A UAV Based Multi-object Detection Scheme to Enhance Road Condition Monitoring and Control for Future Smart Transportation

Jian Yang[1], Jielun Zhang[2], Feng Ye[2](✉), and Xiaohui Cheng[1]

[1] Nanjing University of Posts and Telecommunications, Nanjing, China
{yangj,chengxh}@njupt.edu.cn
[2] University of Dayton, Dayton, OH 45469, USA
{zhangj46,fye001}@udayton.edu

Abstract. Road condition monitoring and control is essential for smart transportation in the era of autonomous driving. In this paper, we propose to apply unmanned aerial vehicle (UAV), wireless communications and artificial intelligence (AI) to achieve multi-object detection for smart road monitoring and control. In particular, the application of UAV enables real-time image view to monitor road condition, such as traffic flow and on-road objects, in an efficient way without disturbing normal traffic. Those raw image data are first offloaded to a road side unit through wireless communications. A computing platform connected to the road side unit can execute the AI based scheme for road condition monitoring and control. The AI based scheme is developed around convolutional neural network (CNN). For demonstration, the objects of interest considered in this work include advertisement billboards, junctions, traffic signs and unsafe objects. Other objects can be extended to the developed system with more collected data. To evaluate the proposed scheme, we launched a UAV to collect real-life road images from multiple road sections of a highway. The AI based scheme is then developed using portion of the raw data. Test of the AI scheme is conducted using the rest of the dataset. The evaluation results have demonstrated that the proposed UAV based multi-object detection scheme can provide accurate results to support efficient road condition monitoring and control in future smart transportation.

Keywords: Artificial Intelligence · Unmanned Ariel Vehicle · Object detection · Traffic monitoring · Internet of Things

1 Introduction

Enabled with advanced information and communications technologies, Smart Cities are a future reality for municipalities around the world, which will significantly transform the way people live, enhance the quality of life, and contribute

S. Han et al. (Eds.): AICON 2019, LNICST 287, pp. 270–282, 2019.
https://doi.org/10.1007/978-3-030-22971-9_23

to the solutions to key challenges in society, economy, and environment [3,27,29]. As one of the basic components of Smart Cities, smart transportation have received increased attentions and have been translating from a vision into reality. Real-time road condition monitoring is essential to ensure driving safety. The World Health Organization (WHO) recently estimated that road injuries are the 8th leading cause of death worldwide, with 1.4 million deaths all around the world [26]. In the U.S., 37,461 people died on roadways in 2016, a 5.6% increase from 2015 [16]. This increase is significant [1] since the number of motor vehicle fatalities remained virtually unchanged varying between 32,479, and 37,461 over the 2010–2016 time period [22]. In the new era of autonomous driving, the advancement in the next-generation mobile networks (also known as 5G), vehicular networks, and Internet-of-Things (IoT) shall be part of a smart road infrastructure to achieve road condition monitoring and control for both safety and efficiency.

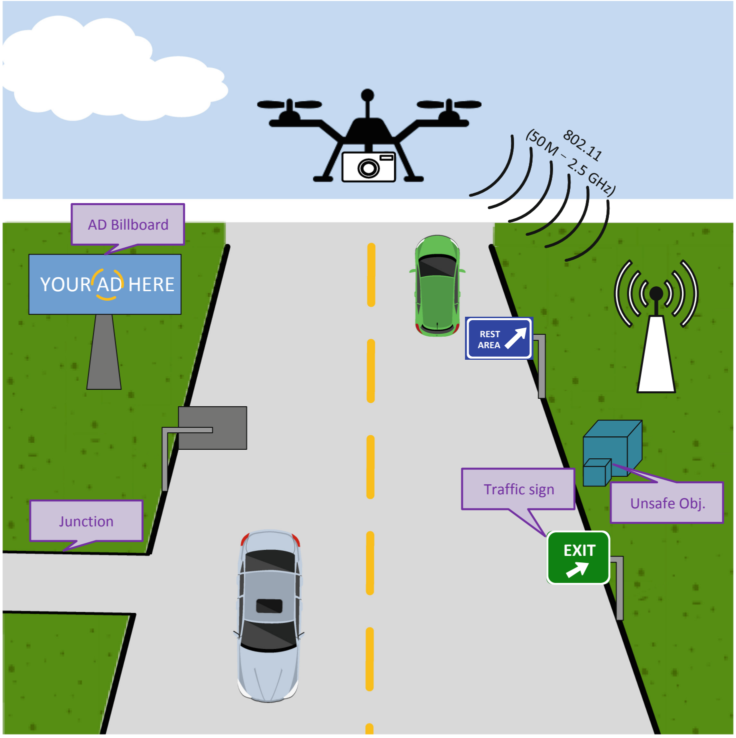


Fig. 1. Overview of the UAV based multi-object detection framework.

