

Received May 28, 2022, accepted June 18, 2022, date of publication June 21, 2022, date of current version June 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3185106

Designing Electronic Traffic Information Acquisition System Using Deep Learning and Internet of Things

RONGRONG HAN 

Heilongjiang Provincial Key Laboratory of Underground Engineering Technology, Harbin University, Harbin 150086, China

e-mail: gongxhrr@hrbu.edu.cn

ABSTRACT In implementing the Intelligent Traffic Monitoring System (ITMS), timely and effective access to road traffic information is an essential link. It requires an effective traffic Information Acquisition System (IAS) to collect real-time data and transmit the collected information to the background for processing. Therefore, this paper studies on-road vehicle information recognition based on Deep Learning (DL). Firstly, a framework of traffic IAS is proposed. Then, an improved MT-GooGleNet model based on Convolutional Neural Network (CNN) is proposed to locate and recognize vehicles in traffic images. Finally, the performance of the model is analyzed by simulation. The experimental results of vehicle position recognition show that the classification accuracy of Multi-Task (MT)-GooGleNet after fine-tuning is 99.5%. Compared with other models, the MT-GooGleNet model proposed is the best in vehicle position recognition, and its positioning accuracy is very high. The results of vehicle identification show that after data enhancement and pre-training, the testing set accuracy of the MT-GooGleNet model is 79.96%. The results show that the model's accuracy has been dramatically improved after processing. The research provides a reference for establishing IAS in the future.


INDEX TERMS Deep learning, Internet of Things, signal acquisition, system design, vehicle identification and positioning.

I. INTRODUCTION

The last several decades have seen rapid socio-economic development and a much-improved quality of life in China, with which household car ownership and industrial vehicle volume have gained a substantial boost. Yet, such civil convenience has come at a cost, such as the stubborn traffic congestion and up-rising accidents. Therefore, there is an imminent demand for traffic congestion alleviation and transportation resource distribution using modern technological means. Under such circumstances, the ubiquitous Internet of Things (IoT) and maturing Artificial Intelligence (AI) technologies might help devise an intelligent urban traffic monitoring and information management system for the well-being of residents [1], [2].

Arguably, for a robust and viable Intelligent Traffic Monitoring System (ITMS), timely and data acquisition efficiency should be the utmost concern. Overall transformational data

can cover vehicle speed, traffic flow, road occupancy rate, and vehicle type. Every link matters for timely and effective data collection, transmission, processing, and forecast [3]–[5]. The first link will be traffic data collection, against which multiple models and systems are available. In particular, real-time traffic information is a prominent and comprehensive traffic situation indicator, which is relatively easier to collect and against which many ready-made video detection tools have been invented. Compared with other traditional detection methods, the video detection approach features intuitive traffic scenes and an excellent corner-cover detection range. Thus, vehicle-oriented Image Recognition (IR) and Video Analysis (VA) technology can effectively realize real-time vehicle tracking and management [6]. On the other hand, the Convolution Neural Network (CNN) is a Deep Learning (DL)-based common image processing feedforward Neural Network (NN) structure with an outstanding local connection and weight sharing mechanism [7], [8]. In a CNN, neurons are organized in some way to respond to overlapping vision domains. With the introduction of more hidden layers,

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

the depth of CNN increases and can be further applied to more advanced scenarios, such as Machine Translation (MT), Speech Recognition (SR), Image Segmentation (IS), and Natural Language Processing (NLP). Its unique operation in the convolution pool has high efficiency in image processing. It has unparalleled advantages in 2D graphics tasks such as IR and positioning [9]. The Residual Neural Network (ResNet) is an innovative form of CNN to make the convolutional layer learn a residual rather than directly learn the target [10]. For example, Zhao *et al.* (2018) [11] analyzed the characteristics of CNN and ResNet to build a ResNet model. They collected image data using the Unmanned Aerial Vehicle (UAV) remote sensing system, based on which the IR performance of the proposed ResNet model was tested. Chen *et al.* (2018) [12] introduced a deep network architecture – the prototype parts network. The network parsed the image by finding the prototype parts and classified the image with the evidence of the prototype. The network only used image-level labels for training without any annotations of the image part. Thus, CNN is often seen in IR and vehicle location recognition.

Background complexity and target diversity increase difficulty in IR, and factors like illumination or shooting angle will result in deviations in the positioning and identifying vehicles from images. Therefore, based on DL, this paper proposes an improved Multi-Task (MT) learning GooGleNet model based on CNN to locate and recognize vehicles in images and studies the model's performance. The main innovation is that CNN in DL is applied to IR, which has contributed to improving the IR accuracy and is the main difference and novelty between the present work and previous research methods. At the same time, the research has a certain reference value for the digitization and intellectualization of urban transportation infrastructure.

