

A hierarchical RCNN for vehicle and vehicle license plate detection and recognition

Chunling Tu¹, Shengzhi Du²

¹Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa

²Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa

Article Info

Article history:

Received Aug 30, 2020

Revised Jun 17, 2021

Accepted Jun 30, 2021

Keywords:

Deep learning

Region convolution neural network

Vehicle detection

Vehicle license recognition

ABSTRACT

Vehicle and vehicle license detection obtained incredible achievements during recent years that are also popularly used in real traffic scenarios, such as intelligent traffic monitoring systems, auto parking systems, and vehicle services. Computer vision attracted much attention in vehicle and vehicle license detection, benefit from image processing and machine learning technologies. However, the existing methods still have some issues with vehicle and vehicle license plate recognition, especially in a complex environment. In this paper, we propose a multivehicle detection and license plate recognition system based on a hierarchical region convolutional neural network (RCNN). Firstly, a higher level of RCNN is employed to extract vehicles from the original images or video frames. Secondly, the regions of the detected vehicles are input to a lower level (smaller) RCNN to detect the license plate. Thirdly, the detected license plate is split into single numbers. Finally, the individual numbers are recognized by an even smaller RCNN. The experiments on the real traffic database validated the proposed method. Compared with the commonly used all-in-one deep learning structure, the proposed hierarchical method deals with the license plate recognition task in multiple levels for sub-tasks, which enables the modification of network size and structure according to the complexity of sub-tasks. Therefore, the computation load is reduced.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Chunling Tu

Department of Computer Systems Engineering, Tshwane University of Technology

2 Aubrey Matlakala St, Soshanguve, Pretoria, South Africa

Email: duc@tut.ac.za

1. INTRODUCTION

Vehicle and vehicle license plate tracking and recognition management systems play an important role in an intelligent transportation system, it is commonly used in automatic driving, vehicle theft prevention, access control security, high-efficiency roadway services, and so on. Most of these systems are based on traffic image or video analysis, where computer vision techniques were commonly employed. A huge recent development was found with the employment of machine learning and deep learning techniques, which demonstrate higher classification accuracy and robustness. Both deep learning and computer vision methods [1]-[15] were developed in various vehicle-relevant applications, such as vehicle detection, vehicle classification, vehicle plate recognition, and road condition monitoring [16]. Comparing with computer vision methods, the newly developed deep learning techniques have some advantages on the generalization capacity and robustness to uncertainties, noise, and occlusion in images, in the cost of higher computation load and demand on sample set size. Several methods were proposed to detect vehicle and vehicle license plates based on these newly developed techniques, such as the dirty number plate detection system based on

you only look once (YOLO) [1], the license number plate recognition system based on convolutional neural network (CNN), and recurrent neural network (RNN) [2], the technique to improve the car plate recognition rate based on support vector machines (SVM) [3], the vehicle detection system based on region convolutional neural network (RCNN) [4], the vehicle detection and counting system using the convolutional regression neural network (CRNN) [5].

However, there are some challenges to deep learning-based techniques in vehicle and vehicle license plate detection and recognition. Firstly, fake objects from a complex environment around the vehicle reduce the accuracy of detection by confusing the deep neural networks, buildings, trees, pedestrians, and other objects surrounded the vehicles. Secondly, the weather and illumination blur the vehicles and license plate, strong sunlight, foggy, rainy, and shadow. Thirdly, the human factor, speed of the vehicle, driver behaviors. All of the above make the difficult to achieve accurate and efficient detection.

In this paper, we proposed a hierarchical RCNN system for vehicle detection and license plate recognition under complex traffic backgrounds. The system contains four steps: i) a high-level RCNN to detect vehicles from traffic images/video, ii) a lower level RCNN is used to locate the license plate from the detected vehicle regions, iii) a smaller RCNN is trained to detect individual letter from the detected license plate, iv) finally an RCNN classified is employed to recognize the individual letter. To reduce computation, the training data were cropped into smaller blocks. The main innovation of the proposed method is manifested in the idea of hierarchical structure with multiple-level networks for sub-tasks. Differentiating from the common all-in-one deep learning structure, the proposed method considers the license plate recognition as multiple sub-tasks, which enables the optimization of network depth and structure for each sub-task. Each RCNN is designed with careful consideration of the complexity of the problem to be solved in the corresponding level, so an RCNN of appropriate depth is chosen for each step. The experiment results show the proposed system has higher speed and accuracy.