As a preliminary work to enhance the smart traffic, in this paper, we propose to apply unmanned aerial vehicle (UAV), wireless communications and artificial intelligence (AI) based schemes to assist road condition monitoring. An overview of the proposed UAV based multi-object detection framework is shown in Fig. 1. Particularly in this work, objects of interest for road monitoring are advertisement billboards, junctions, traffic signs and unsafe objects. Other objects of interest can be easily extended to the proposed framework if required. Those objects need to be monitored and shared by drivers and autonomous driving systems. For example, traffic signs provide important information and instructions for drives, and they need to be maintained and updated when necessary to ensure a safe driving environment. View to junctions may be blocked which would cause safety issues to driving, especially autonomous driving. Unsafe objects on the road side also challenge the drivers. For instance, the glare reflected by some of them can be extremely dangerous to drivers, over-size objects can block traffic signs, etc. To tackle the issues caused by these objects, we propose to use UAV as image data acquisition equipment for objects in our interest to be detected from the collected images. Meanwhile, the UAV is designed to be connected with a road side unit through wireless communications. In the current implementation, IEEE 802.11 is applied for local information exchange. Nonetheless, the future deployment in smart transportation will rely more on 5G and IoT for seamless network connections. The core of the proposed multi-object detection scheme is an AI approach based on convolutional neural network (CNN). The proposed AI based scheme is able to detect multiple objects of interest in one image, or one frame of a video stream. Real-life image data of the highways were collected using UAV for the evaluation of the proposed multi-object detection scheme. With more advanced technology in AI computing chips as well as more efficient algorithms, on-board computation may be available for data pre-processing or even the multi-object detection itself. For simplicity, in this work, it is assumed to offload the computation from the UAV to a more powerful local platform due to the limited on-board computation capability of the testbed.

The rest of this paper is organized as follows. Section 2 introduces the related works that cover the object detection. Section 3 illustrates the proposed multi-object detection scheme. Section 4 demonstrates the evaluation results of the proposed scheme using real-life data. Section 5 gives conclusion and future works.

2 Related Work

IoT is a paradigm that is gaining ground rapidly in the scenario where wireless communications and computing are required [2]. IoT devices have been deployed for road traffic surveillance, accident detection [21], parking spots availability detection [7], traffic offense detection [15] and road traffic noise management in Smart Cities [9]. In [17], the authors designed and implemented a reliable intelligent framework to monitor vehicle traffic condition with a low cost by applying IoT devices. The processed data are transferred to the server through Wi-Fi for the further traffic control. To keep up with the transition towards smart transportation, next generation mobile networks will support the communication requirements for IoT and smart transportation. Thanks to the ultra-reliable low latency communications, 5G transport network is expected to provide the satisfactory capacity, required latency, for real-time road monitoring and control in smart transportation [18,20], which is promising to promote smart drive-assistant services, cooperative driving services, as well as fully autonomous driving services. In this work, we leverage the advancement in the next-generation mobile network and IoT to support the data transmission for the proposed multi-object detection scheme.

Raw data collection is critical to multi-object detection for road monitoring. Traditionally, special utility vehicles are deployed in field to collect data. For example, researchers developed a special vehicle named as S.T.I.E.R. for data collection [5]. A high-speed camera is installed at the tail of the vehicle to collect image data of road surfaces for pavement distress detection in the driving lane. However, such data collection method is limited in coverage as it only monitors one driving lane at a time. Moreover, operating such a vehicle on highways may affect normal traffics. Using UAV is an alternative way to address those issues and collect data [10]. For example, due to the wider field of view from a UAV, a larger area can be covered. Thus more objects besides road surface can be included in the object of interests, such as traffic signs, road junctions, etc. Moreover, the operational cost of deploying UAV for data collection is much cheaper than deploying a specific vehicle.

