Summary of "Development of AI-Based Vehicle Detection and Tracking System for C-ITS Application"

The given study aims to provide a method for deriving traffic information using a camera installed at an intersection to improve the monitoring system for roads. The approach of combining AI and HD map techniques is the main contribution of the given study, which shows a high chance of improving current traffic monitoring systems.

Introduction

- Urban road traffic is a complex phenomenon caused by interactions among various moving entities, such as vehicles and pedestrians.
- The growth in urban population during the past decades has raised the severity of urban traffic congestion, leading to socioeconomic and environmental problems in modern cities.
- To mitigate this issue, brisk trials have been conducted to apply intelligent transportation systems (ITS) on urban roads.
- In this regard, traffic monitoring is one of the most valuable functions of traffic management systems (TMSs).

This paper provides a method for deriving traffic information using a camera installed at an intersection for improving monitoring performance.

The method:

- uses a deep-learning-based (YOLOv4) approach for image processing for vehicle detection and vehicle type classification.
- then estimates lane-by-lane vehicle trajectories by matching the detected vehicle locations with the high-definition map (HD map)

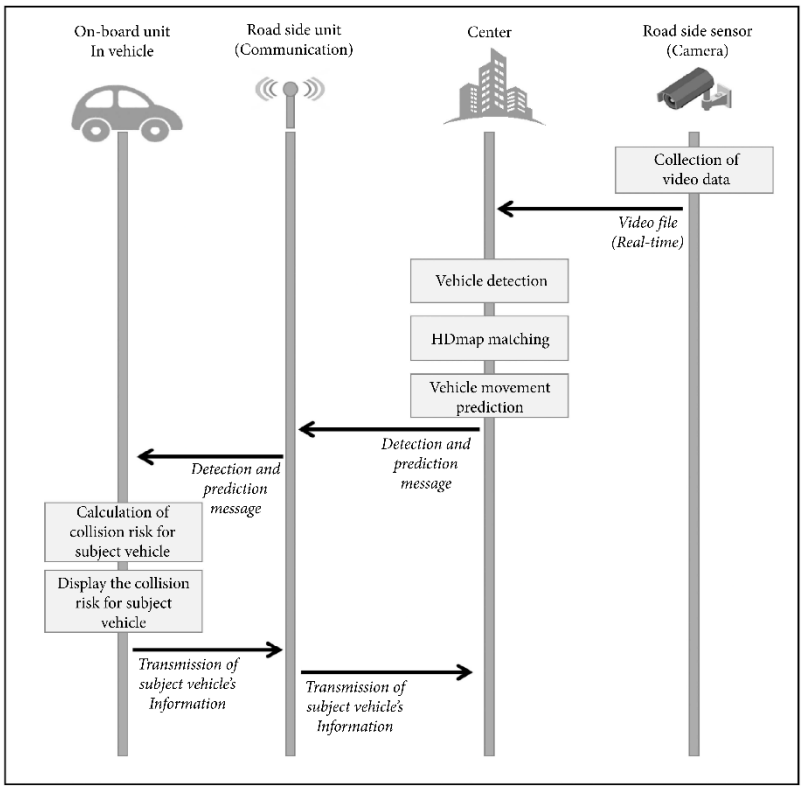
Based on the estimated vehicle trajectories, the traffic volumes of each lane-by-lane traveling direction and the queue lengths of each lane were also estimated.

AI-Based Vehicle Detection System at Intersection

Data Flow Framework

The focus of this study is on how to extract traffic information of multiple vehicles rather than a single vehicle and also on how to deal with traffic information from a traffic monitoring perspective. Hence when designing a camera-based monitoring system, the data flow framework must be considered.

The image below illustrates the data flow framework



Main advantages of the above system are:

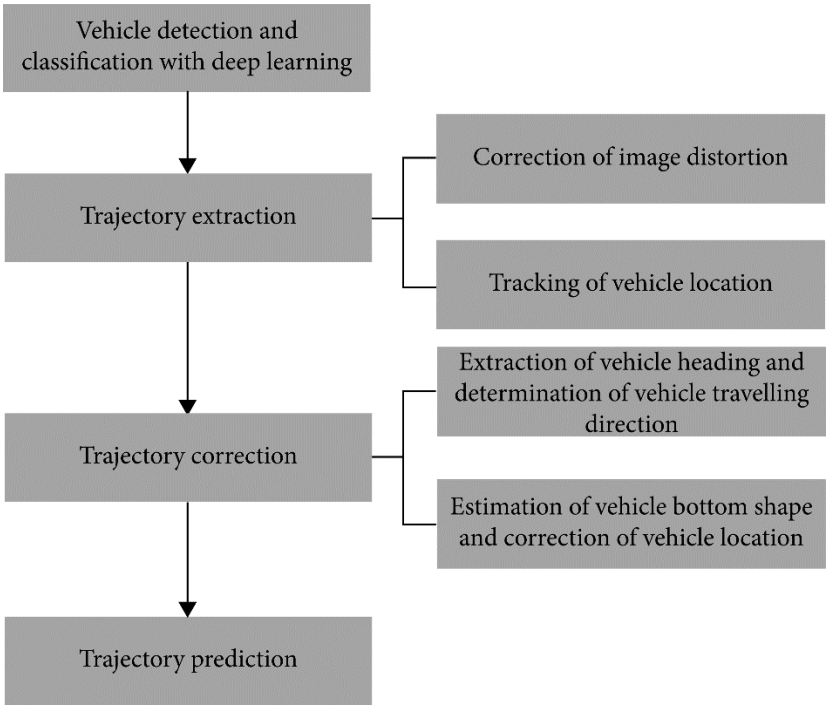
- vehicle-to-vehicle collisions can be prevented by providing vehicles with their detection information traveling through intersections
- a more detailed road status can be provided by extracting lane-by-lane traffic conditions near intersections

This study aims to improve these advantages. The focus of this study is to develop methodologies for AI-based vehicle detection and HD map matching, which are the tasks of the traffic monitoring center described above.

AI-Based Vehicle Detection and Trajectory Prediction

In this study, a deep-learning algorithm is adopted using roadside sensors to extract object information such as vehicle location, movement trajectory, and vehicle speed at intersections and surrounding areas, and useful traffic information, such as traffic volume and queue length, is estimated. The proposed algorithm is based on

- vision data transmitted from the roadside sensors
- to a vision data collecting server located in the traffic monitoring center,
- and the predicted data are stored in a real-time database for real-time data communication.



The proposed algorithm is illustrated in the figure above.

Vehicle Detection and Classification with Deep Learning

They used a deep-learning-based algorithm for vehicle detection as it has higher applicability to real-time traffic monitoring compared to other image processing techniques due to its capability of processing multiple images faster than others. The proposed system performs real-time detection of vehicle location and speed from the vision data sent from the vision data collecting server based on the YOLOv4 deep-learning algorithm and performs vehicle type classification.

Trajectory Extraction

When converting 3D real-world images into 2D images, a distortion occurs, and this distortion causes significant errors between the actual real-world coordinates and the image coordinates. In this study, to remove this error, the corrected vision data were generated from the distorted vision data by inverse application of the camera intrinsic parameters extracted through its calibration. Note that the focal length, principal point, and distortion are the intrinsic parameters of the camera. The values of the intrinsic parameters were determined by projecting a 2D image into 3D world space.

Trajectory Correction

In general, deep-learning-based vehicle detection extracts information in the form of a bounding box, and the central point of the bounding box represents the overall vehicle location information. But sometimes the bounding box does not bind the entire vehicle completely and this results in the central point of the bounding box and the central point of the vehicle to not overlap and cause some error in the location of the vehicle.

This error can lead to another error in the trajectory prediction. This subsequent error can lower the performance of the HD map-matching process, which deals with extracting lane-by-lane traffic information later.

In this study, to reduce the error in center point estimation, real-time correction of vehicle location was performed through the following two steps:

(1) extracting the heading and determining the traveling direction of the vehicle and
(2) estimating the shape of the vehicle bottom and correcting the location.