II. RECENT RELATED WORK

A. CURRENT STATUS OF RESEARCH ON THE IOT-BASED ITMS

Compared with the rapid urbanization, especially the surging car ownership, the development of traffic management has been relatively backward, such as insufficient road planning and construction. These are the leading causes of urban traffic congestion, a modern urban disease demanding an immediate solution. In particular, ITMS is gradually becoming the research direction for solving traffic congestion problems globally. Zhang *et al.* (2020) [13] established the vehicle motion model using OPNET Modeler. They simulated the multi-hop scenario of the vehicle Self-Organizing Network (SON) in the ITMS based on the IoT, finding that the vehicle SON presented a stable overall performance under a low speed. Brincat *et al.* (2020) [14] identified several IoT-related cases in ITMSs, such as connected and autonomous vehicles, cooperative traffic networks, and smart roads, to improve data dissemination create heterogeneous connections in high-capacity environments and low-latency applications. Li *et al.* (2020) [15] integrated Data Mining (DM) technology and distributed parallel Hadoop technology to realize the

TABLE 1. Differences and analysis of different literature methods.

Number	Year	Scholar	Research method
1	2020	Zhang et al	OPNET Modeler vehicle motion model
2	2020	Brincat et al	Case study of ITMS
3	2020	Li et al	Integrated DM and distributed parallel Hadoop
4	2020	Lv et al	DL+ITMS
5	2021	Kaffash et al	Big data algorithm + ITMS literature

ITMS architecture, analyzed the running state, and applied the IoT to obtain data from the perception layer. The simulation results showed that the system had good prediction performance. Lv *et al.* (2020) [16] used the DL method to study, simulate, improve the safety of ITMS and statistically analyzed the data transmission performance, accuracy prediction performance, and path change strategy of the system. The simulation data transmission performance showed a 100% success rate. Kaffash *et al.* (2021) [17] applied big data algorithms to further improve the intelligence of applications in the transportation field by reviewing the literature on big data algorithms and their applications in ITMS. The research outcome determined the research gap and direction in the future. To sum up, scholars have conducted extensive research and used different methods to study IoT-based ITSM, as encapsulated in Table 1.

B. CURRENT STATUS OF RESEARCH ON THE INFORMATION ACQUISITION SYSTEM (IAS) BASED ON DL

Ke *et al.* (2020) [18] proposed an automatic data acquisition and annotation strategy based on target detection to obtain various fine-grained models and quickly label vehicle images. They proposed a dense attention network because the traditional DL network has a serious feature information transmission attenuation between each layer, lacks mutual dependence between features, and ignores important regions and features. Kwan *et al.* (2020) [19] constructed a DL tool –You Only Look Once (YOLO)– for target detection and classification based on real-time DL, pixel-coded exposure measurement, and vehicle detection and classification. They conducted many experiments and proved the feasibility of the real-time framework, which could recognize the objects and their positions in the image and speed up the acquisition of good detection and classification performance.

Considering the time dependence of sensor data, Li *et al.* (2020) [20] established a superimposed Long Short-Term Memory (LSTM) model to detect vehicle maneuvers. The system's performance was compared with previous studies and various Machine Learning (ML) methods, finding that the system accurately detected different vehicle maneuvers. Yu *et al.* (2020) [21] established a traffic safety solution

based on DL for automatic and manual hybrid vehicles in the 5th Generation Mobile Communication Technology (5G) ITSM. In this scheme, the driving trajectory data set and the natural driving data set were used as the network input of the 5G-enabled LSTM ITSM. The softmax function calculated the probability matrix of each intention, which effectively improved the accuracy and real-time performance of intention recognition. It helped to solve the lane change problem in the mixed traffic environment.

Through the above analysis, the application of DL in intelligent traffic is expanding. Vehicle-oriented IR+VA technology can effectively realize real-time vehicle tracking management. However, few studies are on vehicle positioning and type recognition using DL. Therefore, CNN based on DL is of great value for vehicle identification and vehicle type and location.

III. DESIGN OF THE TRAFFIC INFORMATION ACQUISITION SYSTEM (IAS)

A. SYSTEM FRAMEWORK

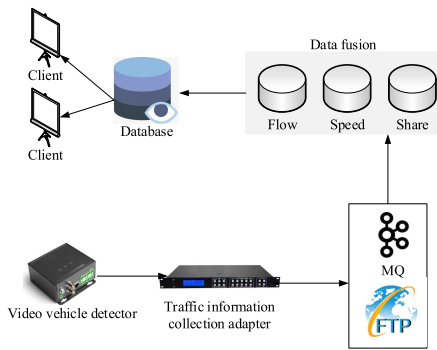


FIGURE 1. Traffic IAS framework.

B. FUNCTIONAL MODULES OF THE TRAFFIC IAS

The primary module in the traffic IAS is the video acquisition and transmission module. It collects video images from the traffic scene through the acquisition device and transmits them to the video processor for processing. Acquisition range and image quality are two main concerns in traffic video acquisition modules. Obviously, a small detection range, blurred or occluded video images, and images with poor contrast will increase image processing difficulty. Thus, there is a need to adjust system parameters during acquisition, such as adjusting the camera's aperture and shutter.

The video processing module is at the core of the entire Traffic IAS. It directly or indirectly collects traffic information through vehicle detection and tracking, so the accuracy of vehicle positioning and recognition will have a direct impact on the acquisition of traffic information signals. Therefore, the present work focuses on vehicle positioning and recognition.