The rest of the paper is organized as follows. Section 2 introduces the relative work. Section 3 explains the proposed method. Section 4 demonstrates the experiments and analysis. Section 5 concludes the paper and indicates future work.

2. RESEARCH METHOD

2.1. Vehicle and license plate detection

Vehicle detection has been popularly studied in literature based on computer vision, which is the fundamental stage for an automatic driver system. There are many algorithms developed. According to the literature, vehicle detection is categorized into moving vehicle detection and static vehicle detection. In this paper, we focus on only moving vehicle detection. The features, background subtractions, frame difference, optical flow, machine learning, and combined methods are used to detect moving detection. A method was proposed [6] using optical flow to detect a moving vehicle. The color and texture features were used to reduce the effects from the complex background, and the fuzzy background subtraction was developed [7] for moving vehicle detection. [8] improved the frame difference method for moving vehicle detection using image contrast and morphological filter. The optical flow was combined with k-nearest neighbor (KNN) to classify the different kinds of moving vehicles [12]. An anti-tracking system [17] was proposed to detect the vehicle based on Haar features and modified the adaptive boosting algorithm.

The vehicle license plate, as an ID, can uniquely identify vehicles. So the automatic license plate detection and recognition is one of the important tasks in an intelligent driving system. Various methods have been developed, a method [10] was developed to detect vehicle plates using salient features, with a success rate of 93.1% reported. The rear vehicle lights were used to detect the range of license plate and then identify the license plated using the histogram algorithm [11], which was only validated by vehicles from five countries. The average success rate is 90.4%. An automatic license plate recognition [12] was proposed using updated YOLO, with a license plate recognition accuracy of 78%. However, the accurate rate still is an issue for real applications. And most of these methods have high computational requirements due to the system's need to search the vehicles and vehicle license plate from large and high-resolution images or videos that were captured from the road.

2.2. Region convolution neural network (RCNN)

Deep learning is one of the algorithms of the machine learning method that was developed based on the neural network [18]. Various deep learning networks structure have been depleted, such as AlexNet [19], Over feat [20], GoogLeNet [21], ResNet [22]. RCNN is the most popular model of deep learning because of the significant success in image processing. The RCNN is made up of neurons that have tunable weights and bias. The input data can be images, the layers of RCNN have 3 dimensions (width, height, and depth). RCNNs are widely used in vehicle and vehicle number plate detection. For instance, Brody [13] detected the

lanes and vehicles using CNN. The front collision was predicated using CNN with adaptive region-of-interest [14]. Several recent studies [15], [23] developed one system for license plate recognition based on RCNN. In this paper, a hierarchical RCNN will be employed to obtain high accuracy and fast speed for vehicle license plate detection.

3. PROPOSED METHOD

Deep learning techniques are commonly a hunger for computation resources. Compared to CNN, the RCNN gets the advantage of higher computation efficiency. The traditional all-in-one method did not leave much space for modifications according to the real-world scenario, such as the task complexity, features to be considered, image qualities, and the relationship among sub-tasks. This paper is intending to propose a hierarchical structure to separate the all-in-one RCNN into multiple levels, with each layer focuses on a single sub-task. Therefore, the network in each layer will be developed according to the real demand of the sub-task only. This will lead to the lower complexity of the total task, which results in a smaller network size, lower computation load, and higher total accuracy.

Using the vehicle plate recognition applications, this paper employs RCNN for 3 level sub-tasks, vehicle detection, plate detection, and character level detection and recognition. These 3 tasks have different complexity on the background and object features. The main idea is to consider RCNNs with different complexity levels for these tasks. Figure 1 depicts the proposed hierarchical structure of the solution for vehicle-plate-character detection and identification, where multiple RCNNs with feasible complexity are considered for different tasks. The hierarchical structure is not only for RCNN, various pre-trained deep neural networks can be considered to replace the RCNN. In this paper, the RCNN is considered just as an example to show the idea.

The vehicle level RCNN is the most complex one because the vehicles are commonly found in busy traffic contexts when smart traffic monitoring systems are concerned. Detecting vehicles with acceptable accuracy is critical for total system performance. In this level, a 15 layer RCNN is employed, as shown in Figure 2, where the RCNN has 3 folds of convolutional-pooling layers, which are pre-trained.