AI has been widely adopted in the recent years for object detection in images and videos [11,23,24]. Object detection approaches can be mainly classified into three types: classic, two-stage, and one-stage approaches [10]. One of the widely adopted classic object detection algorithm is the CNN used for digits recognition that proposed by LeCun in [11]. The classical detectors conduct the detection by operating like a sliding window, where a classifier is applied every time over a image grid that is predefined. Two-stage detection approaches, e.g. region-based convolutional neural network (R-CNN) is being developed over the last few years with several improvements [6,25] and extensions [8,13]. In R-CNN and its extensions, region proposal methods are first applied to generate a sparse set of candidate proposals which contain all objects. The process would filter out the majority of negative locations in an image. The second stage is to perform classifications on these proposals to separate them into foreground classes and background classes. One-stage approach has been adopted more recently. For example, SSD [14], YOLO [23,24] are two main approaches that assume a single detection stage that is closer to the way human beings detect objects. The two approaches both enhance the efficiency as well as detection precision. For instance, the latest YOLO can achieve almost the same detection precision compared with other two-stage detection approaches [4]. In this work, we develop an AI based multi-object detection scheme around a CNN structure.

3 Multi-object Detection Scheme for Road Condition Monitoring

Without loss of generality, the proposed multi-object detection scheme is designed and implemented in two steps. In the first step, we establish a database by collecting raw images from a UAV along the highway in Jiangsu province, China. With the established database, in the second step, we develop a multi-object detection scheme around a CNN based deep learning algorithm. Detailed illustration is given in the rest of this section.

3.1 UAV Based Raw Data Collection

The UAV used for data collection is shown in Fig. 2. For practical monitoring, a UAV is required to cruise at a relatively high speed between 40 km/h to 80 km/h for a long duration. The actual implementation has a hovering height at 40 meter. Meanwhile, the UAV must be able to resist strong wind, e.g. below the level Beaufort 6. Reasonable payload capacity is also required to mount a high-resolution camera, data communication modules, and possible on-board computing devices for data acquisition, transmission and processing, respectively. The communication module in the applied UAV needs to be energy efficiency, high bandwidth, low latency, and secure. Currently, the raw data captured from the on-board camera is beyond 4K resolution. Transmission of raw data in real-time causes a challenge to the current implementation that relies on the IEEE 802.11. A better transmission module, e.g., based on 5G, must be developed for the actual system in future smart transportation. An on-board computing device is yet to be developed so that data pre-processing can be conducted in real-time to reduce the transmission overhead. For simplicity, the specifications of the deployed UAV are listed in Table 1.

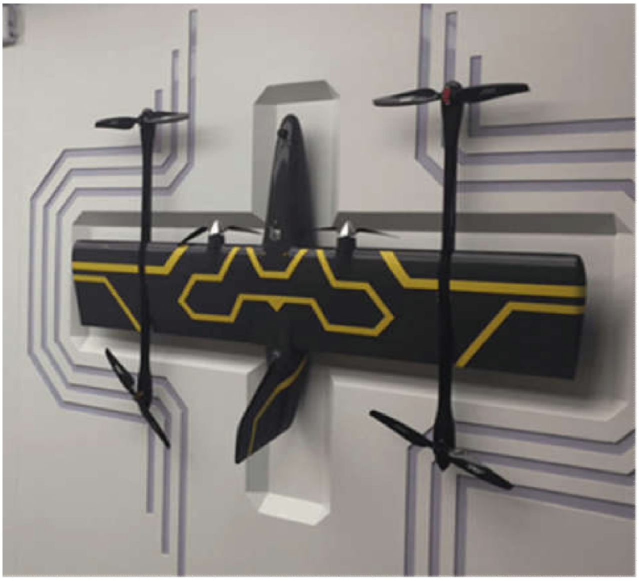


Fig. 2. SwapImagine - the UAV used for data collection.

Table 1. Specifications of the deployed UAV.

