

# Prediction of Queue Dissipation Time for Mixed Traffic Flows With Deep Learning

HUNG-HSUN CHEN<sup>1</sup>, YI-BING LIN<sup>2,3,4,5</sup> (Fellow, IEEE), I-HAU YEH<sup>6</sup>,  
HSUN-JUNG CHO<sup>7</sup>, AND YI-JUNG WU<sup>8</sup>

<sup>1</sup>Program of Artificial Intelligence and Information Security, Fu Jen Catholic University, New Taipei City 24205, Taiwan

<sup>2</sup>College of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

<sup>3</sup>College of Humanities and Sciences, China Medical University, Taichung 404, Taiwan

<sup>4</sup>Miin Wu School of Computing, National Cheng Kung University, Tainan City 701, Taiwan

<sup>5</sup>Computer Science & Information Engineering, Asia University, Taichung City 41354, Taiwan

<sup>6</sup>Elan Microelectronics Corporation, Hsinchu 300092, Taiwan

<sup>7</sup>PiXORD Corporation, Hsinchu 30076, Taiwan

<sup>8</sup>Institute of Computational Intelligence, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

CORRESPONDING AUTHOR: H.-H. CHEN (e-mail: hhchen@nctu.edu.tw)

This work was supported in part by PiXORD Corporation, Taiwan.

**ABSTRACT** Queue dissipation has been extensively studied about traffic signalization, work zone operations, and ramp metering. Various methods for estimating the intersection's queue length and dissipation time have been reported in the literature, including the use of car-following models with simulation, vehicle trajectories from GPS, shock-wave theory, statistical estimation from traffic flow patterns, and artificial neural networks (ANN). However, most of such methods cannot account for the impacts of interactions between different vehicle types and their spatial distributions in the queue length on the initial discharge time and the resulting total dissipation duration. As such, this study presents a system, named TrafficTalk, that applies a deep learning-based method to reliably capture the queue characteristics of mixed traffic flows, and produce a robust estimate of the dissipating duration for the design of the optimal signal plan. The proposed TrafficTalk, featuring the effectiveness in transforming video-imaged traffic conditions into vehicle density maps, has proved its performance under extensive field evaluations. For instance, compared with the benchmark model, XGBoost in the literature, it has reduced the MAPE from 25.8% to 10.4%, and from 31.3% to 10.4% if the queue discharging stream comprises motorcycles.

**INDEX TERMS** Deep learning (DL), traffic queue dissipation time, traffic queue pattern, mixed traffic flows, object detection, traffic signal countdown timer (TSCT).

## I. INTRODUCTION

A RELIABLE estimate of queue dissipation duration is essential information for traffic controls and operations. Delays and other congestion-related measurements based on reliably estimated queue information have been widely used by a large body of researchers for work zones [1]–[3], and ramp metering operations [4]. For instance, Rouhani and

Niemeier argued that accurate estimation of traffic congestion and delay is imperative to studying traffic and flow characteristics as well as formulating effective traffic control strategies [5]. Dion *et al.* categorized five intersection delay models involving different queue estimation methods: deterministic queuing model, shock wave delay model, steady-state stochastic delay model, time-dependent stochastic delay model, and finally microscopic simulation delay model [6]. He also indicated that delay is a parameter that is difficult to estimate without reliably estimated queue information.

The review of this article was arranged by Associate Editor Chi-Hua Chen.

Recognizing the vital role of queue dissipation information in design of traffic control strategies, some studies proposed to approximate such an estimate with microscopic traffic simulation models [7]. The extensive calibration needs of their embedded car-following and intersection discharging behaviors often encumber the estimation work with the simulation to yield the desired level of accuracy.

Intending to address the same issue but from different perspectives, traffic researchers have explored a variety of different estimation methods. For instance, some studies apply the Hidden Markov model based on the processed GPS trajectory data to assess traffic queues and congestion levels [8].

One classical method widely adopted by the traffic control community is to employ the shockwave theory to characterize the queue formation and dissipation patterns and then compute the resulting delay. Examples of methods along this line can be found in the studies by Michalopoulos *et al.* [9] and Wu and Liu [10].

Note that most available models for estimating the queue formation and dissipation times are developed traffic flow of one vehicle type. The entire estimation task becomes much more challenging when the traffic flows comprise mixed types of vehicles such as transit vehicles and motorcycles. For example, the motorcycles may parallelly park in a lane at the signalized intersection, and the dissipation rates for such motorcycles are very different from their surrounding cars. Moreover, the spatial distribution of motorcycles in the queue lines also affects their discharging times and the resulting dissipation rate of the entire queue. More specifically, the total dissipating duration for an intersection's mixed-flow queue length may vary significantly with how motorcycles are distributed in the queued stream.

How to design effective transportation systems is an important issue for a modern city. In traditional approaches, each signalized intersection is provided a fixed signal timing plan on a predetermined basis according to historical traffic flow data. However, the predetermined signal plans cannot capture the characteristics of real-time traffic patterns, which usually lead to unnecessary vehicle waiting times, and make the transportation system inefficient. In addition, the conventional traffic signal control models cannot capture the complexity of the traffic flows in the real world. In recent years, some studies have used reinforcement learning (RL) to realize signal control trained with simulation platforms [11]–[14]. Unfortunately, these simulation-trained RL solutions may not reflect the characteristics of real traffic flows. Also, most simulation platforms do not consider the real effects of motorcycles, and simply convert a motorcycle into a value of passenger car unit (PCU) (i.e., 0.4 PCU [15]) to be manipulated in the simulation platforms. Furthermore, the RL model cannot be used to correctly display the traffic signal countdown timer (TSCT) feature in the real-time scenario.

In many countries, motorcycles account for significant parts of the mixed traffic flows. The irregular behaviors of

motorcyclists usually lead to complicated characteristics of mixed traffic flows, which will make the design of traffic signal timing plans more challenging. The most common method for designing traffic signal timing plans is to use conventional traffic signal control models with offline historical traffic flow data. The average dissipation rate for mixed traffic flows is estimated according to the headway and the occupied space based on statistical data, and the assumptions made on the traffic patterns may be difficult to reflect the actual traffic state. For example, the motorcycles may parallel park in a lane at the signalized intersection, and the dissipation rates for these motorcycles are very different from the cars. Therefore, parameters and the assumptions of conventional traffic signal control models are unable to fully account for the real-world traffic complexity.

Reinforcement learning for traffic signal control was used in the conventional traffic models [11], [12]. However, these models cannot investigate the characteristics of parallel dissipation of motorcycles. According to historical traffic flow data, some solutions [13], [14] utilized traffic simulation platforms to generate animations. The animations serve as inputs to design RL-based signal timing plans without considering the composition of traffic flows and the queue patterns. Although the simulated animations are generated according to historical traffic flow data and real-world road geometry, they still cannot exactly fit the actual traffic flows dissipation state, especially for mixed traffic flows. Also, when a red traffic light in Taiwan is turned on, the road users will see the remaining TSCT seconds. Since RL has to repeat making decisions in a short period (e.g., every 3 or 5 seconds), it is infeasible to apply RL to obtain TSCT in advance. Also, the computational complexity of RL is much higher than deep learning (DL). Therefore, it is expensive to implement practical traffic signal control using RL for timely changing actions in the real world.

To deal with the complexity involved in estimating the mixed flows' queue dissipation duration, this study explores a novel method called TrafficTalk, which can transform the real-world traffic information captured from the real-time streaming videos into "vehicle density maps" and reflect the spatial distribution of different types of vehicles in the queues. Then, such information in turn serves as the input for TrafficTalk to predict the queue dissipation time of mixed traffic flows at signalized intersections. The paper is organized as follows: Section II presents previous studies of queue length estimation and dissipation time prediction; Section III proposes the vehicle density map and deep learning models for TrafficTalk; Section IV describes the TrafficTalk architecture; and Section V demonstrates the experiments and results.