Fig.2 gives the algorithm flow of the designed urban road-oriented video detection model. At first, some parameters are initialized. Then, some basic information is obtained from the

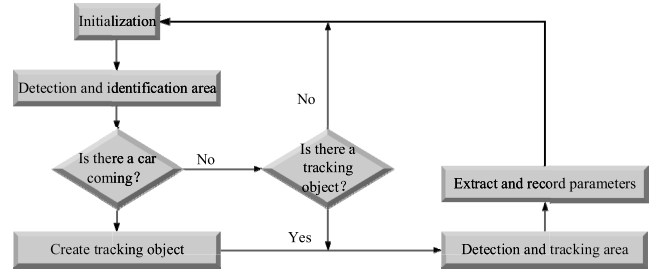


FIGURE 2. Algorithm flow of video detection model.

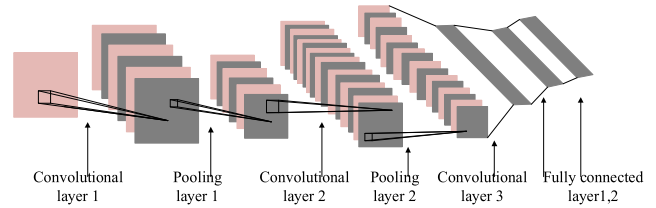


FIGURE 3. Basic structure of the CNN used in IR.

application environment. The virtual line in the lane divides a road into the detection area and the tracking area. The traffic IAS creates a tracking object only when a car is driving in a tracking area. Such a mechanism can reduce the blindness and calculation of tracking.

IV. NETWORK MODULE OF THE DL

A. CONVOLUTIONAL NEURON NETWORKS

CNN is a feedforward Deep Neural Network (DNN). Its essence is the combination of classifier and features extractor. After multi feature learning from the input image, a set of feature vectors will be extracted. Feature vectors are close representations of the original image. Finally, feature vectors are input into the classifier to classify and recognize the data [22], [23]. As in Fig. 3, CNN is generally used for IR and comprises lower sampling, input, convolution, output, and fully connected layers. It is mainly used for IR.

The convolution layer is unique to CNN. Eq. (1) illustrates its activating function ReLu.

$$ReLU(x) = \max(0, x) \quad (1)$$

Eq. (2) demonstrates the feature calculation accorded with the convolution layer.

$$y_n^l = f_1 \left(\sum_{m \in V_n^l} y_m^{l-1} \otimes \omega_{m,n}^l + b_n^l \right) \quad (2)$$

In Eq. (2), y_n^l refers to the n_{th} feature map in the first layer, b_n^l means the offset connected to the first layer, $\omega_{m,n}^l$ expresses the weight of the connection between the m_{th} feature map in the upper layer and the n_{th} neuron in the first layer, and V_n^l represents the set of feature maps connected to the first layer.

The underlying sampling layer contains the same number of features as the previous convolution layer. The underlying sampling layer divides the input into many blocks, and the

values of each block are obtained by determining the sampling method for each pixel and added with bias. Finally, functions are output by stimulating [24], [25]. The bottom sampling can make the features more robust against deformation. If the local features do not change after deformation, the uniformly sampled features are consistent. Meanwhile, the sampling size of the feature map is reduced, thus greatly reducing the amount of data required for subsequent processing and improving the training efficiency [26]. Eq. (3) refers to the characteristic calculation of the down-sampling layer.

$$y_n^l = f_1 \left(z_n^{l-1} \times \omega_n^l + b_n^l \right) \quad (3)$$

In Eq. (3), ω_n^l represents the reflection weight, z_n^{l-1} refers to the value obtained from sampling in the fixed window for features of the l-1 convolution layer, and b_n^l expresses the offset value connected to the l-1 convolution layer.

The pooling layer comes after the convolution layer, which is also unique to CNN. The difference between the pooling and convolution layers is that the former does not activate the function [27]. The convolution and pooling layers can be repeatedly combined within the hidden layers as per actual application needs [28]. The fully connected layer is connected after the convolution layer and pooling layer. In general, the fully connected layer is composed of a sigmoid function or Radial Basis Function (RBF) at the last base of the network. Eq. (4) calculates the output sigmoid function.

$$y_n^l = f_1 \left(\sum_{m=1}^{N_{l-1}} y_m^{l-1} \omega_{m,n}^l + b_n^l \right) \quad (4)$$

In Eq. (4), N_{l-1} refers to the number of neurons in the output layer, $\omega_{m,n}^l$ represents the weight of the connection between the mth feature map in the upper layer with the nth neuron in the first layer, and y_m^{l-1} stands for the mth feature map in the l-1 layer.

Down-sampling layers greatly reduce the training parameter in the CNN. Thus, CNN is trained by the back-propagation algorithm. Generally, feature extraction consumes substantial time and calculation. Down-sampling can reduce computational complexity, so a CNN is constructed by using the supervised learning algorithm [29]–[31]. Then, Gradient Descent Method (GDM) is used to correct the error; each step of the gradient needs to move forward and backward to complete so that it takes a lot of time to train. Fig.4 demonstrates the algorithm flow.

