A Novel Approach of Traffic Density Estimation Using CNNs and Computer Vision

Luong Anh Tuan Nguyen and Thanh Xuan Ha

Abstract — In modern life, we face many problems, one of which is the increasingly serious traffic jam. The cause is the large volume of vehicles, inadequate infrastructure and unreasonable distribution, and ineffective traffic signal control. This requires finding methods to optimize traffic flow, especially during peak hours. To optimize traffic flow, it is necessary to determine the traffic density at each time in the streets and intersections. This paper proposed a novel approach to traffic density estimation using Convolutional Neural Networks (CNNs) and computer vision. The experimental results with UCSD traffic dataset show that the proposed solution achieved the worst estimation rate of 98.48% and the best estimation rate of 99.01%.

Key words — Computer Vision, Convolutional Neural Networks (CNNs), Traffic Density Estimation.

I. INTRODUCTION

Traffic density estimation provides information to Intelligent Transport Systems (ITS) for road planning, intelligent routing, traffic control, scheduling, etc. [1]. Accurate estimation of traffic density is essential for the development of automated signaling and early warning systems, development of statistical, planning, and security applications. Furthermore, traffic density data can be used to help drivers choose the optimal way between multiple routes [2].

In the past, the methods of estimating traffic density were usually vehicle counting methods. Pang et al. [3] introduced a method to count the number of vehicles in the presence of multiple vehicles in the traffic image. Chen et al. [4] developed a new system for counting vehicles in dark environments using their headlights. Pornpanomchai et al. [5] studied video vision for vehicle counting. Mohana et al. [6] studied the counting of vehicles in real-time. Zhao and Wang [7] studied vehicle counting in mixed traffic zones. Today's traffic management systems use image and video processing techniques to extract information from surveillance videos on roads [8].

With the rapid development of computer vision and digital image processing technology, image and video-based traffic flow detection system has become more and more developed and intelligent. Due to these benefits, image and video-based vehicle detection technology are becoming more and more important for ITS [9]. The approach proposed by Gupta et al. [10] is based on traffic calculation by comparing two images,

a reference image, and a real traffic image. Kanojia [11] proposed another method to control traffic signals using image processing techniques. He first selects a reference image that is an image with no or fewer cars and matches the actual image to that reference image each time. Based on the matching rate, the traffic lights are controlled.

In recent years, artificial intelligence and computer vision have made remarkable achievements, which are applied in all fields such as security, economy, commerce, education, etc. Image recognition is intelligent functions from the combination of computer vision and artificial intelligence.

Deep Learning developed from neural networks has many outstanding achievements in the field of computer vision. Convolutional Neural Networks (CNNs) is one of the famous and widely deep learning models in the field of computer vision.

The combination of artificial intelligence and computer vision makes traffic density estimation studies based on image processing techniques more and more effective. Celil Ozkurt and Fatih Camci [12] proposed a method to estimate traffic density and classify vehicles using neural networks. Julian Nubert et al. [13] proposed the work of estimating traffic density by convolutional neural networks (CNNs). The goal of this project is to introduce and present a machine learning application that aims to improve the quality of life of Singaporeans. R. Kaviani et al. [14] propose a method to estimate traffic density based on Topic model, the Topic model is a quite popular method in the natural language processing field.

From the studies of related works, the authors propose a traffic density estimation model for traffic video clips based on computer vision combined with convolutional neural networks (CNNs). The experimental results with UCSD traffic dataset show that the proposed solution achieved the worst estimation rate of 98.48% and the best estimation rate of 99.01%.

The remainder of the paper is organized as follows. Section II presents the design and implementations of the proposed solution. In section III, the numerical results of the experiment are illustrated. Finally, Section IV concludes this paper and figures out the future works.

II. PROPOSED SOLUTION

The system architecture including 3 modules is shown in Fig. 1.

Submitted on August 04, 2021.
Published on August 31, 2021.
Luong Anh Tuan Nguyen, IT Faculty, Vietnam Aviation Academy, Vietnam.
(corresponding e-mail: nlatuan@vaa.edu.vn)

Thanh Xuan Ha, Bao An Technology Company, Ho Chi Minh City, Vietnam.
(e-mail: hathanhxuan@gmail.com)

The first module is the “cut and discards similar frames” module which cuts the input traffic video clip into frames, then discards adjacent similar frames.