After the vehicle is detected by the vehicle level RCNN, we can get the region of the vehicle. The plate level RCNN only focuses on this region, instead of the whole image. This greatly reduces the computation load. Meanwhile, this strategy can avoid mis-hit in plate detection, because other potential candidates in the background are excluded.

Because the problem to detect a license plate from a vehicle is relevantly simpler than detecting vehicles from the background, this plate level RCNN can be smaller (having fewer layers). Figure 3 shows the 12-layer structure of this RCNN, where only 2 folds of convolution-ReLU-Pooling layers are contained. Both the input and the final layers are the same as the vehicle level RCNN.

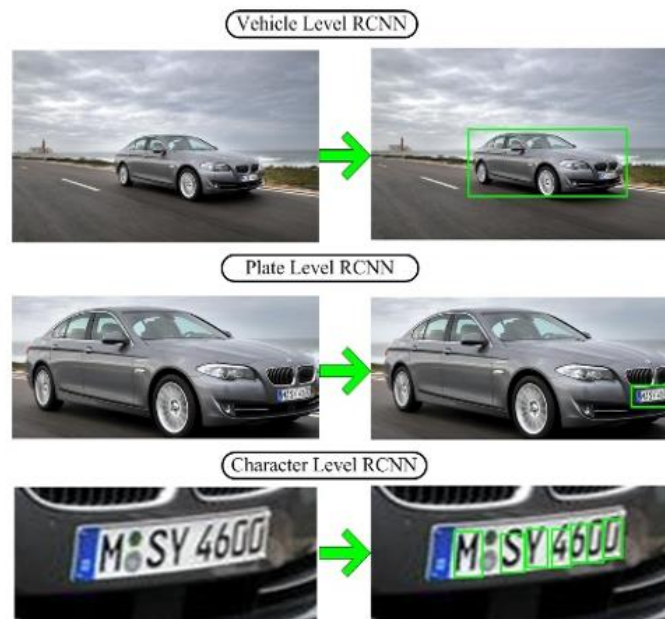


Figure 1. The hierarchical structure of the proposed RCNN vehicle identification system

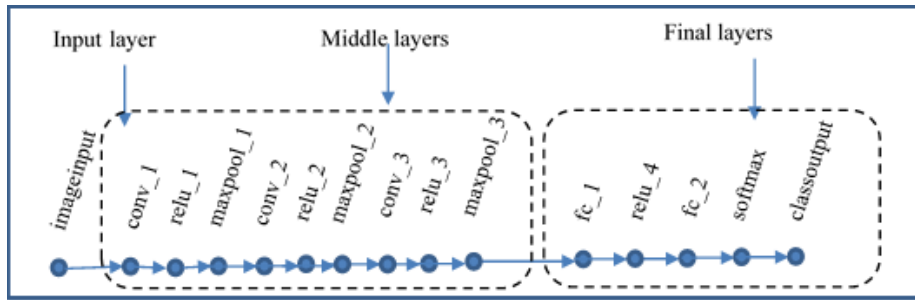


Figure 2. The structure of the vehicle level RCNN

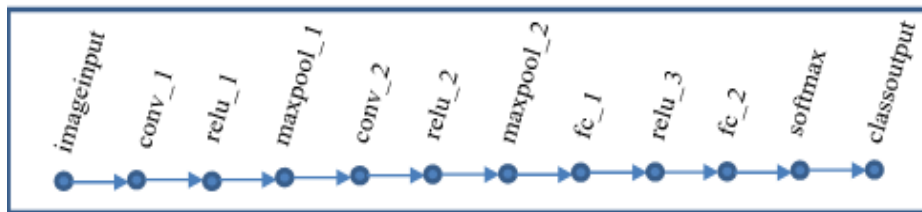


Figure 3. The structure of RCNN in the character and the plate levels

The features used to detect and recognize characters from the plate image are even less. So theoretically the character level RCNN can be smaller than the plate level. However, the plate level RCNN is already very shallow (only 2 convolution layers), it is not feasible to reduce the layers further to avoid the low fitting capacity of the RCNN. The character level RCNN takes the same structure as the plate level one. It should be noted that if the vehicle level deep network has more convolution layers, one can expect the character level RCNN can be smaller than the plate level.

4. EXPERIMENTS AND DISCUSSION

To validate the proposed method, experiments are designed based on the public accessible Canadian Institute for Advanced Research, 10 classes (CIFAR-10) database [24].