CNN is a good DL method in image processing, which helps to extract the features of the image well. Its core method is to use convolution to verify the number of images and then filter the image to extract from each face and direction of the image to ensure that many features can be obtained [32], [33]. Many features can be quickly extracted from the image only by training fewer convolution parameters. Remarkably, weight sharing can optimize CNN by iteratively training the parameters in the convolution kernel with the back-propagation method.

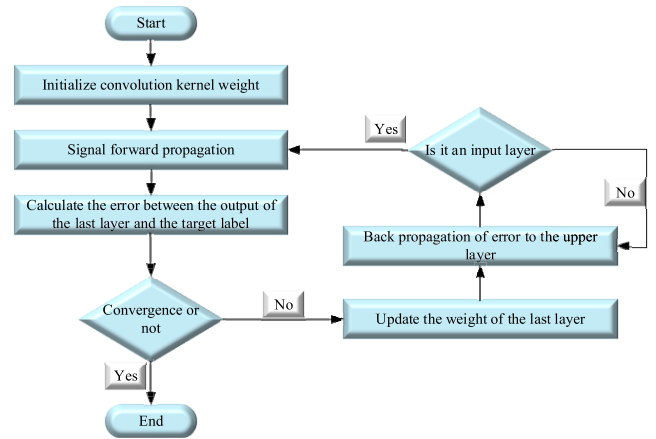


FIGURE 4. Training process of CNN.

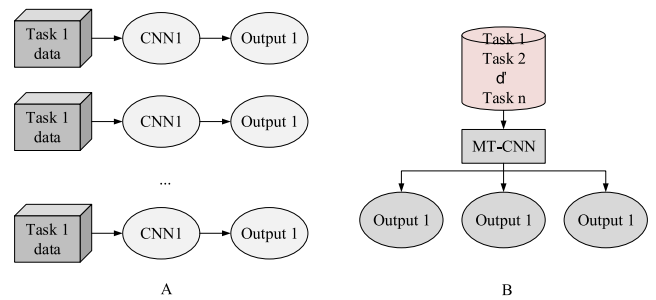


FIGURE 5. Diagrams of DL under different tasks (A: single-task DL; B: multi-task DL).

B. VEHICLE'S LOCATION AND BI-CLASSIFICATION RESEARCH BASED ON MULTI-TASK (MT)-CNN

When the scene is complex, the vehicle is not easily identifiable from an image. To this end, it is necessary to combine the image classification with the vehicle location task to recognize the vehicle. This section uses a multi-task description language model for IR and classification. MT-CNN refers to obtaining higher learning motivation through related tasks. This method can distinguish the differences between different tasks and share relevant features. Fig. 5A refers to a single task DL diagram, and Fig. 5b is a multi-task DL diagram

In the process of vehicle positioning and feature extraction, GooGleNet is adopted. The reason is that GooGleNet uses many 1*1 convolution kernels to enhance the nonlinear fitting and reduce dimension. Generally, GooGleNet has three outputs, but the last layer's output is generally optimal. Therefore, the output of the last layer is used from 22 layers of the network structure, added with the position output. Fig.6 demonstrates the MT-GooGleNet learning algorithm model.

In the vehicle positioning problem, the loss function of the network is modified to the Euclidean distance function. Assuming that there are multi-class problems in class C, and the number of contained training samples is N, the loss function can be expressed in Eq. (5).

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^C (t_k^n - y_k^n)^2 \quad (5)$$

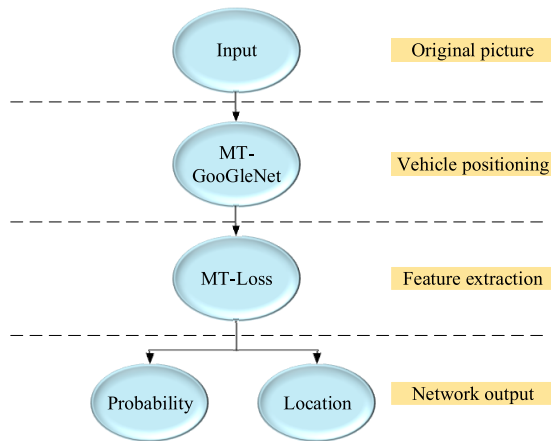


FIGURE 6. MT-GooGleNet learning algorithm.

In Eq. (5), t_k^n refers to the k_{th} dimension of the label according to the n_{th} sample, and y_k^n denotes the k_{th} output according to the n_{th} sample in the network output.

Then, GooGleNet, CaffeNet, and VGGNet are compared in vehicle positioning accuracy. There are 30,000 training pictures used in the experiment, of which 15,000 pictures are used as positive samples, and the remaining 15,000 pictures are used as artificial negative samples. Among these 30,000 pictures, 20,000 are used as training samples, and 10,000 are used as test samples. The positive and negative samples are distributed according to the same proportion. Two methods are used in the binary classification experiment. The first is the full training of MT-GooGleNet, which actually initializes all the network layers of MT-GooGleNet before the bi-classification training. The second is to fine-tune MT-GooGleNet, which specifically uses its initial parameters to initialize the convolution of all the fully connected layers and the last two layers of MT-GooGleNet before the binary classification training. The binary classification experiment is carried out after these two methods. Then, the comparative analysis is conducted.