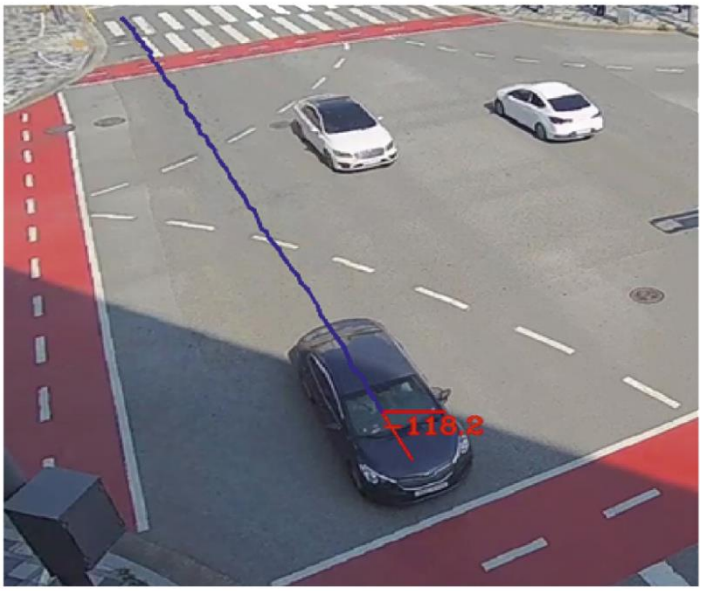
The heading of a vehicle is extracted through the following steps:

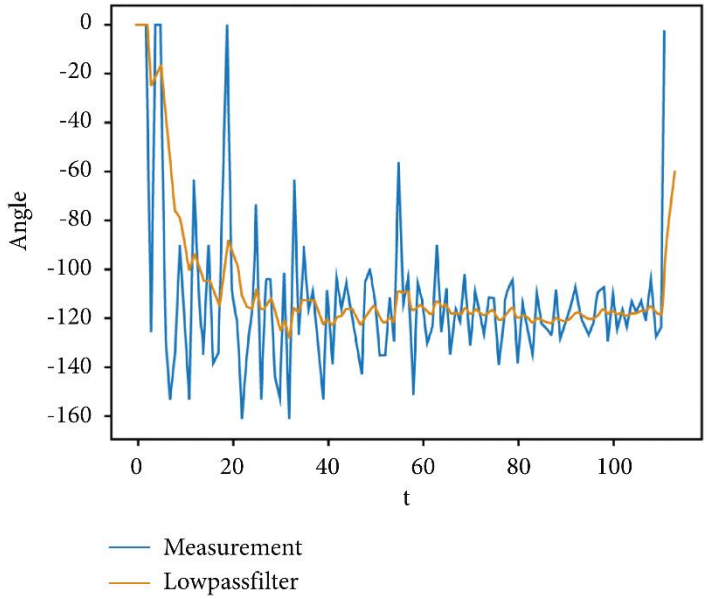
(1) The vehicle position of the previous image frame and the position of the current image frame are converted into coordinates using a transformation matrix.
(2) The angle formed by the two positions is calculated using the Pythagorean equation, and the distance between the two positions is calculated using the coordinate values.

The extracted heading for each frame was corrected based on the low-pass filter as follows:

$$\overline{z_n} = \sigma \cdot \overline{z_{n-1}} + (1 - \sigma) \cdot z_n,$$

where $\overline{z_n}$ is the corrected heading, $\overline{z_{n-1}}$ is the heading at previous time, z_n is the heading at current time, and σ is the weight.





The vehicle traveling direction and the vertical direction are derived using the heading obtained from the real-time estimation and the detected pixel coordinates. The first figure above shows the corrected results of the low-pass filter. The second figure shows the orange and blue colored lines representing the filtered and raw data, respectively. Noisy data points and variations in heading information are smoothed using a low-pass filter.

The bottom surface information is estimated based on the following steps:

(1) The center points of the bounding boxes in the previous image frame and in the current image frame are converted into Transverse Mercator coordinates.
(2) Since the vector formed by the two center points is the moving direction of the vehicle, a hypothetical vector perpendicular to the moving direction is drawn to create a rectangular vehicle bottoms shape (assuming that vehicles have a rectangular shape from the top view).
(3) The height between camera and ground surface, the height of a vehicle, the distance on the surface between camera and vehicle, and the distance on the surface between the camera and the point where the line connecting between the camera and the top of the vehicle meets the

surface are directly obtained from image data. can be calculated by the triangle proportional theorem. Note that the height of the vehicle is assumed to be half of the actual height because the center point of the bounding box detected in the image is half the actual height. Based on this method, the four corner points (in 3D coordinates) of the vehicle bottom are estimated.

(4) The 3D coordinates of the vehicle bottom are then converted into the image coordinates using an inverse transformation matrix, and this finalizes estimating the vehicle bottom. The center point information of the vehicle's bottom surface is extracted based on the estimated pixel information of the bottom surface, and the final pixel-based location information of the vehicle is derived based on this information.

