[1]Joonho Byun
[2]Siwoo Byun

# Deep Learning Model for School Zone Object Detection

Journal of Electrical Systems

***Abstract:*** - Children are more likely to have accidents during school hours because they are slower and less alert than adults. Therefore, more specific information should be proposed for safety management in front of schools. In this paper, we propose a state-of-the-art YOLO-based detection technology that actively responds to various dangerous accidents and events in front of schools, and verify its effectiveness using real-world big data. The dataset used in this experiment consists of Roboflow-based images of objects, which are captured and categorized into 9 types of classes. The experimental results showed a performance with precision of 0.86, mAP50 of 0.85, and processing time of 5.8ms with only 2.7 million parameters. This is probably feasible in real time with low-end portable devices.

***Keywords:*** School Zone; Deep Learning; Yolo; Object Detection; Traffic Sign

## I. INTRODUCTION

There are many accidents involving students around schools. In particular, careless accidents are more likely to happen on the way to school. It is also very dangerous to leave a child who has been involved in an accident unattended in front of the school, even for a short time[1,2].

In particular, the speed limit is 30 km/h within 300 m of the main entrance gate of primary schools and kindergartens, but drivers may not be aware that the road they are traveling on is a school zone. In addition, in non-urban areas, even if there are traffic lights within the protection zone, local authorities may use flashing yellow lights to facilitate traffic flow [3,4].

Children are less likely to look out for vehicles entering school zones, have poor judgment in dangerous situations, and act impulsively. Children are slower and less accurate than those of adults, and their lack of agility makes it difficult for them to take evasive action when faced with a dangerous situation. For these reasons, traffic accidents continue to occur in school zones (Table 1)[3,5].

**Table. 1 Current status of children's accidents by year**

(unit : case, number of people, %)

| division / year | number of accidents | | number of deaths | | number of injured | |
|---|---|---|---|---|---|---|
| | | year-over-year rate of increase | | year-over-year rate of increase | | year-over-year rate of increase |
| 2010 | 14,095 | -5.9 | 126 | -7.4 | 17,178 | -6.5 |
| 2011 | 13,323 | -5.5 | 80 | -36.5 | 16,323 | -5.0 |
| 2012 | 12,497 | -6.2 | 83 | 3.8 | 15,485 | -5.1 |
| 2013 | 11,728 | -6.2 | 82 | -1.2 | 14,437 | -6.8 |
| 2014 | 12,110 | 3.3 | 52 | -36.6 | 14,894 | 3.2 |
| 2015 | 12,191 | 0.7 | 65 | 25.0 | 15,034 | 0.9 |
| 2016 | 11,264 | -7.6 | 71 | 9.2 | 14,215 | -5.4 |
| 2017 | 10,960 | -2.7 | 54 | -23.9 | 13,433 | -5.5 |
| 2018 | 10,009 | 8.7 | 34 | -37.0 | 12,543 | -6.6 |
| 2019 | 11,054 | 10.4 | 28 | -17.6 | 14,115 | 12.5 |
| Annual average rate of change | -2.7 | | -15.4 | | -2.2 | |

[1] Department of Artificial Intelligence, Seogang University, Seoul, Korea

[2]*Department of Software, Anyang University, Anayng, Korea

It is not enough to simply improve CCTV in front of schools to prevent accidents, but at the very least the car's dashcam should be able to identify school zones and warn the driver of the danger. Existing navigation systems do not provide detailed information beyond entering a school zone and do not provide information about pedestrians.

Therefore, an intelligent object detection system should be proposed for safety management in front of schools, and its functional design should be considered for future dash cameras and navigators. In this paper, we propose a detection technology that actively responds to various dangerous events that may occur in front of schools using deep learning, and verify its effectiveness through lightweight deep learning to be used in a small dashcam.

## II. SYSTEM MODEL

### *A. School Zone Safety*

Stop lines, traffic signals, speed bumps, traffic signs, speed cameras, yellow carpets, and pedestrian crossings are used to reduce school zone crashes (Figure 1). However, they can't completely prevent school zone accidents. Since there is no system that can inform drivers about school zones in advance, an AI-based accident prevention system that can recognize traffic conditions in school zones is needed. In other words, AI-based technology that can prevent traffic accidents by informing drivers about school zones and pedestrians in advance is needed[3].