The second module is the “Pre-Processing” module which uses OpenCV library and image segmentation algorithm [15] to remove non-vehicle objects in traffic images.

The third module is the CNN module which estimates traffic density from the output image of the second module.

Fig. 1. System architecture.

A. “Cut and Discard Similar Frames” Module

This module cuts the input traffic video clip into frames, each frame is a traffic image. However, in reality, adjacent frames have almost no content change, so to reduce processing time for the entire system, it is necessary to must discard frames with similar content.

This module is implemented with the following steps:

- 1) Convert K and K+1 frames from color image to grayscale image by averaging method $(R+G+B)/3$.
- 2) Get the histogram of the grayscale image of the K and K+1 frames, then convert these histograms into a one-dimensional array consisting of 256 elements, called vector K and K+1.
- 3) Calculate the similarity between 2 vectors k and vector k+1 according to the following correlation coefficient [16]:

$$R_{xy} = \left| \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2 \sum_{i=1}^N (y_i - \bar{Y})^2}} \right|$$

Where, $X=(X_1, X_2, \dots, X_n)$ is the vector K, $Y=(Y_1, Y_2, \dots, Y_n)$ is the Vector K+1 and $N=256$. This correlation coefficient has a value from 0 to 1. If this correlation coefficient is greater than or equal to 0.95, then frame K and frame K+1 are similar, and frame K+1 will be discarded.

B. “Pre-Processing” Module

The purpose of this module is to detect vehicle objects on

traffic images and retain these vehicle images, filtering out non-vehicle images. Therefore, the output of this module is an image of only vehicle objects, this output image has been discarded for non-vehicle objects and serves as an input to CNN module.

This module uses OpenCV library and image segmentation algorithm to remove non-vehicle objects in traffic images. Fig. 2 illustrates the input image and output image in this module, (a) input image, (b) output image.

Fig. 2. a) input image, (b) output image.

C. CNN Module

This module is shown in Fig. 3. This module includes the following layers:

- 1) The convolutional layer extracts features from input images. Convolution preserves the relationship between pixels by learning image features using small squares of input data.
- 2) ReLU layer: ReLU stands for Rectified Linear Unit for a non-linear operation. The output is $f(x) = \max(0, x)$. Since the real-world data would be non-negative linear values.
- 3) Pooling layers make reduce the number of parameters when the images are too large. The output of this pooling layer is a pooled featured map. Max Pooling and Average Pooling are two widely used pooling techniques.

Fig. 3. CNN Module.

- 4) Fully Connected Layer: The flatten layer is used to transfer pooled feature map into a one-dimensional vector. The flattened matrix goes through a fully connected layer to classify emotions. In the proposed method, using the softmax activation function in the output layer to represent a categorical distribution over class labels, and obtaining the probabilities of each input element belonging to a label. Fully Connected Layer is described in Fig. 4.

Fig. 4. Fully Connected Layer.

III. EXPERIMENTAL RESULTS

The experimental dataset was obtained from UCSD [17] with a total of 253 highway traffic video clips collected by the Washington Department of Transportation through 2 cameras. Each video clip has about 42-52 frames with size 320×240 pixels. These video clips are divided into 3 traffic densities: light traffic, medium traffic, and heavy traffic with weather conditions of overcast, rainy weather, and clear sky. Fig. 5 illustrates three traffic density conditions in this dataset, (a) light traffic, (b) medium traffic, and (c) heavy traffic. Fig. 6 illustrates the weather conditions in this dataset, (a) clear sky, (b) overcast sky, and (c) rainy weather.

Fig. 5. types of traffic density, a) light, b) medium, c) heavy.

Fig. 6. weather conditions, a) clear, b) overcast, c) rainy.

This dataset is divided into the following 2 datasets:

- Training dataset accounts for 80%
- Testing dataset accounts for 20%

This dataset uses 200 video clips to experiment as shown in Table I.

TABLE I: THE NUMBER OF VIDEOS IN THE EXPERIMENTAL DATASET

Content	Number
Total number of used videos	200
Number of videos for training	160 (80%)
Number of videos for testing	40 (20%)
Number of videos in clear weather	36
Number of videos in overcast weather	112
Number of videos in rainy weather	52

The 5-fold cross-validation system with the training dataset accounts for 80% and the testing dataset accounts for 20% as shown in Fig. 7.