4.1. Experiment environment

All the experiments in this paper are completed on a Lenovo Thinkpad X1 laptop, with 16 GB RAM, an Intel(R) Core(TM) i7-8550 CPU at 1 Ghz. The operating system is windows 10 x64. The software is Matlab R2019a with the following toolboxes: computer vision, image processing, deep learning, and statistics and machine learning. The RCNNs are modified and trained based on the CIFAR-10 network [25].

4.2. Reduce the layers of RCNN

Associating the complexity of RCNN to the task is the main advantage of the proposed strategy. This experiment aims to demonstrate the feasibility of the deduction of the CIFAR-10 Network. The original CIFAR-10 Network (15-layer RCNN) has 1 image input layer, 3 folds of Convolution-ReLU-Pooling layers as the middle layer, and the final layers, as shown in Figure 2. The first experiment is to remove 1 fold of the ConvolutionReLU-Pooling layers from the middle layers. The modified network has the structure as shown in Figure 3.

After training using the CIFAR-10 dataset, the weight of the first convolution layer for the original network (3 convolution layers) and the modified network 12-layer RCNN (2 convolution layers) are shown in Figure 4 and Figure 5 respectively. From the weights, one finds that both of the networks have captured the basic features of the images in the dataset. Therefore, the modification of the structure does not significantly damage the feature extraction capacity.

The next experiment depicts the further reduction of the middle layers to a single convolution layer (9-layer RCNN). The weights of the single convolution network shown in Figure 6 shows the strong random distribution, which means the network failed to capture the image features. This is because the removal of two convolution layers has damaged the learning capacity.

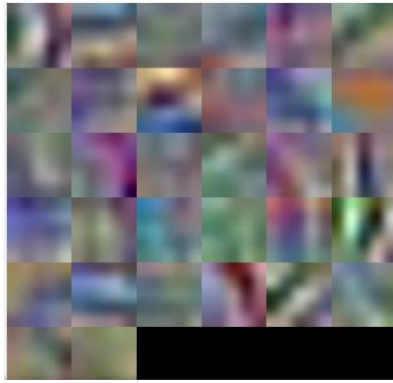


Figure 4. The weight of the first convolution layer in the 15-layer RCNN



Figure 5. The weight of the first convolution layer in the 12-layer RCNN

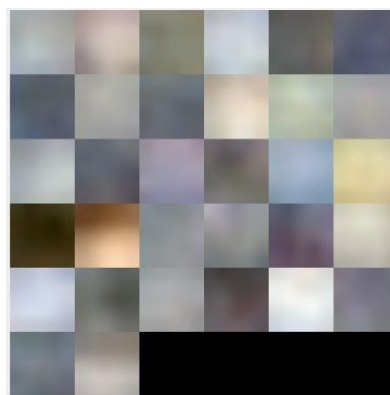


Figure 6. The weight of the first convolution layer in the 9-layer RCNN

The accuracy during the training iterations in Figure 7 confirmed this point, where the 12-layer and 15-layer RCNNs have similar accuracy. This means the reduction of the layers does not significantly affect the learning capacity. However, the accuracy of the 9-layer RCNN was significantly lowered. Considering that the dataset of CIFAR-10 has 10 classes, the accuracy of 10% obtained by the 9-layer RCNN is just a random guess.

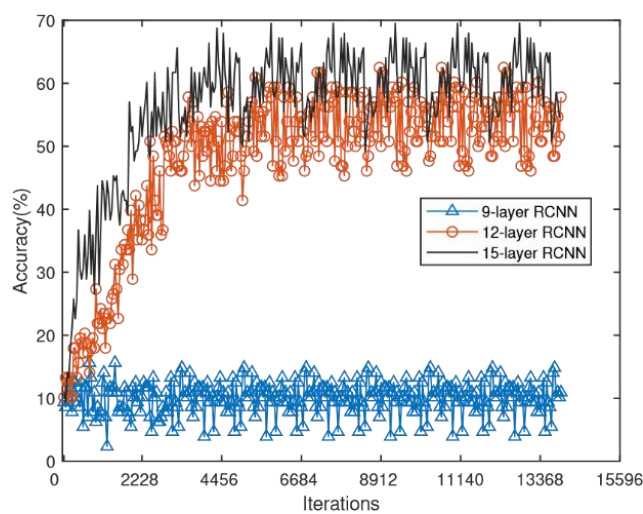


Figure 7. The mini-batch accuracy during the training of RCNN

It should be noted that, due to the advantage of the transfer learning strategy, the RCNNs do not need to reach 100% accuracy when training the middle layers, as long as the middle layers can capture the image features. Table 1 shows the accuracy of the 3 RCNNs after the training of middle layers using the 10 classes. When the computation load is concerned, the modified RCNNs (12-layer and 9-layer networks) have higher efficiency. Considering both accuracy and time efficiency, one finds the 12-layer RCNN can be considered as the lower-level classifiers in the hierarchical structure in Figure 1.