**(a)Detecting stop line violations,**

**(b)Runway-type guidance lights,**

**(c)Focused Lighting**

**Figure 1. Traffic accident prevention facilities**

*B. Interpretive Processing Models*

Image processing technology refers to the entire process of processing and analyzing images acquired from equipment, and consists of functions such as image input and output, preprocessing for digitization, segmentation, and defect detection. [6,7,8]

In the traditional rule-based method, the user models a filter that removes noise from the image to effectively detect objects. However, the disadvantage of this rule-based method is that new filter modeling is required depending on the focal length of the image, the effects of the shooting environment, the shooting quality, the resolution, and so on. In addition, the cameras developed to compensate for this are capable of capturing high resolution images, but are very expensive and therefore not economically viable[6].

To improve the overall performance of these interpretative object detection methods, it is very important to know the statistical characteristics of the target image in advance in order to estimate the relevant parameters appropriately, which usually requires a rather complex implementation to identify additional features of the objects and incorporate them into the design of the algorithm. These analytical approaches have the disadvantage of being inelastic compared to deep learning based methods, which are less robust and can continuously improve their accuracy with additional data, especially when the statistical characteristics of the target images vary due to seasonal and climatic characteristics[9].

*C. AI-based Processing Models*

To compensate for the limitations of traditional techniques, the need for analysis techniques using machine learning such as SVM and KNN[10,11] has become prominent. The SVM is a classification algorithm that classifies two or more pieces of data and is expected to have high generalization performance

The KNN algorithm is a popular method for classification purposes because of its low training effort. The principle is to identify the k labels that are closest to the value you want to learn or predict, and then take the most frequent label as the value. In other words, it assumes a similarity between the new case data and the existing cases, and places the new case in the most similar category.

Deep learning[3,13] has become more prominent in recent years, and related research has been actively conducted. In particular, there have been recent studies on how to inspect and analyze the appearance of large-scale infrastructure using deep learning-based image processing techniques [12,13].

To design an effective model, a sufficient training dataset is required to enable the deep learning model to identify image patterns. Image preprocessing, such as resizing, color correction, and denoising, can improve the model's generalization performance. Image enhancement techniques, including horizontal and vertical flipping, rotation, translation, cropping, color conversion, noise addition, distortion, and deformation, should also be considered.

The choice of deep learning model should be based on the type of objects to be detected and the hardware capabilities available. The more complex the model, the more accurate the predictions, but it may be hardware-intensive or require time-consuming processing that is not feasible in real time. A CNN-based model suitable for image processing can be selected, and after adjusting hyper-parameters such as model structure, learning rate, and number of training runs, the performance of the model can be verified through various evaluation metrics.

For object detection, it is necessary to segment the object using image processing techniques. From an information perspective, the segmentation process converts low-level information (the original image) to higher-level information (the segmented image). It also removes unwanted information from the image, such as noise. Representative segmentation techniques include convolution-based segmentation and autoencoder-based segmentation[9,12].

We have tested many different popular deep learning models such as YOLO and Detectron. YOLO (You Only Look Once) [9], one of the deep learning models that perform object recognition based on CNN. YOLO is a model that focuses on object detection, dividing the image into a grid and simultaneously predicting the bounding box and class probability of the object within each grid cell, and is particularly effective for real-time object detection, providing high accuracy and fast processing speed.

Detectron is a training/inference platform for object detection and instance segmentation. Detectron2 is a model released by FAIR (Facebook Artificial Intelligence Research) and is the latest evolution of the original Detectron1 model. The Detectron1 model was based on the Faster R-CNN. When training with Detectron2, the training process can be abstracted using the engine instead of the training loop we usually implement when building deep learning models, allowing developers to focus on the model development itself.

In this study, we developed a deep learning method and image processing method that can effectively and quickly detect objects and analyze the characteristics of objects in school zones, and verified the performance of the developed method for practical application.