Trajectory Prediction

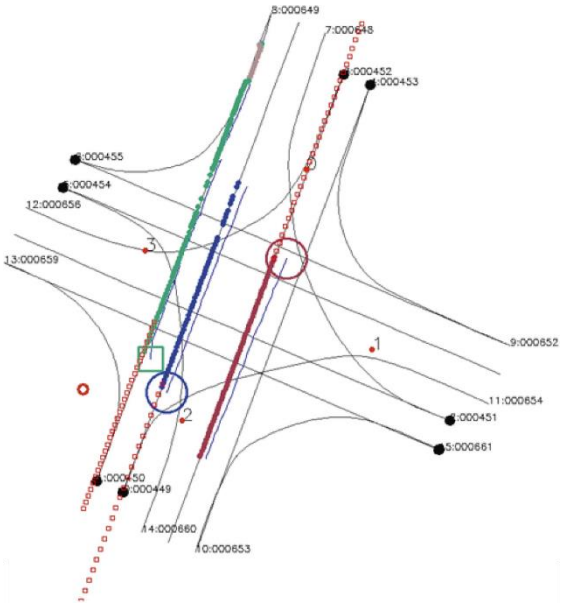
Using the previously derived real-time trajectory data of the vehicle, the upcoming vehicle trajectory information from 1 to 3 seconds was estimated. Location information for each time slot was used to estimate the future trajectory of the vehicle. In addition, a polynomial curve fitting algorithm was used, by applying a linear equation if the past data is a vehicle traveling forward or a quadratic equation for a turning vehicle, to extract the future location of the vehicle.

When estimating the future location of the vehicle based on the detected vehicle location information alone, the result showed that the prediction performance is decreased at the intersection approach where a fewer number of points exist in the trajectory data. To address this limitation, the HD map previously built at the intersection was used, as shown by the solid black lines in the figure given below. Using the location information per link in the HD map, the future vehicle location was estimated assuming that the vehicle trajectory will follow the shape of the HD map link, and the estimated result is shown in the given figure.

Blue Solid Line = Ground truth

Blue- and Green-Dotted Line=Link of the HD map where the detected vehicle is assigned