## II. RELATED WORK

In this section, we review previous queue estimation methods. In [16], the vehicle queue length was investigated in the simulation. The authors conducted a case study using traffic data from an intersection in Beaufort, North Carolina.

However, there is no ground truth in the case study. Also, in the simulation, the study assumed that whether the queue is empty or not is known in advance. Such assumptions may not be practical in real-time traffic situations.

In [17], the cycle-based queue lengths at a signalized intersection were estimated by probe vehicle trajectory data. This study used the maximum likelihood estimation (MLE) method and conducted a performance evaluation of the proposed approach based on simulation and empirical data. However, the estimation of the queue with mixed through-turn lanes was not available. Also, this study did not include cases with mixed traffic flows.

Based on License Plate Recognition (LPR) data, queue length estimation at signalized intersections [18] was proposed. The queue length in the current cycle is predicted through regression analysis using the queue length in the previous cycle.

In [19], real data was collected from the Adalhan junction in a central position in the province of Konya Turkey. The data were used to derive the vehicle arrival and departure distributions. Then the standard M/M/1, M/G/1, and G/G/1 models were used to derive the queue lengths. Dissipation time was not studied in this paper.

The M/M/1 model was also used in [20]. The arrivals are approximated as the Poisson process for M/M/1 and M/G/1. For G/G/1, the arrival process was approximated by the measured mean and variance.

The study in [21] also made the Poisson assumption for its simulation experiments. Note that in a real scenario, the behavior of the traffic in the junction is transient, and the Poisson assumption may not apply.

The study in [22] proposed a real-time queue length estimation method based on probe vehicles' data. Based on the trajectories and stopping information, an integrated parking process together with the Markov model was developed to compute the queue length. As listed in their future work, the impact of queue length estimation accuracy caused by different connected vehicle distributions in the queue should be considered.

In [23], the authors estimated the vehicle queue length at a signalized intersection. Although the discharging time was not considered, the study suggests that the accuracy of queue length estimation has nothing to do with typology and phases of the intersection. Instead, it is affected by the positions of the waiting vehicles and the vehicle types. This observation is consistent with our study.

Vehicle-to-vehicle communication facilitates the exchange of roadside information, enabling easy access and sharing among users. In [24], the study introduced Linear Adaptive Congestion Control (LACC), enhancing the advantages of greedy routing and Data Dissemination Model (DDM). The study in [25] investigated joint queue estimation and max pressure control for urban networks with traffic lights. This approach was investigated by simulation experiments and it is not clear how the performance of this approach is affected by real traffic.

In [26], the authors applied the Lighthill–Whitham–Richards shockwave theory and Robertson's platoon dispersion model to predict the arrival of vehicles in advance at intervals of 5 seconds. This study did not consider vehicle types and did not derive the discharging times.

Using Artificial Neural Networks (ANN) to compute the highly stochastic queue dissipation time has also been attempted by the traffic community. For instance, Murat and Baskan applied the ANN to estimate the vehicle delay time for over-saturated or non-uniform traffic conditions [27], which has the mean average error (MAE) [28] of more than 4 seconds and the mean absolute percentage error (MAPE) [29] of 12.4061%. TrafficTalk proposed in this paper achieves better MAPE (i.e., 10.4%). Motawej *et al.* applied the same method along with the dissipativity-based control to build a time-series model to perform the estimation [30]. In their methods, the lagged traffic flow data are used as the input for predicting the real-time traffic flow. Note that the discrepancies in discharge time between different vehicle types and the impact of parallelly queued and discharged motorcycles in the traffic stream are not considered in this model.

To our knowledge, all of the previous studies considered queue length, not queue dissipation time. Also, they seldom considered the impact of motorcycles. Furthermore, these methods may not reasonably reflect the real traffic situations since the arrival processes and the permutations of the queues were randomly created rather than generated from the real scenarios.

In [31], linear regression was applied to calibrate the functional relations between the total queue dissipation duration and its key contributing variables, including the passenger cars, sport utility vehicles or truck, heavy vehicles and three binary variables. The R-squared value ( $R^2$ ) [32] of the Trans-Log model in [31] is 0.73, which is lower than the  $R^2$  value of TrafficTalk, which is 0.9145.

### III. THE DEEP LEARNING MODELS

This paper considers three major types of vehicles in Taiwan including large vehicles (buses or trucks), passenger cars, and motorcycles. To detect the vehicles and predict their behaviors, TrafficTalk designs a cascade deep learning model consisting of YOLO (for detection) and convolutional neural network (CNN, for prediction) [33], [34]. We first define a *queue pattern* as the order of different types of vehicles queued at the last second of the effective red time in any given signal cycle. According to the data we collected, it is observable that the dissipation characteristics of different queue patterns will affect the required dissipation time. The driver of a large vehicle may spend more perception and reaction time to start up at the beginning of each green time of the traffic signal cycle. Compared with a passenger car or a motorcycle, a large vehicle requires more headway (i.e., the elapsed time between two successive vehicles as they pass a point on the roadway) [35], and the corresponding dissipation time varies for different types of vehicles.



FIGURE 1. Two different arrangements with the same number of vehicles.

Fig. 1 shows the queue patterns with two arrangements of four large vehicles, seven passenger cars, and one motorcycle. In Fig. 1 (a), two large vehicles are queued in the front of the intersection, which requires about 18 seconds of dissipation time; Fig. 1 (b) shows the queue pattern with two passenger cars in front of the queue, which requires about 13 seconds of dissipation time.

TrafficTalk builds a CNN model to recognize the different queue patterns in traffic videos and predict the corresponding dissipation times. This section describes how TrafficTalk detects and extracts the vehicles from an image, and modifies the CNN model for more efficient execution.

#### A. VEHICLE OBJECT DETECTION AND EXTRACTION

In our study, the traffic video datasets 1-3 were collected from three signalized intersections illustrated in Fig. 2 (a). Details of the datasets 1-3 will be elaborated in Section V. We translate the video images into the vehicle density maps illustrated in Fig. 2 (c). The concept of vehicle density map assumes that different types of vehicles have different “densities”. In physics, objects with larger densities have larger static frictions, which results in slower object movement (long headways of the vehicles). By creating a vehicle density map from a vehicle pattern image in a road intersection, we translate the question “how long the vehicles in an intersection are dissipated” into a question like “how long can we squeeze the heterogeneous-density toothpaste out of the tube”.

To create a vehicle density map, we first use YOLOv4 [36] to identify the types of vehicles and detect the positions of the vehicle objects in the traffic videos (Fig. 2 (a)). The detection results (Fig. 2 (b)) were affected by miscellaneous noises (such as buildings and roadside trees) other than the target objects. Such noises reduce the accuracy of model predictions. To resolve this issue, we extract the detected vehicles as colored rectangles and remove the “background information” (Fig. 2 (c)). The extracted objects include passenger cars and large vehicles, which are considered in most traffic management models [37]. We also detect and extract motorcycle objects since they account for a large proportion of mixed traffic flows in Taiwan. In TrafficTalk, vehicle object detection is performed in the YOLO detection module (to be elaborated later in Fig. 6 (3)). Vehicle object extraction

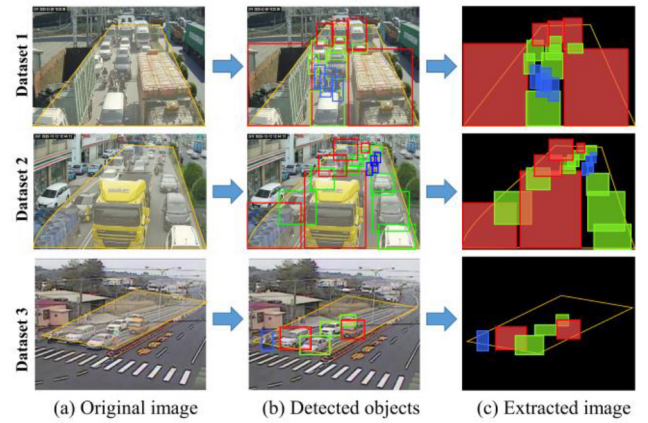


FIGURE 2. Vehicle density map creation through vehicle object detection and extraction.

is performed in the extraction module of the Predictor (to be elaborated later in Fig. 7 (2)).

