Novel deep learning model for traffic sign detection using capsule networks[J]. arXiv preprint arXiv:1805.04424, 2018.

[4] Tian Y, Gelernter J, Wang X, et al. Traffic sign detection using a multi-scale recurrent attention network[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(12): 4466-4475.

[5] Zhu Z, Liang D, Zhang S, et al. Traffic-sign detection and classification in the wild[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2110-2118.

[6] Yuan Y, Xiong Z, Wang Q. VSSA-NET: vertical spatial sequence attention network for traffic sign detection[J]. IEEE transactions on image processing, 2019, 28(7): 3423-3434.

[7] Cao Y, Chen K, Loy C C, et al. Prime sample attention in object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11583-11591.

[8] Ma X, Niu W, Zhang T, et al. An Image Enhancing Pattern-based Sparsity for Real-time Inference on Mobile Devices[J]. arXiv preprint arXiv:2001.07710, 2020.

[9] Wong A, Shafiee M J, Jules M S. Micronnet: A highly compact deep convolutional neural network architecture for real-time embedded traffic sign classification[J]. IEEE Access, 2018, 6: 59803-59810.

[10] Ma X, Niu W, Zhang T, et al. An Image Enhancing Pattern-based Sparsity for Real-time Inference on Mobile Devices[J]. arXiv preprint arXiv:2001.07710, 2020.

[11] Wang X, Xiao T, Jiang Y, et al. Repulsion loss: Detecting pedestrians in a crowd[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7774-7783.

[12] Zhang S, Wen L, Bian X, et al. Occlusion-aware R-CNN: detecting pedestrians in a crowd[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 637-653.

[13] Zamir S W, Arora A, Khan S, et al. Learning Enriched Features for Real Image Restoration and Enhancement[J]. arXiv preprint arXiv:2003.06792, 2020.

[14] Dundar A, Sapra K, Liu G, et al. Panoptic-based Image Synthesis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8070-8079.

[15] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[16] Pouyanfar S, Sadiq S, Yan Y, et al. A survey on deep learning: Algorithms, techniques, and applications[J]. ACM Computing Surveys (CSUR), 2018, 51(5): 1-36.

[17] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[18] Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model[C]//2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2011: 5528-5531.
[19] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
[20] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018.
[21] Bahlmann C, Zhu Y, Ramesh V, et al. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information[C]//IEEE Proceedings. Intelligent Vehicles Symposium, 2005. IEEE, 2005: 255-260.
[22] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. IEEE, 2001, 1: I-I.
[23] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[24] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv preprint arXiv:2004.10934, 2020.
[25] Yao Z, Cao Y, Zheng S, et al. Cross-iteration batch normalization[J]. arXiv preprint arXiv:2002.05712, 2020.
[26] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
[27] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
[28] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
[29] Qi Q, Wang J, Ma Z, et al. Knowledge-driven service offloading decision for vehicular edge computing: A deep reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology, 2019, 68(5): 4192-4203.
[30] Liu Y, Yu H, Xie S, et al. Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(11): 11158-11168.
[31] Chen J, Ran X. Deep learning with edge computing: A review[J]. Proceedings of the IEEE, 2019, 107(8): 1655-1674.