C. IDENTIFICATION OF VEHICLE'S TYPE BASED ON MT-CNN

Then, to collect comprehensive information in the driving process, the present work uses the MT-GooGleNet model to locate the vehicle in the image and applies the vehicle area in the image after shearing as the input of the final classification model. Its essence is to use the MT-GooGleNet to locate the vehicles first and then use the positioning results as the input of the second CNN to classify the vehicle types. Then, two improvements are made on GooGleNet. The first is to enhance the data of GooGleNet, that is, to mirror the training set and then use GooGleNet to classify the original image. The second is to pre-train GooGleNet, that is, to classify GooGleNet on the original graph. The initial parameters are multi-task DL model parameters and then compare and analyze the models. The vehicle type recognition data are chosen

from Stanford University public database and Google Road recognition open dataset and are divided into 196 categories and 16,185 images in total. After data enhancement and model pre-training, the vehicle appearance signal is recognized and processed. The vehicle image is collected by road monitoring equipment, and the vehicle type and position coordinates are returned. The classified vehicle images are input into the model for unified processing. Here, “vehicle identification” refers to the identification and processing of vehicle appearance and data signals, while “vehicle position identification” refers to the road signal identification of vehicles. The two constitute the data acquisition and processing part of the proposed model. CNN is usually integrated with LSTM to maintain a certain time dependence and correlation. Further, to use the CNN network alone for image feature extraction, it is necessary to fine-tune the application and output layer of the CNN model and add the method in the Timedistributed layer of the LSTM model to the CNN model to improve the accuracy of model recognition. In this case, the model is applied to the data of multiple input time steps many times, and a series of image interpretation or image features are provided to the CNN model in turn.

V. RESULTS AND DISCUSSION

A. ANALYSIS OF THE ACCURACY OF VEHICLES' POSITIONING

Fig. 7 compares the vehicle positioning accuracy of MT-GooGleNet, CaffeNet, and VGGNet. Fig.7A reveals the comparison of positive sample accuracies, and Fig.7B indicates the comparison of negative sample error rates.

As in Fig.7A, when the overlapping rate is 0.7, the positive sample accuracies of MT-GooGleNet, MT-VGGNet, and MT-CaffeNet are 97.1%, 96.5%, and 96.5%; when the overlapping rate is 0.8, the positive sample accuracies of MT-GooGleNet, MT-VGGNet, and MT-CaffeNet are 96.3%, 95.5%, and 95.3%; when the overlapping rate is 0.9, the positive sample accuracies of MT-GooGleNet, MT-VGGNet, and MT-CaffeNet are 95.4%, 94.3%, and 94.1%. Apparently, the accuracies of these three models decline as the overlapping rate increases, while the MT-GooGleNet has the highest positive sample accuracy in these three models. As in Fig.7B, when the overlapping rate is 0.7, the negative sample error rates of MT-GooGleNet, MT-VGGNet, and MT-CaffeNet are 0.03%, 0.04%, and 0.04%; when the overlapping rate is 0.8, the negative sample error rates of MT-GooGleNet, MT-VGGNet, and MT-CaffeNet are 0.06%, 0.12%, and 0.13%; when the overlapping rate is 0.9, the negative sample error rates of MT-GooGleNet, MT-VGGNet, and MT-CaffeNet are 0.1%, 0.38%, and 0.4%. Thus, as the overlapping rate increases, the error rates of these three models are surging, while the MT-GooGleNet has the lowest negative sample error rate. To sum up, MT-GooGleNet performs best because the 1×1 filter can enhance the model nonlinearity fitting, thus promoting the model prediction accuracy.

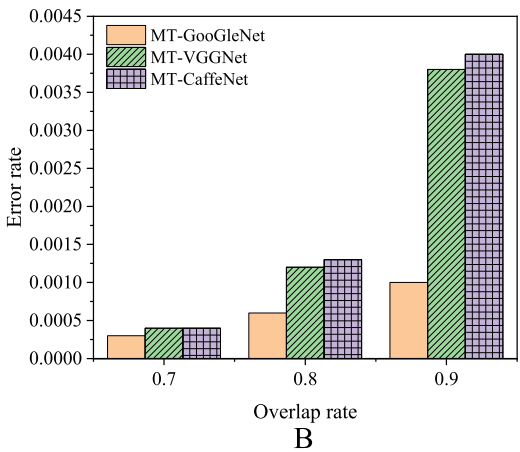
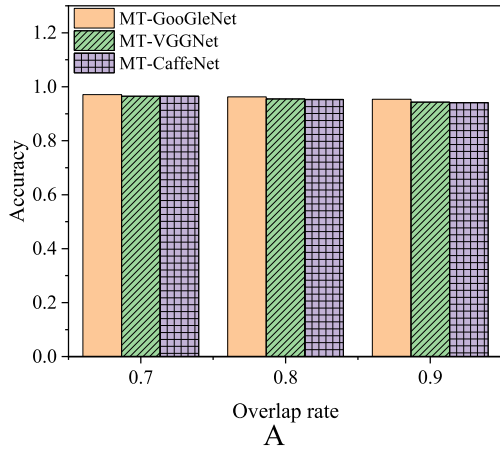


FIGURE 7. Vehicles' positioning accuracy (A: accuracy; B: error rate).