The above two-step image processing creates a vehicle density map bounded by a yellow-boundary region in Fig. 2 (b), where a green rectangle denotes a passenger car (i.e., the density for a passenger car), a blue rectangle denotes a motorcycle, and a red rectangle represents a large vehicle such as a bus or a truck.

#### B. LABELS FOR QUEUE DISSIPATION TIME PREDICTION

TrafficTalk uses CNN models to learn different vehicle density maps (queue patterns in traffic videos) and predict the corresponding dissipation time. As we mentioned in the previous subsection, the output of the extraction module is a vehicle density map, which is the input of the CNN model. In this paper, one of the most common CNN models, VGG16 [38] is used in TrafficTalk. Since we have reduced the complexity of the images through the feature extraction model, when the resulting simplified queue pattern images serve as the inputs of the VGG16 model, we can reduce the computational cost of the model without compromising the prediction accuracy of the dissipation time. The details will be given in the next subsection. To automatically generate the label data of the dissipation time prediction model at the training phase, the SORT [39] algorithm is used for vehicle tracking as shown in Fig. 3.

In the region with the yellow boundaries in Fig. 3, the queue pattern is captured at the start time of the green light, denoted as  $t_s$ . Through SORT, we track all detected vehicles in the queue during the effective green time until they are dissipated, and then record the time  $t_i$  when vehicle  $i$  is discharged for  $i = 1, 2, \dots, n$ , where  $n$  is the number of the detected vehicles in the queue. Then we compute the queue-pattern dissipation time,  $t_l$ , as the label data of the queue dissipation time prediction model as follows:

$$t_l = \max(t_i) - t_s \text{ for } i = 1, 2, \dots, n \quad (1)$$

Eq. (1) computes  $t_l$  as the time for the queue to dissipate in the specific intersection.

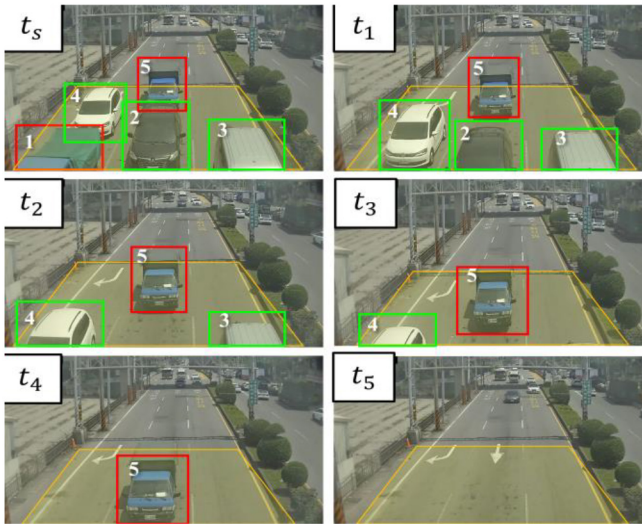


FIGURE 3. The procedure for generating the label data of the dissipation time prediction model (for training).

### C. TRANSFER LEARNING AND MODEL REDUCTION

To predict the dissipation time from the vehicle density maps, we first apply transfer learning [40] to modify the VGG16 model for prediction. The TrafficTalk CNN architecture is divided into the input block (Fig. 4 (1)), the convolutional block (Fig. 4 (2)), and the fully connected block (FC block; Fig. 4 (3)). In Fig. 4, a white circle represents a convolutional layer with the activation function of ReLU [41], where the stride is 2 and the kernel size is 3x3. A black circle represents a max pooling layer. Fig. 4 (a) illustrates the pre-trained VGG16 model with the input size 512x512, the convolutional block of thirteen convolutional layers, and the FC block of two fully connected layers.

The above FC block is a modification of the original VGG 16 model through transfer learning. For the purpose of description, we use the notation  $FC(x, y)$  to represent a fully connected layer with the kernel number  $x$  and the activation function  $y$ . The original VGG 16 model consists of two fully connected layers of 4096 kernels with the ReLU activation function (Fig. 5 (a)), and the second  $FC(4096, ReLU)$  is followed by the fully connected layer of 1000 kernels with the softmax activation function.

Through transfer learning, we replace the FC block in Fig. 5 (a) with the simplified FC block in Fig. 5 (b). The simplified FC block consists of a connected layer of 256 kernels with the ReLU function followed by a fully connected layer of one kernel with the ReLU activation function. In the second connected layer, ReLU is selected because the output of TrafficTalk prediction is dissipation time instead of multiple classes (that are typically identified by softmax). Our experiments indicate that transfer learning reduces the computation complexity of the FC block. It also improves the MAPE by 2.059%, reduces the MAE by 13.3255%, and improves  $R^2$  by 17.5581%.

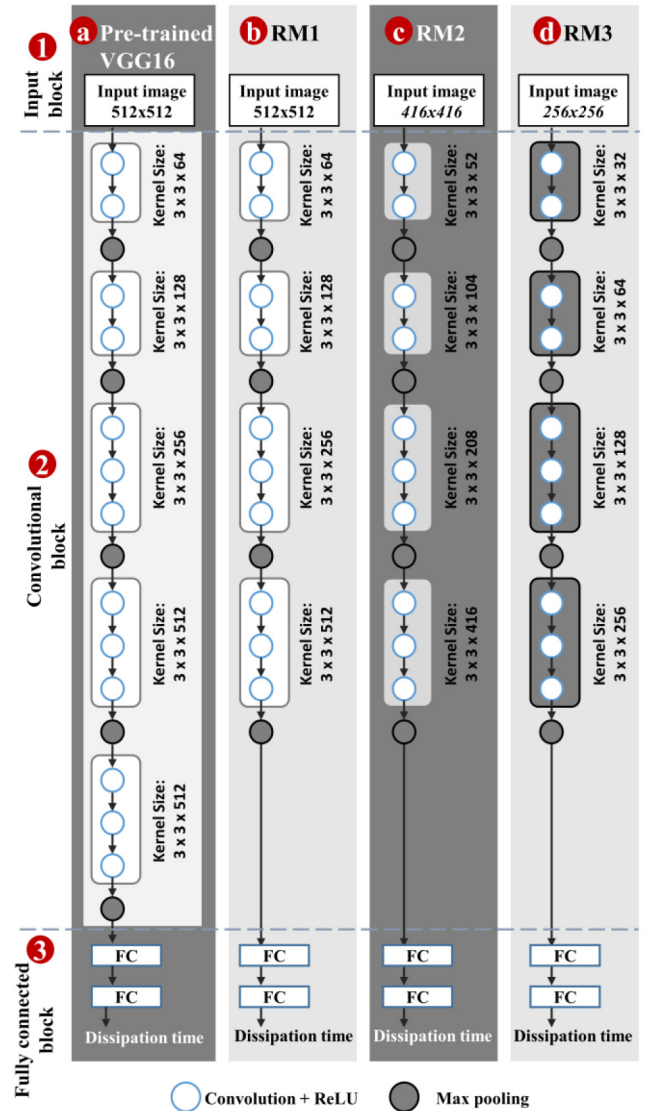


FIGURE 4. The structures of the pre-trained VGG16 and the reduced models.