## III. METHODS AND RESULTS

### *A. Dataset*

The dataset used in this experiment is an object dataset consisting of 1213 camera images taken in 2024 and published on a computer vision dataset platform called Roboflow[14], which is aimed at object detection. It is a dataset of RGB images annotated in Coco format and resized to 640 x 640. It is also pre-processed with 50% horizontal flip and 50% vertical flip.

This data set contains the label of the image and the coordinates of the four corners *x_min*, *x_max*, *y_min* and *y_max* of the object. The following code creates a dataset for testing the YOLO v10 model.

```
from roboflow importRoboflow
rf = Roboflow(api_key="2abGJ3rNxbbVOQY7T1gu")
project = rf.workspace("jmk").project("sc-objects")
dataset = project.version(1).download("yolov10")
```

Out of the total images, 87% images were used as training dataset for training, 8% images were used as validation dataset for hyper parameter tuning, and 5% images were used as test dataset for final performance evaluation of the model. The dataset is classified into nine types of objects: child, green light, parking sign, red light, school zone, slow sign, traffic light, and background. Figure 2 is an example image of the data set.


**Figure 2. Example images**

### *B. Deep Learning Models Used*

YOLO[15] is a well-known deep learning model, and v10[16] is the latest version of the YOLO object detection model, which is built on top of the Ultralytics Python package and performs well in a variety of computer vision tasks. YOLO v10 uses the Swin-transformer as its backbone architecture, with strengths in feature representation, context capture, and accuracy, and uses EfficientNetV2 to balance efficiency and accuracy.

To improve the feature extraction ability, YOLO pre-trained the model with a large unlabeled dataset using a self-supervised learning method, and used gradient accumulation and mixed precision learning methods to allow training with larger batch sizes without memory constraints. In addition, the use of focal loss, which can focus more on difficult tasks and compensate for class imbalance, is one of the reasons why it was chosen as a suitable model for our experiment.

Non-maximum suppression (NMS) is an important post-processing technique in object detection that removes multiple redundant prediction boxes, leaving only the most reliable predictions. However, over-reliance on NMS in object recognition models negatively affects the model's ability to perform optimally, and YOLO v10 is designed to utilize consistent double mapping to enable training that does not rely on NMS, which has the advantage of reducing the model's inference speed. Figure 3 shows the model structure of YOLO v10, and Figure 4 shows performance metrics for several models on standard benchmarks such as COCO[16].