Table 1. Comparison of accuracy and computation of the RCNNs

RCNN	15-layer	12-layer	9-layer
Training Time (min)	49	44	33
Test Set Accuracy (%)	56.7	52.5	10

4.3. Vehicle license plate detection and recognition

This experiment focuses on the specific application, the vehicle license plate detection and recognition. Table 2 shows the accuracy of each RCNN in the proposed structure. From Table 2, one finds that the RCNNs in vehicle and plate levels have similar accuracy, although the classification capacity was lowered when a convolution layer was removed. This is because the problem complexity for plate detection is lower than the vehicle detection, therefore, a smaller RCNN also can get similar accuracy. This also means there are some spaces to modify the all-in-one RCNN without significant damage to the accuracy. The character level accuracy is lower than the other two levels although the characters have fewer features than the plate and vehicle. This is because the classification problem in the character level becomes a multiple-class (36 classes) problem. The final layers of the character level should be improved, and more training sets should be considered. The experiments demonstrated the reduction of RCNNs and the performance of the total system, which validated the proposed strategy.

Table 2. Accuracy of the proposed method for detection and recognition

RCNNs	Vehicle Level	Plate Level	Character Level
Accuracy(%)	98.5	97	85

5. CONCLUSION

This paper proposed a hierarchical strategy for vehicle and vehicle license plate detection and identification based on RCNN. The complexity of the RCNN was associated with the tasks. In this way, multiple complex level RCNNs can be employed in the same system. As an example, a sample vehicle detection system was developed for smart traffic monitoring, thereafter the license plate recognition RCNN was considered for vehicle identification.

The vehicle level RCNN with the highest complexity was employed to detect the vehicle from the background. In this RCNN, the task can be considered as a two-class (vehicle and background) classification problem. The detected vehicle region (a portion of the original image) was then inputted to the license plate level RCNN, where the license plate was detected. This is also a two-class classification problem (license plate and vehicle body as background). In this way, the computation load is greatly reduced. Furthermore, the disturbances from outside of the vehicle were excluded, which improved the success rate of the plate level RCNN. Finally, the detected license plate became the input of the character level RCNN, where the individual characters were detected and classified. This RCNN solved a multiple class classification problem, where the numbers ('0' to '9') and letters('a' to 'z') are the classification targets. The future work includes: (1) separate the total task into more levels and design the network in each level according to the specific features in the corresponding sub-task; (2) improve the final layers for the character level RCNN to improve the accuracy of plate recognition.




REFERENCES

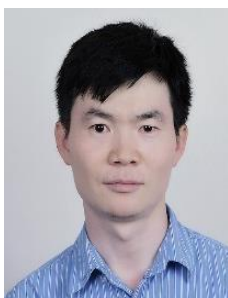
- [1] R. Laroca *et al.*, "A robust real-time automatic license plate recognition based on the YOLO detector," *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1-10, doi: 10.1109/IJCNN.2018.8489629.
- [2] T. K. Cheang, Y. S. Chong, and Y. H. Tay, "Segmentation-free vehicle license plate recognition using ConvNet-RNN," 2017, *arXiv:1701.06439*.
- [3] M. S. Faraq, M. M. M. El Din, and H. A. Elshenbary, "Parking entrance control using license plate detection and recognition," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 15, no. 1, pp. 476-483, Jul. 2019, doi: 10.11591/ijeecs.v15.i1.pp476-483.