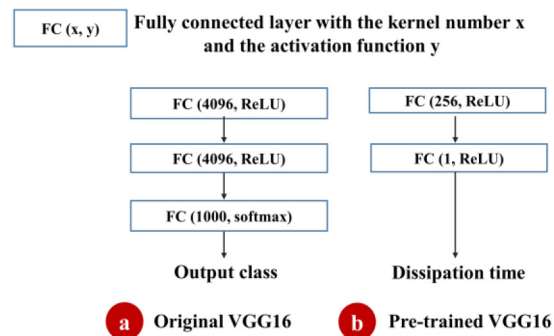


FIGURE 5. The fully connected blocks of the original VGG16 and the pre-trained VGG16 models.

In TrafficTalk, we use the weights obtained from the VGG16 model pre-trained on ImageNet to construct a fixed convolution feature extractor ((2) of Fig. 4 (a)) to extract specific features of the image (e.g., the color and the shape of

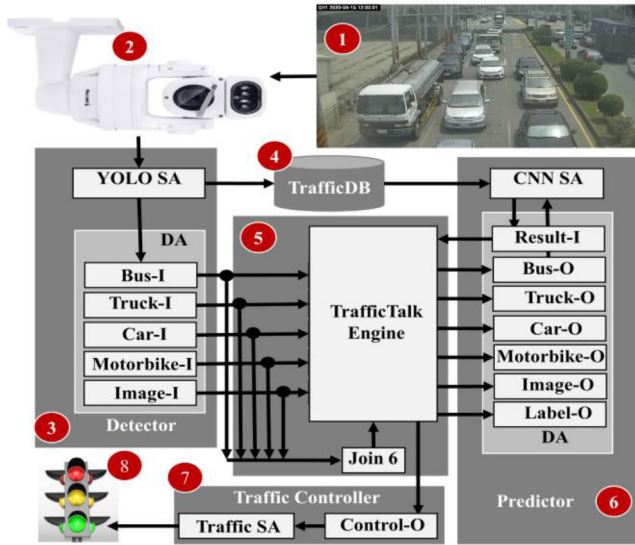


FIGURE 6. The TrafficTalk architecture.

the images in Fig. 2). These features are used for fine-tuning the model.

Through feature extraction in Section III-A, we have reduced the complexity of the input for the VGG16 model from the queue pattern images to the vehicle density maps. Therefore, we can reduce the computational cost of the VGG16 model by adjusting the model structure including the input size of queue-pattern images ((1) of Fig. 4), the number of layers in the convolutional block, and the number of kernels in each convolutional layer ((2) of Fig. 4). Three reduction models are described as follows:

- In Reduction Model 1 (RM1; see Fig. 4 (b)), the last three convolutional layers in (2) of Fig. 4 (a) are removed. Because the vehicles in the input images are transformed into simple geometric shapes, and the color distribution of the input images is simplified into red, green, and blue densities, the simplified RM1 can still effectively extract features with fewer convolutional layers.
- Reduction Model 2 (RM2; see Fig. 4 (c)) reduces the resolutions and the kernel sizes of RM1's convolutional block are replaced. Note that the layers in the convolutional block are the same for both RM1 and RM2. The number of kernels in each layer of the convolutional block is reduced from (64, 128, 256, 512; see (2) in Fig. 4 (b)) to (52, 104, 208, 416; see (2) in Fig. 4 (c)), and the input size of queue pattern images is reduced from 512x512 ((1) in Fig. 4 (b)) to 416x416 ((1) in Fig. 4 (c)). Since the objects in the vehicle density maps have been simplified, we can resize input images with a lower resolution (i.e., 416x416), and can still capture the same features as the input images with higher resolution (i.e., 512x512).

Reduction Model 3 (RM3; see Fig. 4 (d)) further reduces the resolutions and the kernel sizes of RM2's convolutional

block; i.e., (1) and (2) of Fig. 4 (c) are replaced. Note that the number of layers in the convolutional blocks for both RM2 and RM3 are the same. The number of kernels in each convolutional layer is reduced from (52, 104, 208, 416) to (32, 64, 128, 256), and the input size of queue pattern images is reduced from 416x416 to 256x256. Since the objects in the pre-processed images have been simplified, we try to resize input images with a lower resolution (i.e., 256x256), and we expect that the model still captures the same features as the input images with the higher resolution of 416x416. The effects of model reduction evaluated in Section IV indicate that RM3 does not provide any improvement over RM2.

#### IV. THE TRAFFICTALK ARCHITECTURE

This section describes the TrafficTalk architecture. TrafficTalk detects the real-time queue patterns of mixed traffic flows and predict the corresponding dissipation time. The training data are collected from the real-world videos of Closed-Circuit Television (CCTV) and PiXORD bullet network cameras.

Based on an IoT application development tool called IoTalk [42], TrafficTalk analyzes the video data collected by cameras and controls the signal lights based on dissipation time prediction. In TrafficTalk, the cameras (Fig. 6 (2)) provide real-time video streaming of traffic flows in the intersection (Fig. 6 (1)). Two AI tools are deployed as software IoT devices [43], including YOLOv4 and CNN. The Detector (Fig. 6 (3)) continuously takes snapshots from the video stream in real time, detects vehicle objects from the background using the YOLOv4 model and stores them in the Traffic DataBase (TrafficDB; Fig. 6 (4)). The Detector sends the coordinates of the extracted features (Bus-I, Truck-I, Car-I, Motorcycle-I) and the HTML file path of the image (Image-I) to the TrafficTalk Engine (Fig. 6 (5)). The features appended with "-I" are called input features, and the measured data of input features are sent to the TrafficTalk Engine. The engine may process the received data (to be elaborated later), and then forwards them to the Predictor (Fig. 6 (6)). The data sent from the engine to the Predictor are received by the output features appended with "-O". The CNN model uses the data of the output features (the file paths) to retrieve the images in the TrafficDB, and then dynamically predicts the dissipation time. The results are sent to the Traffic Controller (Fig. 6 (7)) through the TrafficTalk engine. The traffic Controller then determines the control of the traffic lights (Fig. 6 (8)). In this system, the TrafficTalk engine is responsible for dispatching messages among the software IoT devices (i.e., the Detector, the Predictor, and the Traffic Control devices).

In TrafficTalk, the software of an IoT device consists of two parts, the Device Application (DA) and the Sensor & Actuator Application (SA). The DA is responsible for the connection to the TrafficTalk server (that is, (3)→(5), (6)→(5), (5)→(6) and (5)→(7)) using the HTTPS and the

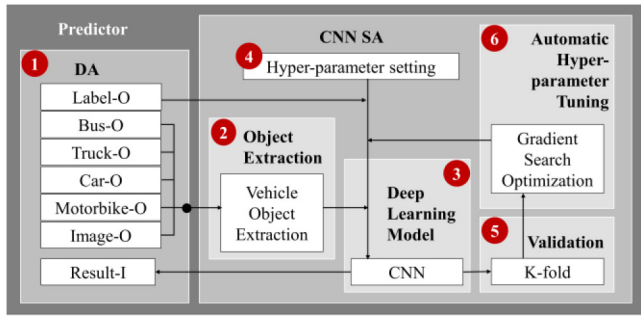


FIGURE 7. The CNN SA.