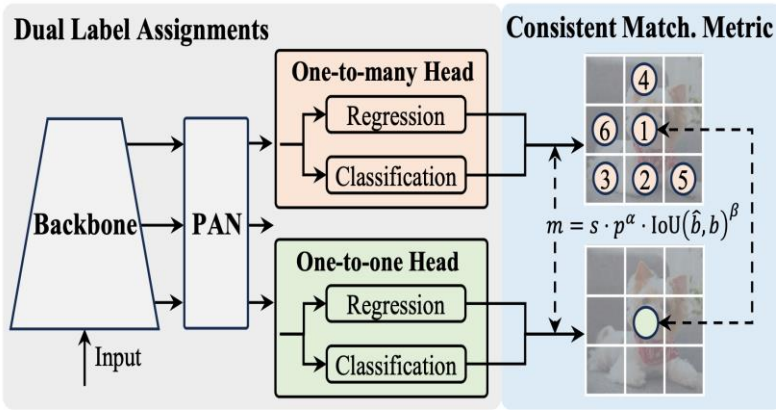


**Figure 3. YOLO v10 Model Structure Diagram**

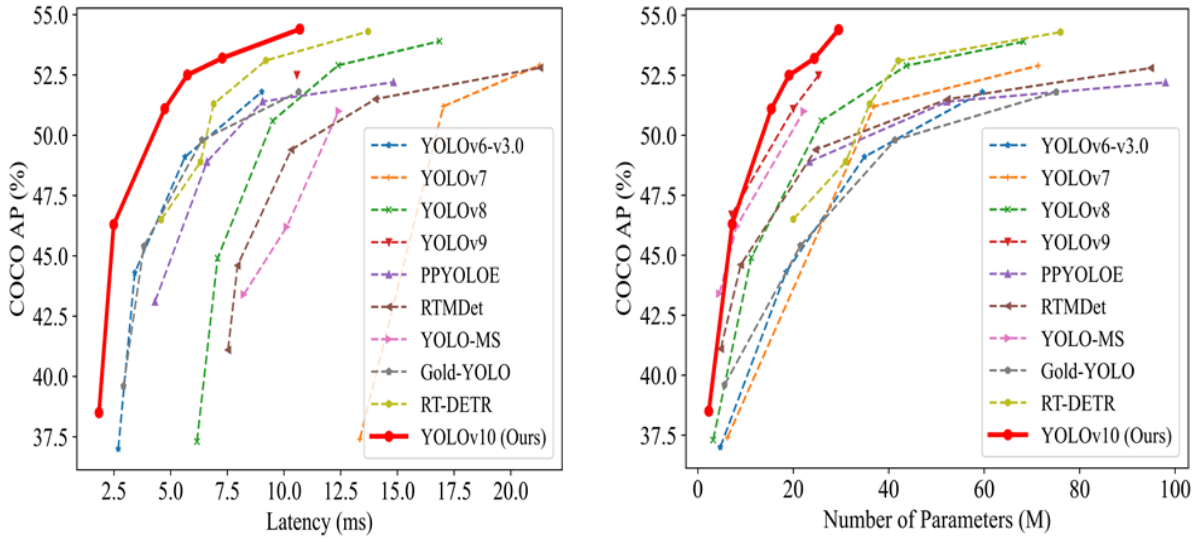


**Figure 4. Comparing the performance of different models**

## IV. EXPERIMENTAL RESULTS

We trained Yolo, which is often used for image processing via CNN, on the imported dataset with the following code. We used *image_size* 640x640, *batch_size* 32, and pre-trained the model with the yolov10.pt file. The Yolo model consists of 285 layers, including convolutional, block, and connection layers, and 2.7 million parameters tuned through training.

```
HOME = '/content/sc-objects-1'

!yolo task=detect mode=train model=yolov10.pt data={HOME}/data.yaml epochs=180
imgsz=640,640 batch=32 --device=0
```

After training the model, the learning progress graphs of Yolo v10 are shown in Figure 5 and Figure 6. The training time required to achieve similar performance generally varied with hyper parameters such as batch size, epoch, and optimize. The model has approximately 5.5ms inference time after 0.3ms preprocessing.