UAV specification	Value
Cruise duration (min)	70
Wind-resistance (Beaufort)	>6
Payload (kg)	6–8
Cruising speed (km/h)	40–80
Com. specification	Value
Modulation	COFDM, 16QAM
Frequency (Hz)	50M–2.5G
Bandwidth (Mbps)	2–8
Output power (W)	0.5–2
Encryption	AES 128
Range (km)	8 (40 for the directional)

3.2 AI Based Multi-object Detection Scheme for Road Monitoring

Our proposed AI based multi-object detection scheme is built around the core structure of a CNN. A typical CNN has the architectures designed for processing grid-like data [12]. Compared with a traditional feedforward neural network where each weight is used only once, a typical CNN uses kernels at all positions of the input data. A well developed CNN would capture the features for every location in the data, especially for image data [4].

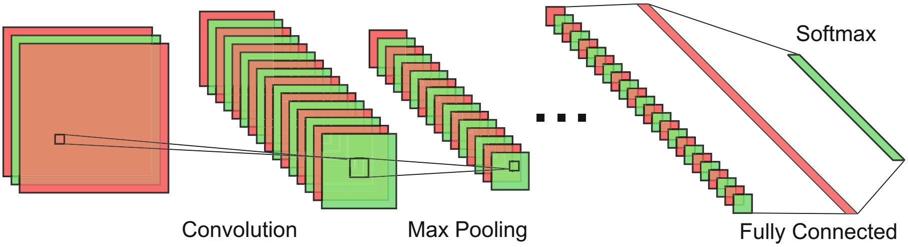


Fig. 3. CNN standard architecture.

As shown in Fig. 3, convolutional layers are the core components of a CNN. The key function of the convolutional layers is to perform feature extraction from the input image data. They are composed of a set of spatial filters which are usually known as convolutional kernels. Each convolutional layer in a CNN is followed by an activation layer which gives nonlinear transformation to the previously obtained feature map. The activation function enhances computational efficiency and helps to eliminate gradient vanishing in the training progress [19].

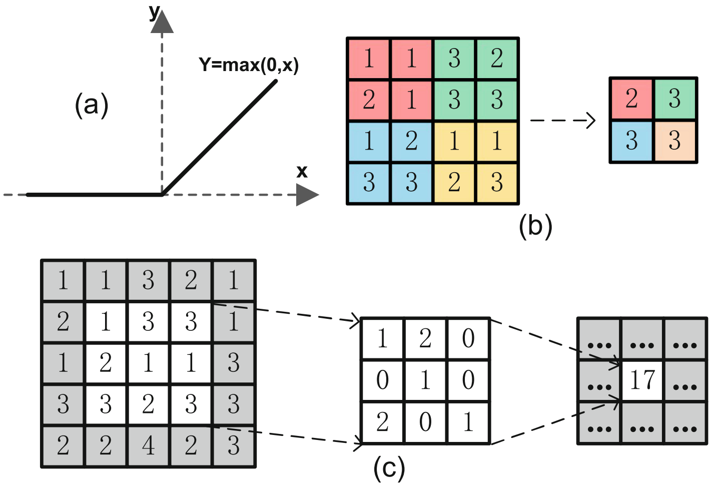


Fig. 4. Operations in CNN layers.

Rectified linear unit (ReLU) expressed in Eq. (1) is usually used as the activation function, which is illustrated in Fig. 4(a).

$$f(x) = \max(0, x) = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{else.} \end{cases} \tag{1}$$

A pooling layer also follows the convolutional layer to reduce the dimension of the extracted feature map without losing essential information. Max pooling and average pooling are two common pooling methods used in CNN [28]. Assuming a filter of 2 × 2, the max pooling computes as follows:

$$\text{max pooling} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} = \max(x_1, x_2, x_3, x_4). \tag{2}$$

And the average pooling computes as follows:

$$\text{average pooling} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} = \frac{1}{4} \sum_{i=1}^{4} x_i. \tag{3}$$

In comparison, max pooling reduces the feature map size by picking the maximum value within a sliding window, while average pooling takes the average value in the sliding window to streamline the feature map. Figure 4(b) demonstrates a max pooling process with a 2-by-2 filter and stride equals 2. An example of a completed convolution operation is given in Fig. 4(c). In this example, an input image data with a size of 5-by-5 is convoluted by a 3-by-3 filter. The stride, a hyper-parameter that determines the sliding interval of the kernel, is set to be 1. The convolution operation produces a two-dimension feature map.