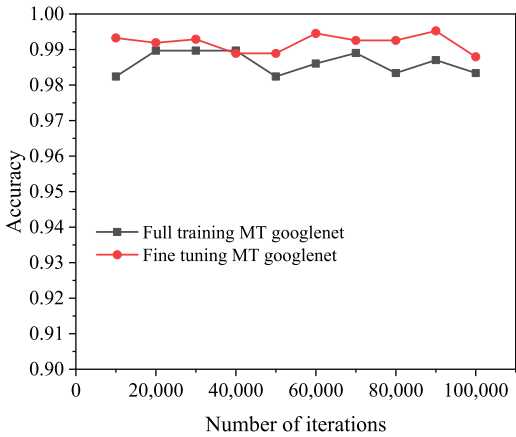


FIGURE 8. Bi-classification results of MT-GooGleNet.

B. RESULTS ANALYSIS ON BI-CLASSIFICATION OF VEHICLES

Fig.8 presents the bi-classification results of MT-GooGleNet after fine-tuning and full training.

Fig.8 manifests that both the classification accuracies of the two experiments reach more than 98% under the same recall rate. The classification accuracy of the fine-tuned MT-GooGleNet reaches up to 99.5%. The fine-tuned experimental results are significantly better than

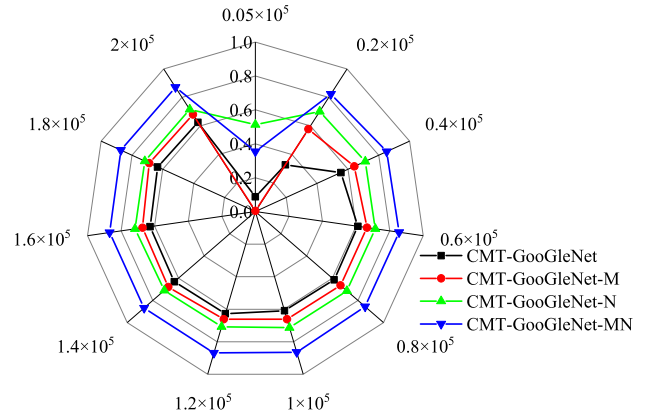


FIGURE 9. Accuracy comparison of recognition of vehicles' type based on MT-GooGleNet model.

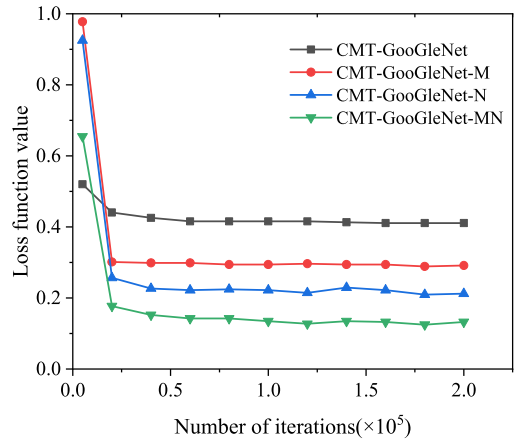


FIGURE 10. Loss function variation of cascaded MT-GooGleNet model.

the full-trained experimental results, indicating that the initialization of parameters had a great impact on the training results.

C. RECOGNITION RESULTS OF VEHICLES' TYPE

The fine-tuned GooGleNet, CaffeNet, and VGGNet are directly used for classification on the original image. Fig.9 and Fig.10 reveal the comparison results of recognition accuracy and the comparison of loss functions separately.

Fig.9 reveals that the total number of iterations is 200,000, and the initial learning rate is 0.005. Obviously, the network converges at the 10,000th iteration, and the data enhancement and pre-training can accelerate the network convergence. Meanwhile, vehicle identification has the highest accuracy of 87.2% after data enhancement and pre-training. More data help to train the network with higher accuracy in vehicle identification.

Fig.10 signifies that the cascaded MT-GooGleNet model converges the fastest after data enhancement and pre-training, and the final loss function basically converges to 0.