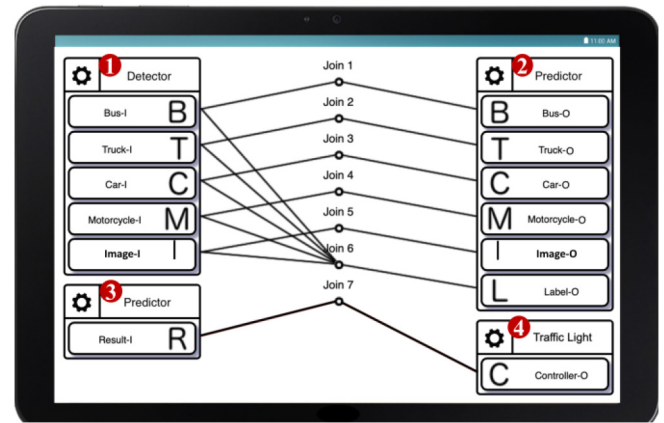


FIGURE 8. The TrafficTalk GUI.

MQTT protocols. The lower-layer communication technologies for the DA can be wired (Ethernet) or wireless (LTE, 5G, or WiFi).

The SA implements the intelligence of the IoT device. For example, the YOLO SA in Fig. 6 (3) implements the YOLOv4 model to identify the types of vehicles and detect their positions in the snapshots of the streaming videos. With the detected types and locations of the vehicles, The Predictor (Fig. 6 (6)) derives the queue patterns emerging in the traffic videos. Details of the CNN SA for the Predictor are illustrated in Fig. 7. The vehicle detection results in Fig. 7 (1) are sent to the extraction module (Fig. 7 (2)) for image preprocessing of vehicle density map generation (Fig. 2 (c)) described in Section III. The dissipation time label (Label-O; the first output feature in Fig. 7 (1)) and hyper-parameter setting (Fig. 7 (4)) are used in the queue dissipation time prediction model described in Section III-C (Fig. 7 (3)). The results generated by the preliminary model are used to conduct stratified k-fold cross validation (Fig. 7 (5)) to adjust the hyper-parameters (Fig. 7 (6)) by gradient search optimization, and then fed back to the CNN model for fine-tuning. The dissipation time prediction results are sent to the Traffic Controller (Fig. 6 (7)) through the input device feature Result-I in Fig. 7 (1).

The connections among the IoT devices are easily configured using the TrafficTalk graphical user interface (GUI) illustrated in Fig. 8. This GUI can be conveniently accessed by an arbitrary computing device with a Web browser. In the TrafficTalk GUI, a software IoT device with input features (for example, the Detector) is represented by a “device model” icon placed in the left of the GUI window (Fig. 8 (1)). The input features are represented by small icons grouped within the device model icon. A software IoT device with output features (for example, the Traffic Controller) is represented by a device model icon placed in the right of the GUI window (Fig. 8 (4)). An IoT device with both input and output features (for example, the Predictor) is represented by two device model icons placed in the right (Fig. 8 (2)) and the left (Fig. 8 (3)) of the window.

To create a path (3)→(5)→(6) in Fig. 6, we simply drag a line from an input feature in the Detector icon to an output feature in the Predictor icon in Fig. 8. For example, Bus-I

and Bus-O are connected through the Join 1 line. Therefore, to create the configuration described in Fig. 6, we simply make the connections Joins 1-7 in Fig. 8.

The Join 6 connection in Fig. 8 merits further discussion. In the CNN model, the labels are required to validate the prediction results for training. Through the vehicle objects identified by the Detector, the vehicle tracking algorithm implemented in Section III-C calculates the actual dissipation time using Eq. (1) to produce the ground truth labels. By clicking the circle in the middle of the Join 6 link in Fig. 8, a window pops up for one to write the Python vehicle tracking function (i.e., Eq. (1)) with the inputs received from input features of the Detector. The reader is referred to [42] for the details.

Based on the queue patterns, the Predictor computes the corresponding queue dissipation times. By connecting Result-I to Controller-O through Join 8, the Predictor sends the predicted dissipation time to the Traffic Controller for making decisions to switch traffic signal lights.

Implementation of TrafficTalk (Figs. 6, 7 and 8) are guaranteed by three tools provided by IoTalk. The VerificationTalk tool [48] guarantees the codes in Figs. 6 and 7 are made safe from failure. The BigraphTalk tool [49] guarantees that the connections in Fig. 8 are correct.

## V. EXPERIMENTS AND RESULTS

This section first describes how we collected the queue pattern data. Then we conduct experiments to investigate the performance of TrafficTalk in terms of the queue dissipation time prediction and the inference time complexity.

### A. QUEUE PATTERN DATASETS

We established three datasets by collecting the traffic video data from different signalized intersections:

- Dataset 1 collects the data from the PiXORD bullet network cameras at Ping’an Rd. and Zhongzheng E. Rd. intersection in Dayuan Dist., Taoyuan City (Fig. 9 (a)). The collection periods range from 7:00 a.m. to 5:00 p.m. on 04/15/2020, 02/24/2020, 03/01/2020, 12/08/2020, 12/09/2020, and 12/12/2020.

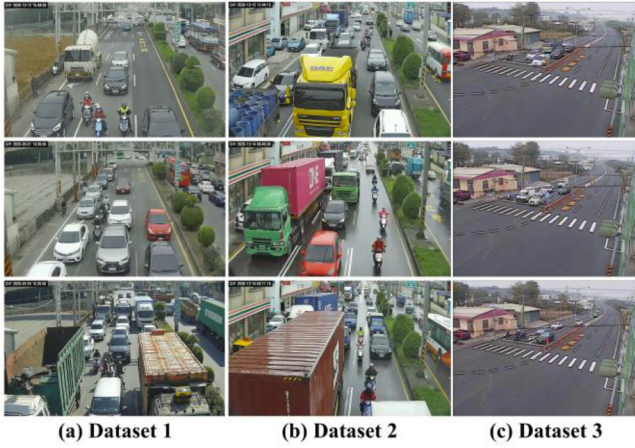


FIGURE 9. Locations for the queue pattern datasets.

- Dataset 2 collects the data from the PiXORD bullet network cameras at Hengnan Rd. and Zhongzheng E. Rd. intersection in Dayuan Dist., Taoyuan City (Fig. 9 (b)). The collection period ranges from 7:00 a.m. to 5:00 p.m. on 12/12/2020.
- Dataset 3 collects the data from the CCTV at Aly. 1, Ln. 1, Xipu Rd. intersection in Dashu Dist., Kaohsiung City (Fig. 9 (c)). The collection periods range from 8:00 a.m. to 5:00 p.m. on 10/30/2019, 10/31/2019, 02/20/2020, 02/24/2020, and 02/25/2020.

By excluding the images with empty queues and outliers caused by traffic accidents or other special incidents, there are 981,295,415 valid images. In TrafficTalk, 60% of the data are used for training, 20% are used for validation, and 20% are used for testing.

## B. EXPERIMENTS FOR QUEUE DISSIPATION TIME PREDICTION

The MAE in Eq. (2), the MAPE in Eq. (3), and the  $R^2$  in Eq. (4) are used to evaluate the accuracy of the TrafficTalk model, where

$$MAE = \left[ n^{-1} \sum_{i=1}^n |P_i - O_i| \right] \quad (2)$$

$$MAPE = \left[ 100 * n^{-1} \sum_{i=1}^n \left| \frac{P_i - O_i}{O_i} \right| \right] \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

In Eqs. (2)–(4),  $P_i$  is the  $i$ -th model prediction,  $O_i$  is the pairwise matched observation for  $i = 1, 2, \dots, n$ , and  $\bar{O}$  is the mean value of all observations.

According to [44], a model with MAPE less than 10% is a highly accurate forecasting model; a model with MAPE 10%-20% is a good forecasting model; a model with MAPE 20%-50% is a reasonable forecasting model, and a model with MAPE more than 50% is an inaccurate forecasting model.