A fully connected layer and a softmax layer are the final layers of a CNN architecture. The fully connected layer is known as multilayer perceptron where neural nodes are fully connected to the output of the previous layer. The softmax layer is used to normalize the probability value for classification decision. The softmax computes an output as follows:

$$S_j = \frac{e^{a_j}}{\sum_{k=1}^{T} e^{a_k}}, \tag{4}$$

where S_j is the probability of the j_{th} class to be the classification result among all T possible classes. a_i are the output of the former fully connected layer. The cross entropy is further calculated for classification.

$$E = -\sum_{j=1}^{T} y_i \log(S_j), \tag{5}$$

where $y_i = 1$ only if the i_{th} class it the true class matching the classification result, otherwise $y_i = 0$.

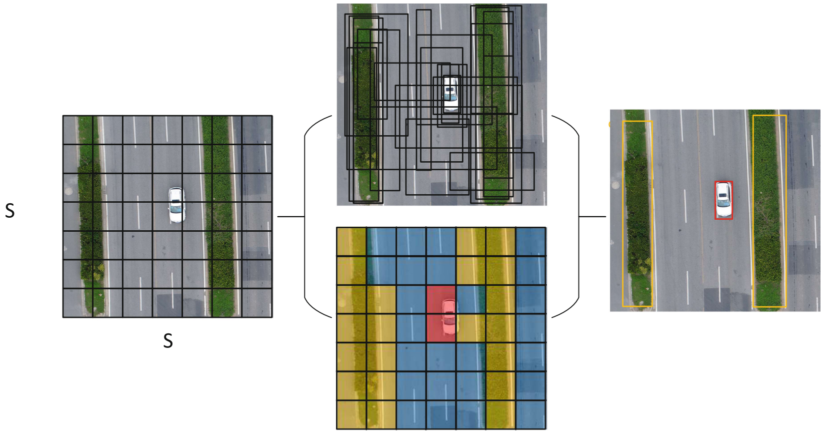


Fig. 5. Work flow of the AI based multi-object detection scheme.

The proposed AI based scheme is built around the CNN structure described above to achieve multi-object detection in one image. For simplicity, the work flow of the implemented AI scheme is illustrated in Fig. 5. An input image is first divided into $S \times S$ grid cells. Boundary boxes and probabilities will be calculated to make detection on the relevant objects. Each of the grid cells is composed of $(\mathbf{x}, \mathbf{y}, \mathbf{w}, \mathbf{h}, \mathbf{c})$, where $(\mathbf{x}, \mathbf{y})$ is the coordinates stands for the centers of the position of B detection boundary boxes of the corresponding grid cell; $\mathbf{w}$ is the width of the detection boundary boxes; $\mathbf{h}$ stands for the height; and $\mathbf{c}$ is the confidence score which is a column vector represents the predicted probabilities of respective C categories. Success of a detection is computed as follows:

$$\mathbf{o} = \frac{1}{2} + \frac{1}{2}\frac{pI - \eta}{|pI - \eta|}, \tag{6}$$

where $\mathbf{o}$ is either 1 or 0 indicates a successful detection of targeted objects or object detection failure respectively. In Eq. (6), p reflects the probability of a targeted object existing in the grid cell, I is the value of the term as Intersection over Union (IOU) which is defined as the overlapping rate of the predicated bound and the ground truth bound. The production pI is the confidence score of the prediction and η is the threshold to filter out low confidence scores. The remaining boundary boxes will be processed by non-maximum suppression to finalize the object detection. Finally, IOU is computed as follows:

$$I = \frac{A_p \cap A_t}{A_p \cup A_t}, \tag{7}$$

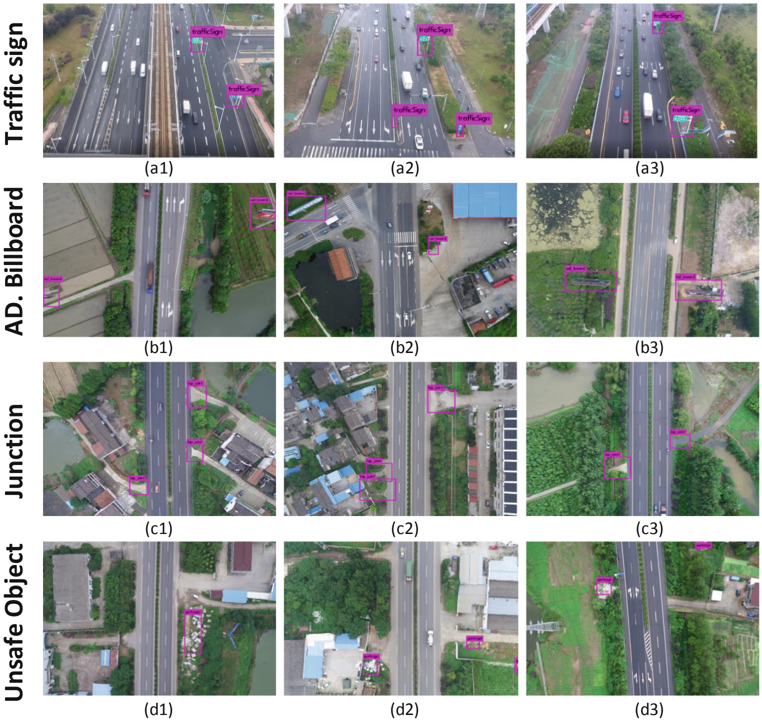
where A_p and A_t are the area of predicted bounding box and the ground-truth bounding box respectively.

4 Evaluation Results

In this section, we first establish our database which includes sets of sample images collected by operating the UAV above partial highway road sections in

Table 2. The applied network architecture in the scheme design.

Layer		Filters	Size/Stride	Output data dimension
1	conv	16	3 × 3/1	416 × 416 × 16
2	max		2 × 2/2	208 × 208 × 16
3	conv	32	3 × 3/1	208 × 208 × 32
4	max		2 × 2/2	104 × 104 × 32
5	conv	64	3 × 3/1	104 × 104 × 64
6	max		2 × 2/2	52 × 52 × 64
7	conv	128	3 × 3/1	52 × 52 × 128
8	max		2 × 2/2	26 × 26 × 128
9	conv	256	3 × 3/1	26 × 26 × 256
10	max		2 × 2/2	13 × 13 × 256
11	conv	512	3 × 3/1	13 × 13 × 512
12	max		2 × 2/1	13 × 13 × 512
13	conv	1024	3 × 3/1	13 × 13 × 1024
14	conv	1024	3 × 3/1	13 × 13 × 1024
15	conv	30	1 × 1/1	13 × 13 × 30
16	detection			



(a1) (a2) (a3)

(b1) (b2) (b3)

(c1) (c2) (c3)

(d1) (d2) (d3)

Fig. 6. Examples of the multi-object detection results.

Jiangsu, China. The operating height of the UAV is fixed at 40 m. The captured videos are in 4K resolution, which is 4608 × 3456 in specific. Each image extracted from each frame has a file size around 25 Megabytes. The applied dataset includes multiple video clips collected from 25 different road sections of the monitored highway. The raw data is captured and stored from the UAV in the current implementation, and processed in an offline computing platform. In order to achieve real-time monitoring and control, data pre-processing and filtering will be further investigated in the future work.

The AI based multi-object detection scheme first extracts a single frame of from a raw video clip. The extracted image is then resized to a fixed size of 416 × 416, which is used as the input to the multi-object detection scheme. The architecture of the proposed AI scheme in the current implementation is illustrated in Table 2. The current AI scheme is evaluated based on 1, 457 images extracted from multiple raw video clips captured from the UAV described earlier. Figure 6 shows some results of the multi-object detection scheme.

For better illustration, the overall performance of the proposed scheme is provided in Table 3. The evaluation results show that the detection accuracy is higher when the detection objects are traffic signs or junctions connected to the highway. The detection accuracy of advertisement billboards and unsafe

Table 3. Evaluated detection performance

Object class	Number of labeled objects	Number of successful detection	Detection accuracy
Traffic sign	978	818	83.64%
AD. Billboard	33	23	69.70%
Junction	2571	2212	86.04%
Unsafe object	733	506	69.03%

objects on road sides are low due to their varying colors and shapes. Nonetheless, the preliminary results are presented to demonstrate the efficacy of combining UAV, AI, next-generation mobile networks and IoT to support the future smart transportation.