D. REAL-TIME COMPARISON OF THE MODELS

Next, to verify the real-time performance of the cascaded MT-GooGleNet model proposed here, the experimental

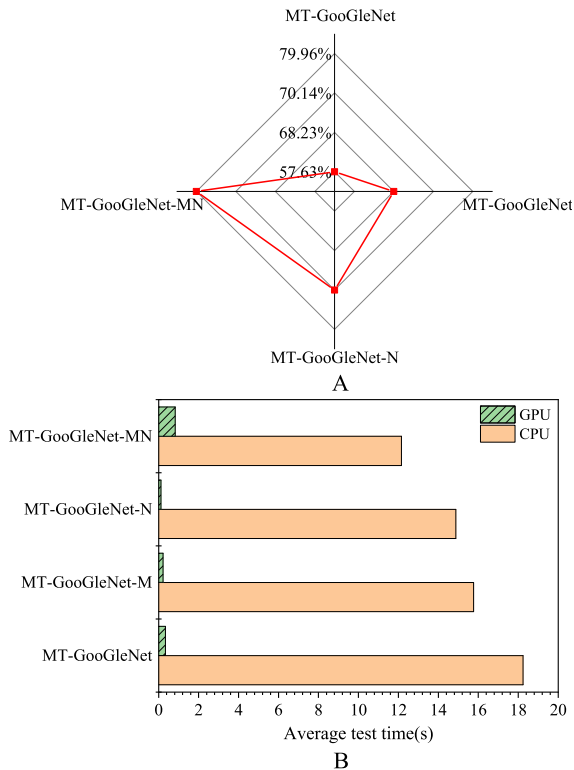


FIGURE 11. Test results of MT-GooGleNet model (A: accuracy of the testing set; B: comparison of the average test time).

environment is configured as follows: Ubuntu Operating System (OS), the Nvidia-K40 Graphics Processing Unit (GPU), 22-layer network, and 227×227 image size. The same input is used to compare different models. Fig.11 indicates the results, where Fig.11A refers to the comparison of the accuracy of the testing set, and Fig.11B displays the comparison of the average test time of the model.

Fig.11A implies that the MT-GooGleNet model has a testing set accuracy of 68.23%, with only data enhancement. The MT-GooGleNet model has a testing set accuracy of 70.14% with only pre-training; the MT-GooGleNet model has a testing set accuracy of 79.96% with both data enhancement and pre-training. Therefore, the accuracy of the MT-GooGleNet model has been greatly improved after processing. Fig.11B suggests that under the same data set, there is a difference in CPU and GPU's test time with different network structures. Thanks to GPU, the test speed of the model is significantly improved.

VI. CONCLUSION

Vehicle information acquisition from complex scenes is extremely difficult. Accordingly, the present work aims to collect comprehensive vehicle on-road driving information. Consequently, an MT-GooGleNet model is proposed to locate and recognize vehicles in images. Firstly, the experimental results of vehicle location and recognition show that the MT-GooGleNet model proposed has the best effect and high positioning accuracy in vehicle location and recognition

compared with other literature methods. Secondly, the experimental results of the fine-tuned MT GooGleNet model are significantly better than the fully trained model, indicating that the initialization of parameters has a significant impact on the training results. Thirdly, the results of vehicle identification show that after data enhancement and pre-training, the testing set accuracy of the MT-GooGleNet model is 79.96%. Therefore, the accuracy of the MT-GooGleNet model has been greatly improved after processing. The research findings provide a reference for establishing IAS in the future. The guiding principles of the research are comprehensive data collection and accurate model construction. However, the research has some deficiencies, mainly because the experimental data set used has a single source. The dataset does not contain enough image types, which is also the main limitation affected by the experimental environment. Therefore, the future research work will further optimize the system structure, optimizes the data processing process, and improves the accuracy of model recognition and the signal acquisition efficiency of the proposed traffic IAS by selecting larger datasets and deeper CNN network

REFERENCES

- [1] R. Kumar, S. Goel, V. Sharma, L. Garg, K. Srinivasan, and N. Julka, "A multifaceted vigilare system for intelligent transportation services in smart cities," *IEEE Internet Things Mag.*, vol. 3, no. 4, pp. 76–80, Dec. 2020.
- [2] M. B. Mollah, J. Zhao, D. Niyato, Y. L. Guan, C. Yuen, S. Sun, K.-Y. Lam, and L. H. Koh, "Blockchain for the Internet of Vehicles towards intelligent transportation systems: A survey," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4157–4185, Mar. 2021.
- [3] M. D. Lytras, K. T. Chui, and R. W. Liu, "Moving towards intelligent transportation via artificial intelligence and Internet-of-Things," *Sensors*, vol. 20, no. 23, p. 6945, Dec. 2020.
- [4] N. Suthanithra Vanitha, J. Karthikeyan, G. Kavitha, and K. Radhika, "Modelling of intelligent transportation system for human safety using IoT," *Mater. Today, Proc.*, vol. 33, pp. 4026–4029, Jan. 2020.
- [5] R. Dirnfeld, F. Flammini, S. Marrone, R. Nardone, and V. Vittorini, "Low-power wide-area networks in intelligent transportation: Review and opportunities for smart-railways," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–7.
- [6] B. Hu, J.-H. Lai, and C.-C. Guo, "Location-aware fine-grained vehicle type recognition using multi-task deep networks," *Neurocomputing*, vol. 243, no. 21, pp. 60–68, Jun. 2017.
- [7] H.-W. Huang, Q.-T. Li, and D.-M. Zhang, "Deep learning based image recognition for crack and leakage defects of metro shield tunnel," *Tunnelling Underground Space Technol.*, vol. 77, pp. 166–176, Jul. 2018.
- [8] C. M. J. M. Dourado, S. P. P. da Silva, R. V. M. da Nobrega, P. P. R. Filho, K. Muhammad, and V. H. C. de Albuquerque, "An open IoHT-based deep learning framework for online medical image recognition," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 541–548, Feb. 2021.
- [9] J. Huixian, "The analysis of plants image recognition based on deep learning and artificial neural network," *IEEE Access*, vol. 8, pp. 68828–68841, 2020.
- [10] J. Toldinas, A. Venčkauskas, R. Damaševičius, Š. Grigaličnas, N. Morkevičius, and E. Baranauskas, "A novel approach for network intrusion detection using multistage deep learning image recognition," *Electronics*, vol. 10, no. 15, p. 1854, Aug. 2021.
- [11] K. Zhao, T. He, S. Wu, S. Wang, B. Dai, Q. Yang, and Y. Lei, "Application research of image recognition technology based on CNN in image location of environmental monitoring UAV," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–11, Dec. 2018.
- [12] C. Chen, O. Li, C. Tao, A. J. Barnett, J. Su, and C. Rudin, "This looks like that: Deep learning for interpretable image recognition," 2018, *arXiv:1806.10574*.