TABLE 1. MAE and MAPE comparison between different models.

Model	MAE	MAPE	$R^2$
Pre-trained VGG16 (Original Image)	3.7508	33.1414	0.5758
Pre-trained VGG16	3.1215	28.4868	0.6769
RM1	2.9563	21.2266	0.8299
RM2	<b>2.4516</b>	<b>10.3771</b>	<b>0.9145</b>
RM3	3.2982	30.9507	0.6391
Linear Regression	3.6123	20.3392	0.8435
Random Forest	3.6922	27.9360	0.7099
SVR	3.3761	17.6532	0.8740
XGBoost	3.1847	25.7848	0.7190

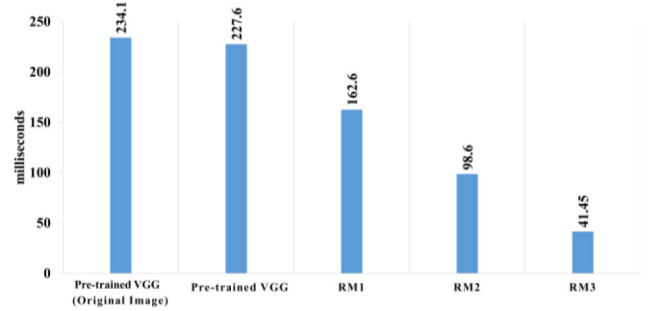


FIGURE 10. The inference time comparison between different models.

Table 1 shows the MAE/MAPE/ $R^2$  values and Fig. 10 shows the inference times for TrafficTalk based on VGG16 and model reductions RM1, RM2, and RM3. Without the image preprocessing of object detection and extraction, the MAPE of VGG16 using the original images is 33.1414%, and the inference time is 234.1 milliseconds. On the other hand, the MAPE of VGG16 using the vehicle density maps is 28.4868%, which indicates that object detection and extraction can significantly reduce the error on predicting the queue pattern dissipation time. Since the vehicle density map significantly reduces the complexity of an image, the pre-trained VGG16 may overfit in predicting the queue pattern dissipation time. Therefore, it is required to conduct model reduction to avoid overfitting in predicting the queue pattern dissipation time. Table 1 and Fig. 10 show that with model reduction, the MAPE is 21.2266% for RM1, and the inference time is decreased from 234.1 milliseconds to 162.6 milliseconds. The MAPE is 10.3771% for RM2, and the inference time is decreased to 98.6 milliseconds, which is an improvement of 57% over the VGG16 model. The experiments show that the effectiveness and efficiency of queue pattern dissipation time prediction can be significantly improved with model reduction RM2. On the other hand, the MAPE of RM3 is 30.9507%, which is worse than that of RM2. This result indicates that RM3 has over-reduced the resolutions and the kernel sizes of RM2's convolutional block.

We have also implemented four regression models (e.g., linear regression, random forest [45], support vector regression (SVR) [46], and XGBoost [47]). Table 1 lists the MAEs, the MAPEs and the  $R^2$  values for all models considered in this paper. The MAPE is 20.3392% for linear



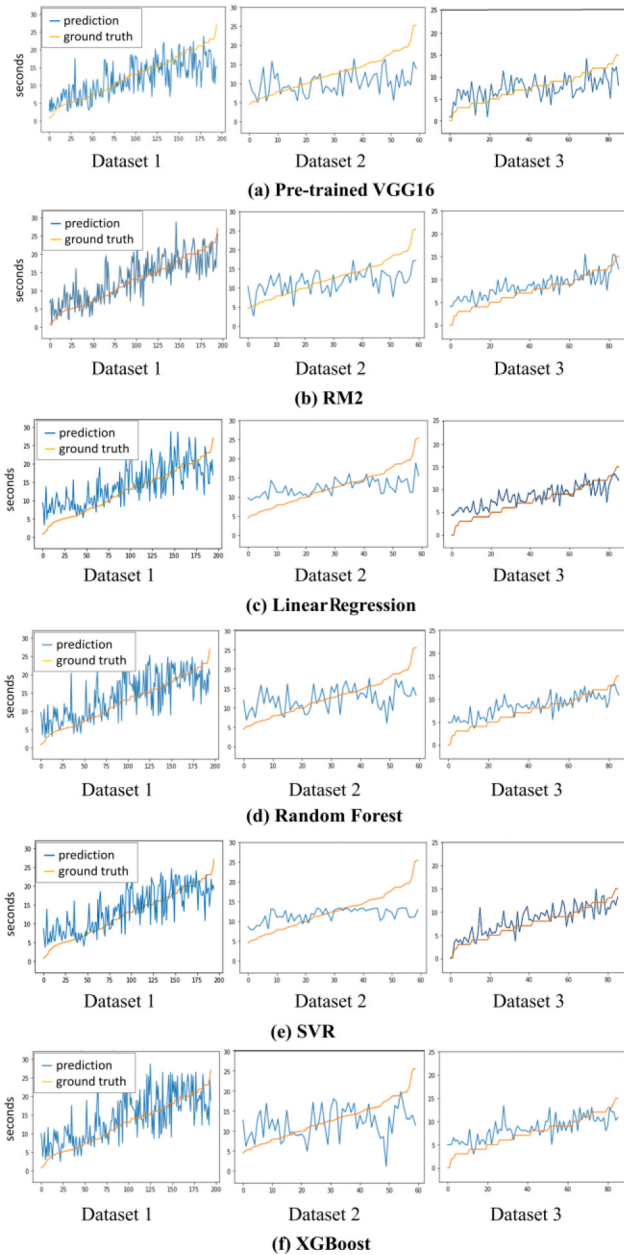


FIGURE 11. The prediction results of the six models.

regression, 27.9360% for random forest, 17.6532% for SVR, and 25.7848% for XGBoost.

Fig. 11 illustrates the queue dissipation time prediction results for the various machine learning models based on Dataset 1, Dataset 2, and Dataset 3. In this figure, the orange lines represent the ground truths of the queue dissipation time, and the blue lines stand for the queue dissipation time prediction results. We observe that the queue dissipation time prediction results generated by RM2 (Fig. 11 (b)) are more in line with the ground truths of the queue dissipation time.

We conclude that RM2 has better MAE, MAPE and  $R^2$  than other models. The computation overhead for RM2 is also reasonably small.

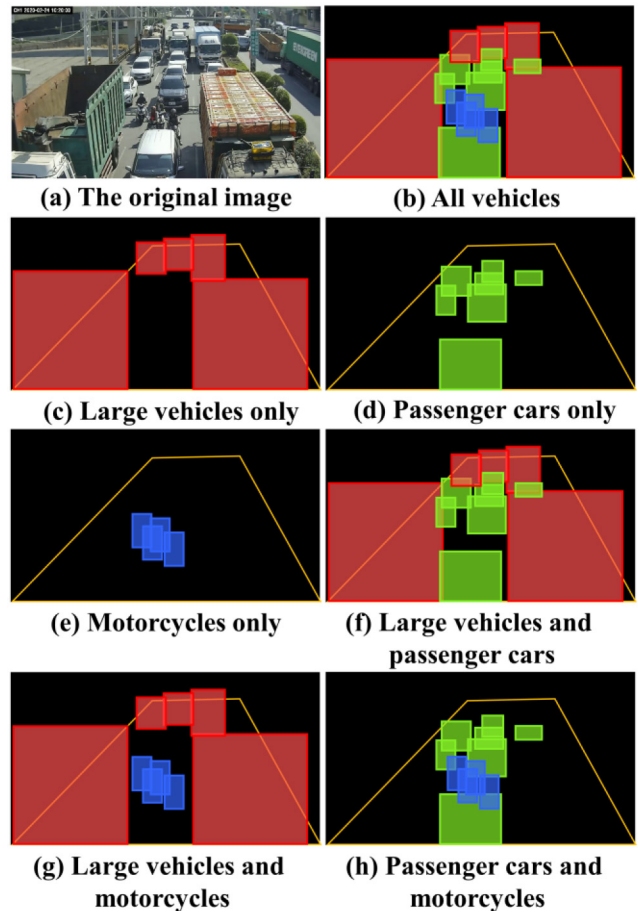


FIGURE 12. The object extraction results for various selected types of vehicles combinations.