Red-Dotted Line= Estimated future location of the line

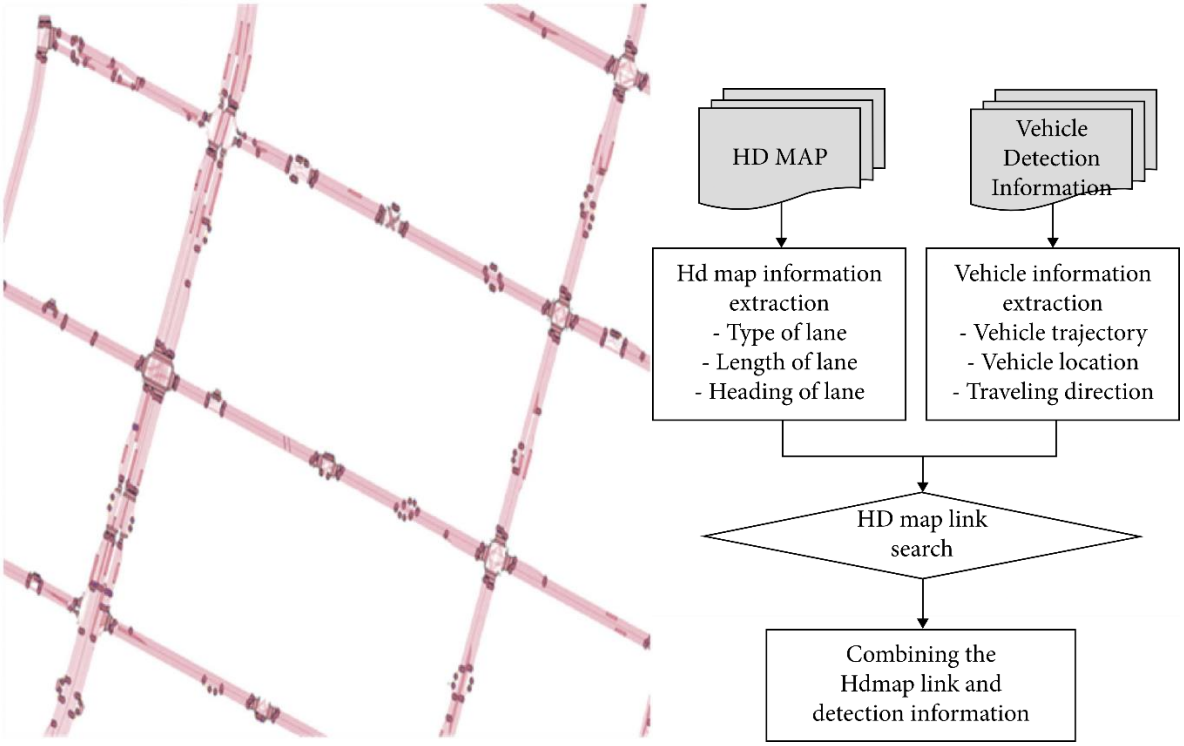


Provision of V2X Communication-Based Detection Information

Generation of HD Map-Based Information

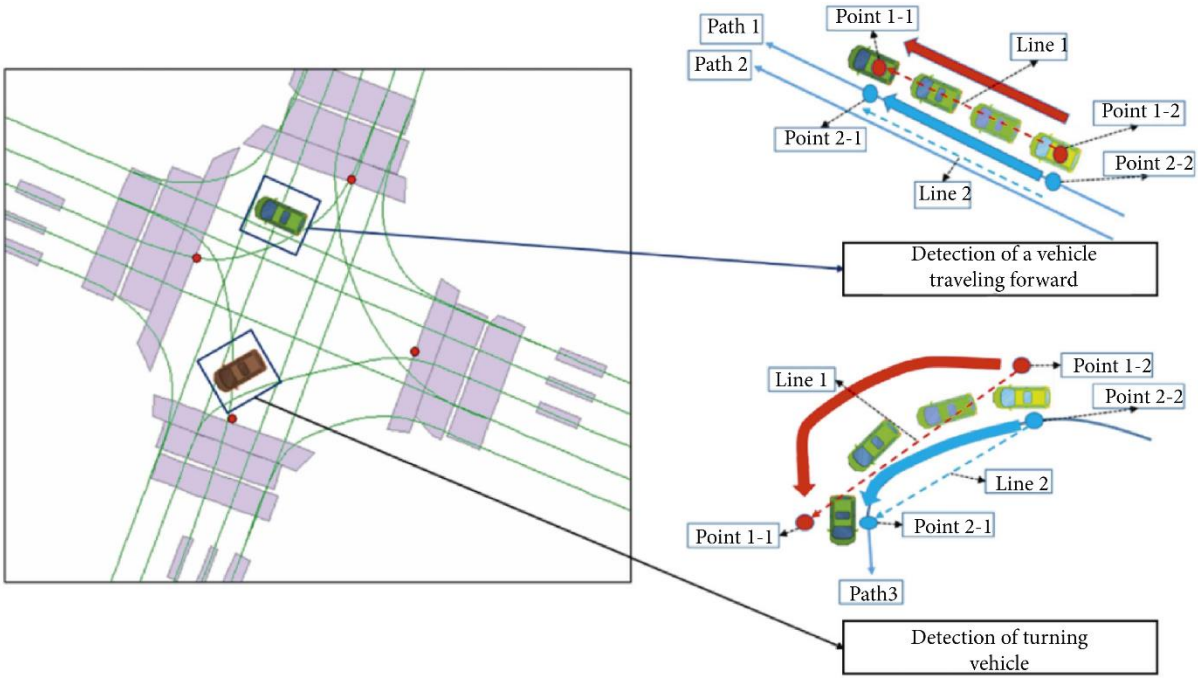
The current C-ITS provides information such as unforeseen incidents or accidents via messages that include longitude and latitude data. This becomes disadvantageous as the number of messages increases sharply with the number of related pieces of information. The computational load increases rapidly as the number of messages increases when matching the predicted vehicle trajectory information and point of event occurrence for each event.