Fig. 7. Description of cross-validated experimental data.

The 5-fold cross-evaluation experimental system with the results of the confusion matrix [18] evaluates the performance of the system is depicted in Fig. 8.

Round 1				Round 2				Round 3			
Predicted Class	Actual Class			Predicted Class	Actual Class			Predicted Class	Actual Class		
	Light	Medium	Heavy		Light	Medium	Heavy		Light	Medium	Heavy
	98.96%	0.59%	0.83%		99.38%	0.78%	0.92%		99.31%	0.88%	0.49%
	0.47%	99.07%	1.78%		0.29%	97.88%	1.22%		0.28%	97.93%	0.85%
	0.57%	0.34%	97.390%		0.13%	1.34%	97.860%		0.41%	1.19%	98.660%
Round 4				Round 5							
Predicted Class	Actual Class			Predicted Class	Actual Class						
	Light	Medium	Heavy		Light	Medium	Heavy				
	99.52%	0.34%	0.69%		99.64%	0.68%	0.73%				
	0.36%	98.61%	0.58%		0.25%	98.42%	1.0%				
	0.12%	1.05%	98.730%		0.11%	0.84%	98.280%				

Fig. 8. Illustration of Confusion Matrix for The Result of Classification.

The average classification rate obtained over 5 rounds of cross-checking is shown in Fig. 9. The overall average correct classification rate is 98.66%.

	TP	TN	FP	FN	AC
Light	99.40%	98.61%	1.39%	0.60%	99.01%
Medium	98.39%	98.59%	1.41%	1.61%	98.49%
Heavy	98.18%	98.78%	1.22%	1.82%	98.48%

Fig. 9. The Experimental Results of UCSD Dataset.

The estimation rate of each type of traffic density is shown in Fig. 10.

With the UCSD Traffic dataset [17], the work [14] achieved the traffic density estimation rate shown in Table II.

TABLE II: THE ESTIMATION RATE OF THE WORK [14]

	Light	Medium	Heavy
Estimation rate	100%	83.3%	93.3%

Fig. 10. The estimation rate of UCSD dataset.

However, with the UCSD Traffic dataset [18], the proposed model achieves the traffic density estimation rate as shown in Table III.

TABLE III: THE ESTIMATION RATE OF THE PROPOSED MODEL

	Light	Medium	Heavy
Estimation rate	99.01%	98.49%	98.48%

From the data of Table II and III, it finds that the traffic density estimation rate of the work [14] is not evenly at different density labels. With the proposed model, the traffic density estimation rate is more evenly, which shows that the proposed model is more stable.

Moreover, the traffic density estimation rate "Light" of the proposed model is lower than the work [14] (99.01% compared to 100%), but the lower rate is not much, so can be remedied in the future; While the traffic density estimation rate "Medium" and "Heavy", the work [14] is much lower than the proposed model. Fig. 11 depicts the traffic density estimation rate on the UCSD Traffic dataset [17] of the work [14] and the proposed model.

Fig. 11. Compare traffic density estimation rate.

IV. CONCLUSION

The authors have studied the theoretical basis of image processing, computer vision, image feature extraction methods, OpenCV library, models, and methods to estimate traffic density from images and traffic video clips. From the researched knowledge, the authors propose the traffic density estimation model using convolutional neural networks (CNNs) and computer vision with the input data of traffic video clips. This proposed model is tested on the dataset of video clips from UCSD (University of California San Diego)

[17]. The experimental results show that the proposed solution achieved the worst estimation rate of 98.48% and the best estimation rate of 99.01%. The proposed model is more stable than the work [14]. With the traffic density estimation rate "Medium" and "Heavy", the work [14] is much lower than the proposed model.

In the future, the authors will experiment with many traffic video clips with different situations and environments to evaluate and improve the proposed model more and more completely.

ACKNOWLEDGMENT

The authors wish to thank the authors of the dataset which is publicly available at:

<http://www.svcl.ucsd.edu/projects/traffic/>.

The authors thank reviewers for their reading of our manuscript and their insightful comments and suggestions.