### C. EFFECTS OF TYPES OF VEHICLES

This subsection conducts experiments to investigate the effects of mixed traffic flows by considering selected types of vehicles. Specifically, after vehicle object extraction (see Section III-A), the image in Fig. 12 (a) is extracted as a vehicle density map with various types of vehicles. Specifically, Fig. 12 (b) is the vehicle density map  $\{L, P, M\}$  including all types of vehicles; Fig. 12 (c) is the vehicle density map  $\{L\}$  with the large vehicles only; Fig. 12 (d) is the vehicle density map  $\{P\}$  with the passenger cars only; Fig. 12 (e) is the vehicle density map  $\{M\}$  with the motorcycles only; Fig. 12 (f) is the vehicle density map  $\{L, P\}$  without the motorcycles; Fig. 12 (g) is the vehicle density map  $\{L, M\}$  without the passenger cars; Fig. 12 (h) is the vehicle density map  $\{P, M\}$  without the large vehicles.

Table 2 lists the MAE, MAPE and  $R^2$  measures of the RM2 model for various queue patterns in Fig. 12. Denote “>” as “better than”. Then Table 2 indicates that  $\{L,P,M\} > \{L,P\} > \{P,M\} > \{L,M\} > \{P\} > \{L\} > \{M\}$ .

Compared with merely considering the large vehicles and passenger cars, additionally consideration of the motorcycles can reduce the MAPE from 31.2912% to 10.3771%, which is an improvement of over 66%.

**TABLE 2.** MAE and MAPE of RM2 for patterns with extracted types of vehicles.

Dataset	MAE	MAPE	R <sup>2</sup>
{L}	4.2327	42.7598	0.3976
{P}	4.0044	34.7495	0.5268
{M}	4.8827	51.1418	0.3417
{L,P}	3.3620	31.2912	0.6107
{L,M}	4.1421	38.6837	0.4908
{P,M}	3.7832	33.9117	0.5402
{L,P,M}	<b>2.4516</b>	<b>10.3771</b>	<b>0.9145</b>

## VI. CONCLUSION

We proposed a novel method called TrafficTalk, which predicts the dissipation time for traffic signal timing plan design of signalized intersections for multiple types of vehicles including motorcycles. TrafficTalk can estimate the dissipation time of various queue patterns consisting of different types of vehicles. Compared with the previous approaches, TrafficTalk has better feasibility and stability by further considering the mixed traffic flows including, for example, the characteristics of parallel dissipation of motorcycles. Moreover, TrafficTalk provides appropriate settings for TSCT. Experiments show that TrafficTalk accurately and efficiently predicts the dissipation time of different queue patterns. By extracting the detected vehicle objects in an image to produce a vehicle density map, TrafficTalk reduces the complexity of the VGG16 model and obtains more accurate prediction results with model reduction RM2. Specifically, RM2 achieves the lowest MAPE of 10.4% among all machine learning models considered in this paper (i.e., VGG16, linear regression, random forest, SVR and XGBoost). By considering motorcycles in the traffic flows, the MAPE can be improved over 66%, the MAE can be reduced over 35%, and the R<sup>2</sup> can be improved over 49%.

In the future, we will investigate TrafficTalk by using other country's traffic datasets such as the KTTI dataset. Also, we are negotiating with the Kaohsiung city to open the city traffic datasets for public access.