Therefore, in this study, to overcome the limitation in sending location information based on longitude and latitude, the AI-based detection and prediction information provided in the previous subsection was combined with the HD map link information, as shown in given figure



The second figure shows the HD map link allocating algorithm

First, in the process of extracting HD map link information, information such as the length, linearity, and type of link and the longitude and latitude of the start and end points are extracted from the link attribute information of the HD map. This information of the HD map is compared with the detected location coordinates of the vehicle, and matching is performed with the nearest link, extracting the lane on which the vehicle is currently traveling. The below figure shows an example of the HD map link allocation based on the trajectories of the forward-traveling vehicle and turning vehicle. As shown in the figure, information on whether the vehicle travels forward or turns is extracted based on the vehicle trajectory for the past 1 s. Based on this information, if the vehicle is determined to be traveling forward, links with forward-type traveling are extracted from the HD map links, and the extracted candidate links and vehicle trajectory for 1 s are matched based on the start and end points, thereby extracting the HD map link with the closest matching. Finally, the HD map link extracted based on the distance is compared with the heading of the vehicle traveling direction, and when the latter shows consistency within a set threshold, the HD map link is allocated.



To enhance the applicability of the extracted information based on AI, the information extracted from the vision sensor is allocated in HD map link units. Then, the number of vehicles present in the link representing density, the most necessary information in traffic management, and queue length information are generated by the link.

The density is calculated as the difference between the number of vehicles entering the starting point and those leaving the endpoint of the link. As for the queue length of a vehicle, when the average speed over the last 1 s is smaller than the set speed for each HD map link, the corresponding vehicle is classified as the vehicle in the queue. To improve the applicability of the information, the queue length is expressed based on the offset of the HD map.

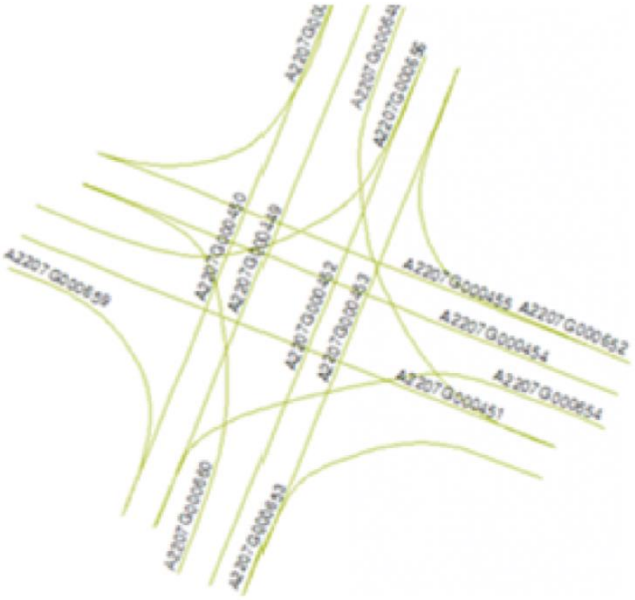
Data Design for V2X Communication-Based Information Provision

Data converted based on the link format of the HD map are stored in the server in two tables which have their own format.

The first table shows the storage format of vehicle information, which is used for storing and sharing object information (vehicle type, longitude, and latitude coordinates) extracted from AI. However, to improve the applicability of the information and accuracy of matching with the vehicle trajectory, the information allocated to the HD map link is combined. In addition, providing predicted information of vehicle objects based on the HD map link ID facilitates the calculation of the probability of collision in the future traveling direction of an autonomous vehicle.

The second shows the storage format of data, primarily processed to facilitate the application of the information extracted from AI-based detection information to the traffic management field. As described above, information on the number of vehicles present in the link (density), queue length information, and average speed information is generated with reference to the HD map link. Similar to the storage format of the vehicle information, the predicted information is provided to facilitate the calculation of the collision probability in the future traveling direction of an autonomous vehicle.

Target Site for Application of the Proposed Method and Evaluation



The above figure shows the target site for applying the technique

The proposed system was evaluated using data collected for three days, and data for accuracy verification were generated in two steps as follows. The ground truth data was captured using a drone, field survey, and manual observation. This data is useful for calculating the accuracy of the system.