- [4] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining," *Sensors*, vol. 17, no. 2, pp. 1-17, 2017, Art. no. 336, doi: 10.3390/s17020336.
- [5] H. Tayara, K. Gil Soo, and K. T. Chong, "Vehicle detection and counting in high-resolution aerial images using convolutional regression neural network," in *IEEE Access*, vol. 6, pp. 2220-2230, 2018, doi: 10.1109/ACCESS.2017.2782260.
- [6] Y. Chen and Q. Wu, "Moving vehicle detection based on optical flow estimation of edge," *2015 11th International Conference on Natural Computation (ICNC)*, 2015, pp. 754-758, doi: 10.1109/ICNC.2015.7378085.
- [7] X. Lu, T. Izumi, T. Takahashi, and L. Wang, "Moving vehicle detection based on fuzzy background subtraction," *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2014, pp. 529-532, doi: 10.1109/FUZZ-IEEE.2014.6891578.
- [8] S. Sarikan and A. M. Ozbayoglu, "Anomaly detection in vehicle traffic with image processing and machine learning," *Procedia Computer Science*, vol. 140, pp. 64-69, 2018, doi: 10.1016/j.procs.2018.10.293.
- [9] B. Huval *et al.*, "An empirical evaluation of deep learning on highway driving," 2015, *arXiv:1504.01716*.
- [10] M. R. Asif, Q. Chun, S. Hussain, M. S. Fareed, and S. Khan, "Multinational vehicle license plate detection in complex backgrounds," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 176-186, 2017, doi: 10.1016/j.jvcir.2017.03.020.
- [11] A. Kumar, A. Sharma, and R. K. Singla, "An efficient license plate text extraction technique," *Ambient Communications and Computer Systems*, vol. 904, pp. 487-495, 2019, doi: 10.1007/978-981-13-5934-7_43.
- [12] Hendry and R. C. Chen, "Automatic license plate recognition via sliding-window darknet-YOLO deep learning," *Image and Vision Computing*, vol. 87 pp. 47-56, 2019, doi: 10.1016/j.imavis.2019.04.007.
- [13] J. Pyo, J. Bang and Y. Jeong, "Front collision warning based on vehicle detection using CNN," *2016 International SoC Design Conference (ISOCC)*, 2016, pp. 163-164, doi: 10.1109/ISOCC.2016.7799842.
- [14] C. Gou, K. Wang, Y. Yao and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1096-1107, Apr. 2016, doi: 10.1109/TITS.2015.2496545.
- [15] M. A. Rafique, W. Pedrycz, and M. Jeon, "Vehicle license plate detection using region-based convolutional neural networks," *Soft Computing*, vol. 22, pp. 6429-6440, 2018, doi: 10.1007/s00500-017-2696-2.
- [16] T. Siriborvornratanakul, "An automatic road distress visual inspection system using an onboard in-car camera," *Advances in Multimedia*, vol. 2018, 2018, Art. no. 2561953, doi: 10.1155/2018/2561953.
- [17] O. Oheka and C. Tu, "Fast and Improved Real-Time Vehicle Anti-Tracking System," *Applied Sciences*, vol. 10, no. 17, 2020, doi: 10.3390/app10175928.
- [18] Y. Le Cun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015, doi: 10.1038/nature14539.
- [19] F. N. Landola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*.
- [20] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.
- [21] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.
- [22] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [23] P. Marzuki, F. Radzi, Y. C. Wong, N. A. Hamid, N. Ali, and M. M. Ibrahim, "A design of license plate recognition system using convolutional neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 3, pp. 2196-2204, 2019, doi: 10.11591/ijece.v9i3.pp2196-2204.
- [24] A. Krizhevsky, V. Nair, and G. Hinton. "The CIFAR-10 dataset." Toronto.edu. <https://www.cs.toronto.edu/~kriz/cifar.html> (accessed Dec. 17, 2020).
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, vol. 1, pp. 1097-1105.

BIOGRAPHIES OF AUTHORS



Chunling Tu    received a bachelor degree in computer science from Tianjin University of Technology and Education, China in 2002; MTech and MSc degrees in Electrical Engineering from Tshwane University of Technology (South Africa) and ESIEE Paris University (France) in 2010; DTech and Ph.D. degrees of Electrical Engineering from the Tshwane University of Technology and University Paris East, France in 2015. Her research interests include image processing, AI, industrial control, machine learning, deep learning, and pattern recognition. She can be contacted at email: duc@tut.ac.za.



Shengzhi Du    received an M.S. degree in control theory and control engineering from Tianjin Poly Technology University, Tianjin, China, in 2001 and the Ph.D. degree in control theory and control engineering from Nankai University, Tianjin, China, in 2005. He is currently a professor in the French South Africa Institute of Technology (F'SATI), Tshwane University of Technology, South Africa. His research interests include computer vision, AI, pattern recognition, and human in loop systems. He can be contacted at email: dus@tut.ac.za.