- [13] H. Zhang and X. Lu, "Vehicle communication network in intelligent transportation system based on Internet of Things," *Comput. Commun.*, vol. 160, pp. 799–806, Jul. 2020.
- [14] A. A. Brincat, F. Pacifici, S. Martinaglia, and F. Mazzola, "The Internet of Things for intelligent transportation systems in real smart cities scenarios," in *Proc. IEEE 5th World Forum Internet Things (WF-IoT)*, Apr. 2019, pp. 128–132.
- [15] W. Li, J. Zhu, Y. Zhang, and S. Zhang, "Design and implementation of intelligent traffic and big data mining system based on Internet of Things," *J. Intell. Fuzzy Syst.*, vol. 38, no. 2, pp. 1967–1975, Feb. 2020.
- [16] Z. Lv, S. Zhang, and W. Xiu, "Solving the security problem of intelligent transportation system with deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4281–4290, Jul. 2020.
- [17] S. Kaffash, A. T. Nguyen, and J. Zhu, "Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis," *Int. J. Prod. Econ.*, vol. 231, Jan. 2021, Art. no. 107868.
- [18] X. Ke and Y. Zhang, "Fine-grained vehicle type detection and recognition based on dense attention network," *Neurocomputing*, vol. 399, pp. 247–257, Jul. 2020.
- [19] C. Kwan, D. Gribben, B. Chou, B. Budavari, J. Larkin, A. Rangamani, T. Tran, J. Zhang, and R. Etienne-Cummings, "YOLOs," *Electron.*, vol. 9, no. 6, p. 1014, 2020.
- [20] P. Li, M. Abdel-Aty, Q. Cai, and Z. Islam, "A deep learning approach to detect real-time vehicle maneuvers based on smartphone sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 1–10, Apr. 2020.
- [21] K. Yu, L. Lin, M. Alazab, L. Tan, and B. Gu, "Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4337–4347, Jul. 2021.
- [22] H. Chen, A. Chen, L. Xu, H. Xie, H. Qiao, Q. Lin, and K. Cai, "A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources," *Agricult. Water Manage.*, vol. 240, Oct. 2020, Art. no. 106303.
- [23] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, May 2018.
- [24] F. O. Ozkok and M. Celik, "A hybrid CNN-LSTM model for high resolution melting curve classification," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103168.
- [25] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, and N. Z. Jhanjhi, "A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images," in *Proc. IRBM*, Jun. 2021.
- [26] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, Mar. 2021.
- [27] F. Bertoni, G. Citti, and A. Sarti, "LGN-CNN: A biologically inspired CNN architecture," *Neural Netw.*, vol. 145, pp. 42–55, Jan. 2022.
- [28] J. Siavashi, A. Najafi, M. Ebadi, and M. Sharifi, "A CNN-based approach for upscaling multiphase flow in digital sandstones," *Fuel*, vol. 308, Jan. 2022, Art. no. 122047.
- [29] G. Vrbančić and V. Podgorelec, "Efficient ensemble for image-based identification of pneumonia utilizing deep CNN and SGD with warm restarts," *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115834.
- [30] A. Moradzadeh, H. Teimourzadeh, B. Mohammadi-Ivatloo, and K. Pourhossein, "Hybrid CNN-LSTM approaches for identification of type and locations of transmission line faults," *Int. J. Electr. Power Energy Syst.*, vol. 135, Feb. 2022, Art. no. 107563.
- [31] M. F. Aslan, K. Sabanci, and A. Durdu, "A CNN-based novel solution for determining the survival status of heart failure patients with clinical record data: Numeric to image," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102716.
- [32] Ö. Inik, M. Altok, E. Ülker, and B. Koçer, "MODE-CNN: A fast converging multi-objective optimization algorithm for CNN-based models," *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107582.
- [33] J. Xie, K. Hu, Y. Guo, Q. Zhu, and J. Yu, "On loss functions and CNNs for improved bioacoustic signal classification," *Ecol. Informat.*, vol. 64, Sep. 2021, Art. no. 101331.

• • •