5 Conclusions

Road condition monitoring and control will be critical in future smart transportation. In this paper, we explored the possibility to apply UAV and AI based scheme for on-road multi-object detection. In the studied scenario, the application of UAV for image acquisition allows us to collect top views of roads in an efficient way. The AI scheme is built around modern CNN-based architectures that can achieve multi-object detection in an image. Evaluation of the studied work was conducted using real-life data collected by a UAV. The evaluation results demonstrated that our proposed AI based scheme, together with UAV, wireless communications and IoT can provide a good solution to future road condition monitoring and control in smart transportation. In future work, we will further explore the schemes for accurate and efficient multi-object detection for road monitoring. For example, different objects of interest should be detected simultaneously in the AI based scheme. Moreover, we will also validate the schemes in various road monitoring techniques such as the live web-cam along the highway and on-board dash-cam.

References

1. Essex, A., Shinkle, D., Miller, A., Pula, K.: Traffic safety trends—state legislative action 2017. In: National Conference of State Legislatures (2018)
2. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
3. Clarke, R.Y.: Smart cities and the internet of everything: the foundation for delivering next-generation citizen services. IDC Government Insights (2013)
4. Ćorović, A., Ilić, V., Durić, S., Marijan, M., Pavković, B.: The real-time detection of traffic participants using yolo algorithm. In: 26th Telecommunications Forum (TELFOR), pp. 1–4. IEEE (2018)

5. Eisenbach, M., et al.: How to get pavement distress detection ready for deep learning? A systematic approach. In: International Joint Conference on Neural Networks (IJCNN), pp. 2039–2047, May 2017. https://doi.org/10.1109/IJCNN.2017.7966101
6. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587, June 2014. https://doi.org/10.1109/CVPR.2014.81
7. Goyal, R., Kumari, A., Shubham, K., Kumar, N.: IoT and XBee based smart traffic management system. J. Commun. Eng. Syst. **8**(1), 8–14 (2018)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Kazmi, A., Tragos, E., Serrano, M.: Underpinning IoT for road traffic noise management in smart cities. In: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 765–769, March 2018. https://doi.org/10.1109/PERCOMW.2018.8480142
10. Kharchenko, V., Chyrka, I.: Detection of airplanes on the ground using YOLO neural network. In: IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET), pp. 294–297, July 2018. https://doi.org/10.1109/MMET.2018.8460392
11. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989). https://doi.org/10.1162/neco.1989.1.4.541
12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436 (2015)
13. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: CVPR, vol. 1, p. 4 (2017)
14. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
15. Mihelj, J., Kos, A., Sedlar, U.: Source reputation assessment in an IoT-based vehicular traffic monitoring system. Procedia Comput. Sci. **147**, 295–299 (2019)
16. Moore, W., et al.: Transportation statistics annual report, 2016. United States, Bureau of Transportation Statistics (2017)
17. Nagmode, V.S., Rajbhoj, S.: An intelligent framework for vehicle traffic monitoring system using IoT. In: International Conference on Intelligent Computing and Control (I2C2), pp. 1–4. IEEE (2017)
18. Nakao, A., et al.: End-to-end network slicing for 5G mobile networks. J. Inf. Process. **25**, 153–163 (2017)
19. Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S.: Activation functions: comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378 (2018)
20. Öhlén, P., et al.: Data plane and control architectures for 5G transport networks. J. Lightwave Technol. **34**(6), 1501–1508 (2016)
21. Patel, R., Dabhi, V.K., Prajapati, H.B.: A survey on IoT based road traffic surveillance and accident detection system (a smart way to handle traffic and concerned problems). In: Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1–7, April 2017. https://doi.org/10.1109/IPACT.2017.8245066
22. Patil, D., Rosekind, M.: Traffic fatalities data has just been released: a call to action to download and analyze. US Department of Transportation (2015)
23. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

24. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. arXiv preprint (2017)
25. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
26. World Health Organization, et al.: The top 10 causes of death (2016). https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death. May 2018
27. Ye, F., Qian, Y., Hu, R.Q.: Smart service-aware wireless mixed-area networks. IEEE Netw. **33**(1), 84–91 (2019). https://doi.org/10.1109/MNET.2018.1700399
28. Yi, Z.: Evaluation and implementation of convolutional neural networks in image recognition. In: Journal of Physics: Conference Series, vol. 1087, p. 062018. IOP Publishing (2018)
29. Zhang, J., Ye, F., Qian, Y.: A distributed network QoE measurement framework for smart networks in smart cities. In: IEEE International Smart Cities Conference (ISC2), pp. 1–7, September 2018. https://doi.org/10.1109/ISC2.2018.8656854