Accuracy

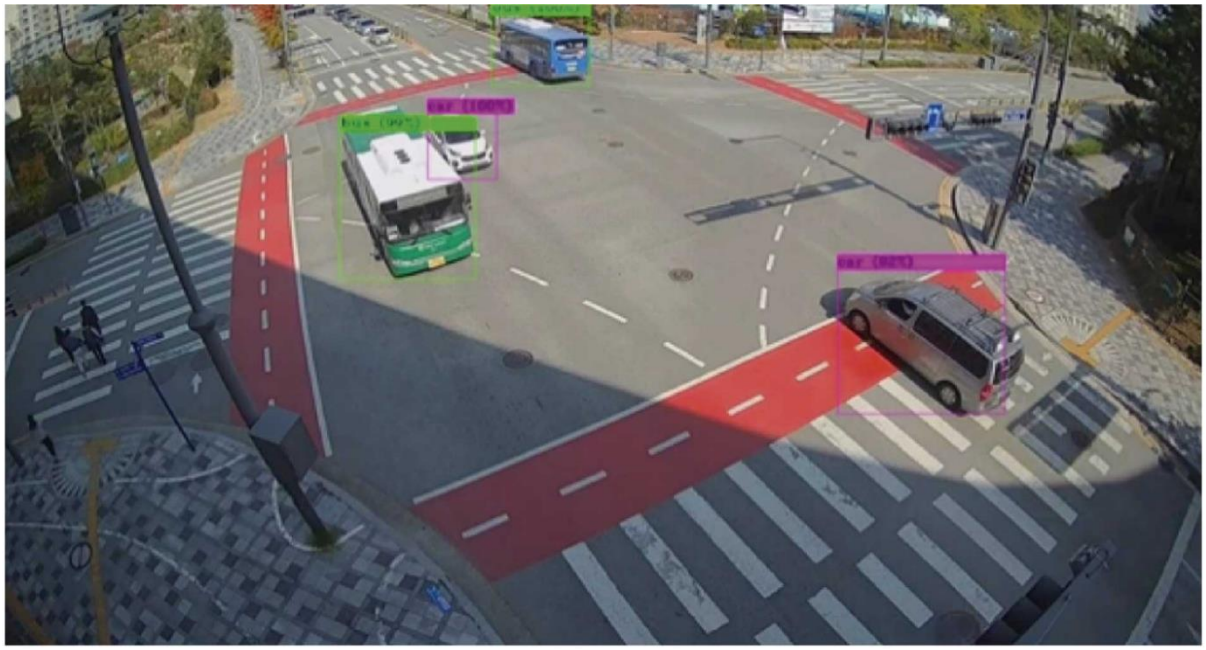
To evaluate the vehicle detection performance, the detection rate was calculated to determine whether all vehicles were successfully detected regardless of vehicle type. The detection rate is defined as the ratio of the total number of detected vehicles to the total number of ground truths

The performance of the vehicle trajectory prediction was evaluated by comparing the predicted and actual trajectories.

Traffic volume was estimated by comparing the number of vehicles counted by the image processing (estimated value) technique and that counted manually (actual value). The evaluation was performed by calculating the root mean square error (RMSE) and mean absolute percentage error (MAPE).

The evaluation of the performance of the queue length estimation is similar to that of the traffic volume estimation. This is done by comparing the queue length in meters derived by the image processing (estimated value) technique and that collected from a drone image (actual value).

Results



The figure above shows the results of applying the AI-based vehicle detection and trajectory prediction

Detection performance of 99% for 6,804 attempts is achieved, which can be judged to be highly consistent.

The performance of vehicle classification is performed in terms of MAP. With 6,804 test samples, the MAP values were 95%, 87%, and 81% for cars, trucks, and buses, respectively. Hence, the proposed method also shows reasonable performance in classifying the vehicle types.

In terms of trajectory prediction, the average Euclidean distance was 1.138 m when 60,531 samples were tested. Such a low degree of error indicates the high performance of the proposed method.

In terms of both traffic volume and queue length estimations, the absolute differences are only 4.20 vehicles for vehicle counting and 3.08 m in queue length estimation upon the RMSE values for more than 60,000 test samples. The MAPE values are less than 20%, which means that the performance of the proposed method is reasonable, particularly when estimating the lane-by-lane traffic information.

Overall, based on the analyses of the five different evaluation criteria, the method proposed in this study shows the feasibility of collecting detailed traffic information with a camera installed at an intersection. In addition, the average time taken from image collection, data processing, and data storage in the server is 0.034 seconds, showing that the performance of the entire process can be completed within 0.1 seconds in general.

Conclusion

It has shown high accuracy in

(1) real-time vehicle detection and classification based on deep-learning-based image processing and
(2) estimating lane-by-lane vehicle trajectories by matching the detected vehicle locations with the HD map

This study is not without limitations. The error rates for both lane-by-lane traffic volume and queue length estimations are greater than 15% even though the vehicle detection showed a 99% performance, which is reasonable but not sufficient in terms of the reliability of traffic information. This is due to intermittent mismatches between the vehicle locations of the camera images and the HD map coordinates.