REFERENCES

- [1] Ruixue Mao, and Guoqiang Mao. "Road traffic density estimation in vehicular networks." In *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 4653-4658. IEEE, 2013.
- [2] Serap Kazan, Musa Balta, Kevser Ovaz, and Erman Ozer. "Traffic Intensity Estimation Using Real Time Computer Vision System.", 2013.
- [3] Chun Cheong Pang, William Wai Leung Lam, and Nelson Hon Ching Yung. "A Method for Vehicle Count in the Presence of Multiple-Vehicle Occlusions in Traffic Images." *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 3, September 2007.
- [4] Thou-Ho (Chao-Ho) Chen, Jun-Liang Chen, Chin-Hsing Chen and Chao-Ming Chang. "Vehicle Detection and Counting by Using Headlight Information in the Dark Environment." *IIHMSP '07 Proceedings of the Third International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)* – Vol. 02, pp. 519-522. 2007
- [5] Chomtip Pornpanomchai, Thititnut Liamsanguan, Vissakorn Vannakosit. "Vehicle Detection and Counting From a Video Frame." *Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition*, Hong Kong, 30-31 Aug. 2008
- [6] H.S. Mohana M. Ashwathakumar G. Shivakumar. "Vehicle Detection and Counting by Using Real Time Traffic Flux through Differential Technique and Performance Evaluation." *International Conference on Advanced Computer Control*. Page(s): 791-795, 2009.
- [7] Rui Zhao, Xiaogang Wang. "Counting Vehicles from Semantic Regions." *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 2, June 2013.
- [8] Erhan, I. N. C. E. "Measuring traffic flow and classifying vehicle types: A surveillance video based approach." *Turk. J. Electr. Eng. Comput. Sci* 19 (2011): 607-620.
- [9] Zhaoxiang Zhang, Yuqing Hou, Yunhong Wang, and Jie Qin. "A traffic flow detection system combining optical flow and shadow removal." In *Intelligent Visual Surveillance (IVS), 2011 Third Chinese Conference on*, pp. 45-48. IEEE, 2011.
- [10] Pratishtha Gupta, G.N Purohit, Sweta Pandey, "Traffic Load Computation for Real Time Traffic Signal Control", *International Journal of Engineering and Advanced Technology*, Vol. 2, Issue 4, April 2013
- [11] Sabya sanchi kanojia, "Real –time Traffic light control and Congestion avoidance system", *International Journal of Engineering Research and Applications (IJERA)*, pp.925-929, Vol. 2, Issue 2, Mar-Apr 2012
- [12] Celil Ozkurt and Fatih Camci, (2009) "Automatic Traffic Density Estimation And Vehicle Classification For Traffic Surveillance Systems Using Neural Networks", © *Association for Scientific Research, Mathematical and Computational Applications*, Vol. 14, No. 3, pp. 187-196, 2009.
- [13] Julian Nubert, Nicholas Giai Truong, Abel Lim, Herbert Ilhan Tanujaya, Leah Lim, Mai Anh Vu, (2018), "Traffic Density Estimation using a Convolutional Neural Network", *Machine Learning Project – National University of Singapore*, arXiv:1809.01564v1 [cs.LG] 5 Sep 2018.

- [14] R. Kaviani, P. Ahmadi and I. Gholampour, "A new method for traffic density estimation based on topic model", *2015 Signal Processing and Intelligent Systems Conference (SPIS)*, pp. 114-118, 2015.
- [15] Otsu's Thresholding with OpenCV, Available: <https://learnopencv.com/otsu-thresholding-with-opencv/>
- [16] correlation coefficient, Available: <https://businessjargons.com/karl-pearsons-coefficient-of-correlation.html>.
- [17] UCSD Traffic Dataset. <http://www.svcl.ucsd.edu/projects/traffic/>.
- [18] Kohavi, R. and F. Provost (1998). Glossary of terms. Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process.

Luong Anh Tuan Nguyen is the head of Information Technology Faculty, Vietnam Aviation Academy. He received B.Sc. and M.Sc. in Computer Science from the University of Science – Vietnam National University Ho Chi Minh City. He also received Ph. D. degree in Control Engineering and Automation. His current research interests include Artificial Intelligence, Soft Computing, Control Engineering, and Security.

Thanh Xuan Ha is a technical supervisor at Bao An technology company, Ho Chi Minh city, Vietnam. He received B.Sc. in Computer Science from Ho Chi Minh City University of Technology. He is also preparing to receive M.Sc. in Computer Science. His current research interests include Artificial Intelligence, Computer Vision, Network System, and Security.