## REFERENCES

- [1] Y. Jiang, "Traffic capacity, speed, and queue-discharge rate of Indiana's four-lane freeway work zones," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1657, no. 1, pp. 10–17, Jan. 1999.
- [2] M. Chitturi and R. F. Benekohal, "Work zone queue length and delay methodology," *Transp. Lett. Int. J. Transp. Res.*, vol. 2, no. 4, pp. 273–283, Oct. 2010, doi: [10.3328/TL.2010.02.04.273-283](https://doi.org/10.3328/TL.2010.02.04.273-283).
- [3] H. Ramezani and R. F. Benekohal, "Analysis of queue formation and dissipation in work zones," *Procedia Soc. Behav. Sci.*, vol. 16, pp. 450–459, Dec. 2011, doi: [10.1016/j.sbspro.2011.04.466](https://doi.org/10.1016/j.sbspro.2011.04.466).
- [4] X.-Y. Lu, P. Varaiya, R. Horowitz, D. Su, and S. E. Shladover, "A novel freeway traffic control with variable speed limit and coordinated ramp metering," *Transp. Res. Rec.*, vol. 26, no. 2229, pp. 55–65, 2011, doi: [10.3141/2229-07](https://doi.org/10.3141/2229-07).
- [5] O. M. Rouhani and D. Niemeier, "Urban network privatization: Example of a small network," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2221, no. 1, pp. 46–56, Jul. 2011, doi: [10.3141/2221-06](https://doi.org/10.3141/2221-06).
- [6] F. Dion, H. Rakha, and Y.-S. Kang, "Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections," *Transp. Res. Part B Methodol.*, vol. 38, no. 2, pp. 99–122, Feb. 2004, doi: [10.1016/S0191-2615\(03\)00003-1](https://doi.org/10.1016/S0191-2615(03)00003-1).
- [7] F.-B. Lin and D. Cooke, "Modeling of queue dissipation for signal control," *J. Transp. Eng.*, vol. 112, no. 6, p. 593, Nov. 1986.
- [8] S.-M. Sun, J. Chen, and J. Sun "Traffic congestion prediction based on GPS trajectory data," *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 5, 2019, Art. no. 550147719847440, doi: [10.1177/1550147719847440](https://doi.org/10.1177/1550147719847440).
- [9] P. G. Michalopoulos, G. Stephanopoulos, and G. Stephanopoulos, "An application of shock wave theory to traffic signal control," *Transp. Res. B, Methodol.*, vol. 15, no. 1, pp. 35–51, 1981.
- [10] X. Wu and H. X. Liu, "A shockwave profile model for traffic flow on congested urban arterials," *Transp. Res. B, Methodol.*, vol. 45, no. 10, pp. 1768–1786, 2011. [Online]. Available: <https://doi.org/10.1016/j.trb.2011.07.013>
- [11] J.-C. Jin, "Advance traffic signal control systems with emerging technologies," Ph.D. dissertation, Dept. Archit. Built Environ., KTH Roy. Inst. Technol., Stockholm, Sweden, 2018.
- [12] P. Balaji, X. German, and D. Srinivasan, "Urban traffic signal control using reinforcement learning agents," *IET Intell. Transp. Syst.*, vol. 4, no. 3, pp. 177–188, 2010.
- [13] H.-J. Jeon, J.-C. Lee, and K.-M. Sohn, "Artificial intelligence for traffic signal control based solely on video images," *J. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 433–445, 2018.
- [14] H. Wei, G.-J. Zheng, H.-X. Yao, and Z.-H. Li, "IntelliLight: A reinforcement learning approach for intelligent traffic light control," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD)*, Jul. 2018, pp. 2496–2505.
- [15] L. Beeston, R. Blewitt, S. Bulmer, and J. Wilson, *Traffic Modelling Guidelines*, Transp. London, London, U.K., Sep. 2021, pp. 118–119.
- [16] Z. Amini, R. Pedarsani, A. Skabardonis, and P. Varaiya, "Queue-length estimation using real-time traffic data," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2016, pp. 1476–1481, doi: [10.1109/ITSC.2016.779575](https://doi.org/10.1109/ITSC.2016.779575).
- [17] C. Tan, J. Yao, K. Tang, and J. Sun, "Cycle-based queue length estimation for signalized intersections using sparse vehicle trajectory data," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 91–106, Jan. 2021, doi: [10.1109/TITS.2019.2954937](https://doi.org/10.1109/TITS.2019.2954937).
- [18] X. Luo, D. Ma, S. Jin, Y. Gong, and D. Wang, "Queue length estimation for signalized intersections using license plate recognition data," *IEEE Intell. Transp. Syst. Syst. Mag.*, vol. 11, no. 3, pp. 209–220, Jun. 2019, doi: [10.1109/MITS.2019.2919541](https://doi.org/10.1109/MITS.2019.2919541).
- [19] F. Gunes, S. Bayrakli, and A. H. Zaim, "Flow characteristics of traffic flow at signalized intersections and performance comparison using queueing theory," in *Proc. 4th Int. Symp. Multidisciplinary Studies Innovat. Technol. (ISMSIT)*, 2020, pp. 1–9, doi: [10.1109/ISMSIT50672.2020.9255044](https://doi.org/10.1109/ISMSIT50672.2020.9255044).
- [20] Y. Li, H. Chen, and M. Feng, "A novel model for the traffic of urban roads based on queueing theory," in *Proc. Int. Conf. Intell. Comput. Autom. Syst. (ICICAS)*, 2020, pp. 190–194, doi: [10.1109/ICICASS1530.2020.00046](https://doi.org/10.1109/ICICASS1530.2020.00046).
- [21] E. Harahap, D. Darmawan, Y. Fajar, R. Ceha, and A. Rachmiate, "Modeling and simulation of queue waiting time at traffic light intersection," *J. Phys. Conf. Ser.*, vol. 1188, Jan. 2019, Art. no. 012001.
- [22] H. Liu, W. Liang, L. Rai, K. Teng, and S. Wang, "A real-time queue length estimation method based on probe vehicles in CV environment," *IEEE Access*, vol. 7, pp. 20825–20839, 2019, doi: [10.1109/ACCESS.2019.2898424](https://doi.org/10.1109/ACCESS.2019.2898424).
- [23] K. Gao, S. Huang, F. Han, S. Li, W. Wu, and R. Du, "An integrated algorithm for intersection queue length estimation based on IoT in a mixed traffic scenario," *Appl. Sci.*, vol. 10, no. 6, p. 2078, 2020, doi: [10.3390/app10062078](https://doi.org/10.3390/app10062078).
- [24] A. K. Sangaiah, J. S. Ramamoorthi, J. J. P. C. Rodrigues, M. A. Rahman, G. Muhammad, and M. Alrashoud, "LACCVoV: Linear adaptive congestion control with optimization of data dissemination model in vehicle-to-vehicle communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 5319–5328, Aug. 2021, doi: [10.1109/TITS.2020.3041518](https://doi.org/10.1109/TITS.2020.3041518).
- [25] S. Zhang *et al.*, "Joint queue estimation and max pressure control for signalized urban networks with connected vehicles," in *Proc. Forum Integr. Sustain. Transp. Syst. (FISTS)*, 2020, pp. 211–217, doi: [10.1109/FISTS46898.2020.9264889](https://doi.org/10.1109/FISTS46898.2020.9264889).
- [26] B. Li, W. Cheng, and L. Li, "Real-time prediction of lane-based queue lengths for signalized intersections," *J. Adv. Transp.*, vol. 2018, Dec. 2018, Art. no. 5020518. [Online]. Available: <https://doi.org/10.1155/2018/5020518>
- [27] Y. S. Murat and Ö. Baskan, "Modeling vehicle delays at signalized junctions: Artificial neural networks approach," *J. Sci. Ind. Res.*, vol. 65, pp. 558–564, Jul. 2006.

- [28] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, pp. 79–82, Dec. 2005, doi: [10.3354/cr030079](https://doi.org/10.3354/cr030079).
- [29] A. D. Myttenaere, B. Golden, B. L. Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016. [Online]. Available: <https://arxiv.org/abs/1605.02541>
- [30] F. Motawej, R. Bouyekhf, and A. E. Moudni, "A note on artificial intelligence techniques and dissipativity-based approach in traffic signal control for an over-saturated intersection," *IFAC Proc. Vol.*, vol. 44, no. 1, pp. 10721–10726, 2011.
- [31] O. M. Rouhani, "Queue dissipation shockwave speed for signalized intersections," Munich Personal RePEc Archive, Munich, Germany, MPRA Paper No. 53161, 2014.
- [32] A. C. Cameron and F. A. G. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models," *J. Econometrics*, vol. 77, no. 2, pp. 329–342, 1997.
- [33] C. Nebauer, "Evaluation of convolutional neural networks for visual recognition," *IEEE Trans. Neural Netw.*, vol. 9, no. 4, pp. 685–696, Jul. 1998, doi: [10.1109/72.701181](https://doi.org/10.1109/72.701181).
- [34] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, pp. 958–963, doi: [10.1109/ICDAR.2003.1227801](https://doi.org/10.1109/ICDAR.2003.1227801).
- [35] L. Gao and B. M. Alam, "Optimal discharge speed and queue discharge headway at signalized intersections," in *Proc. Transp. Res. Board 94th Annu. Meeting*, 2015, Art. no. 15–5397.
- [36] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [37] S. Calvert, M. Minderhoud, H. Taale, I. Wilmink, and V. L. Knoop, "Traffic assignment and simulation models, state-of-the-art background document, version 2.0," TrafficQuest Centre, Delft, The Netherlands, TrafficQuest Rep., 2016.
- [38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [39] A. Bewley, G. Zongyuan, F. Ramos, and B. Uproft, "Simple online and realtime tracking," in *Proc. ICIP*, 2016, pp. 3464–3468.
- [40] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, p. 9, May 2016. [Online]. Available: <https://doi.org/10.1186/s40537-016-0043-6>
- [41] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
- [42] Y.-B. Lin, Y.-W. Lin, C.-M. Huang, C.-Y. Chih, and P. Lin, "IoTtalk: A management platform for reconfigurable sensor devices," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1552–1562, Oct. 2017, doi: [10.1109/JIOT.2017.2682100](https://doi.org/10.1109/JIOT.2017.2682100).
- [43] Y.-W. Lin, Y.-B. Lin, and C.-Y. Liu, "Altalk: A tutorial to implement AI as IoT devices," *IET Netw.*, vol. 8, no. 3, pp. 195–202, May 2019.
- [44] C. D. Lewis, *Industrial and Business Forecasting Methods*. London, U.K.: Butterworths Sci., 1982, p. 40.
- [45] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [46] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines*. Berkeley, CA, USA: Apress, 2015, pp. 67–80.
- [47] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794, [Online]. Available: <https://arxiv.org/abs/1603.02754>, doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [48] M.-Z. Shieh, Y.-B. Lin, and Y.-J. Hsu, "VerificationTalk: A verification and security mechanism for IoT applications," *Sensors*, vol. 21, no. 22, p. 7449, 2021.
- [49] A. Blair, M.-Z. Shieh, Y.-H. Hu, S. Michele, and Y.-B. Lin, "BigraphTalk: Verified design of IoT applications," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2955–2967, Apr. 2020.