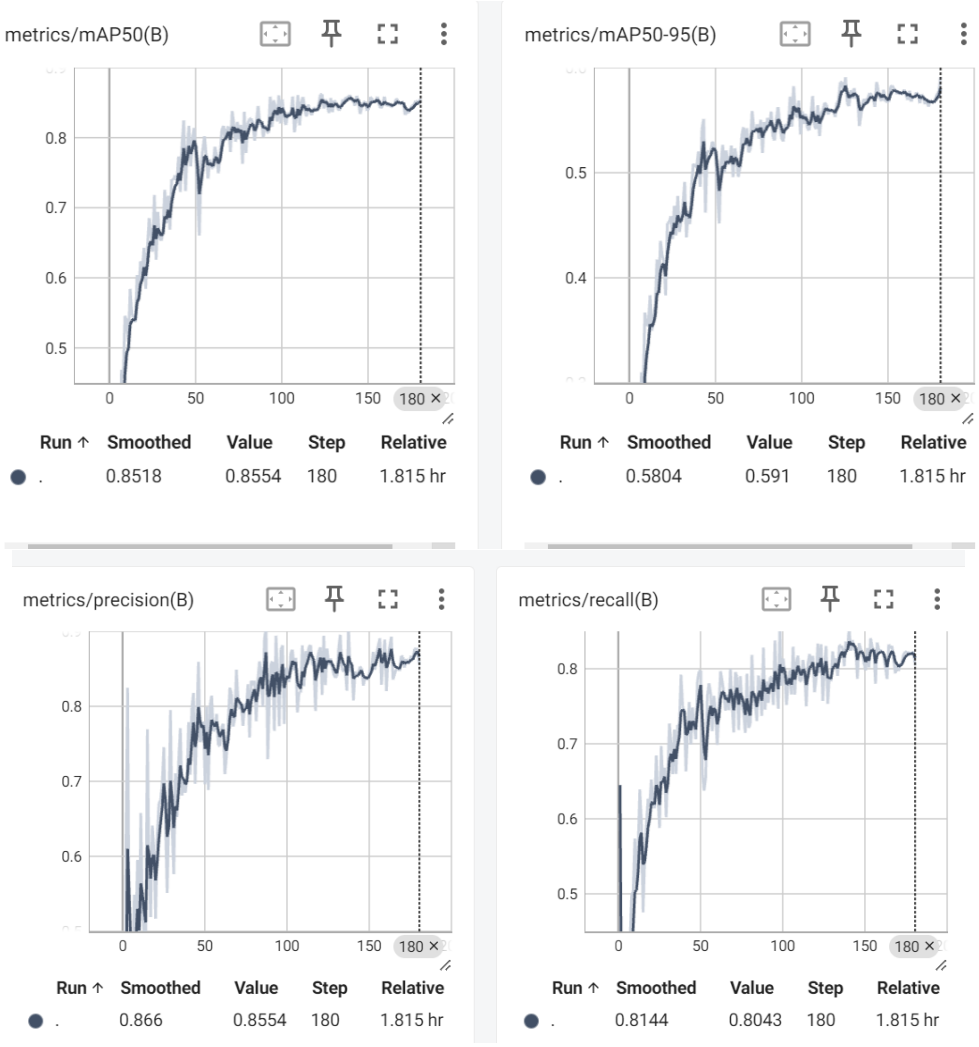


**Figure 5. Example of Training Results**

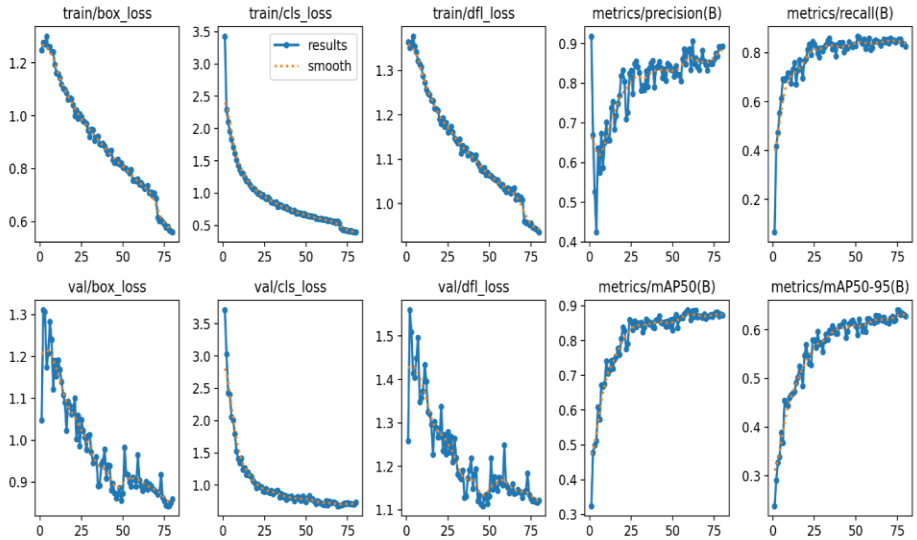


**Figure. 6. Learning Progress Graphs of Yolo v10**

In Figure 6, *train/box_om* means Training Box Omissions, and *train/dfl_om* means Training Distribution Focal Loss Omissions. The *train/box_oo* means Training Box Overlaps, and *metrics/precision(B)* means Precision of Bounding Boxes. The *metrics/mAP50(B)* means Mean Average Precision at IoU 50, and *val/box_om* means Validation Box Omissions.

The experimental results showed a performance with a precision of 0.86, mAP50 of 0.85 and a processing time of 5.8ms. This is probably feasible in real time with low-end portable devices. Figure 7 shows detection examples.

There are several effective solutions to improve this model. One is to prepare a high quality training dataset with a wider variety of objects. Another is to use data augmentation techniques to improve the generalization performance of the model. In addition, lightweighting the model will enable it to perform well even on low-end hardware. This will help to make detection more widespread and feasible, contributing greatly to safe driving. Improved detection can be achieved by using 3D computer vision or depth estimation technology to estimate the depth of objects.

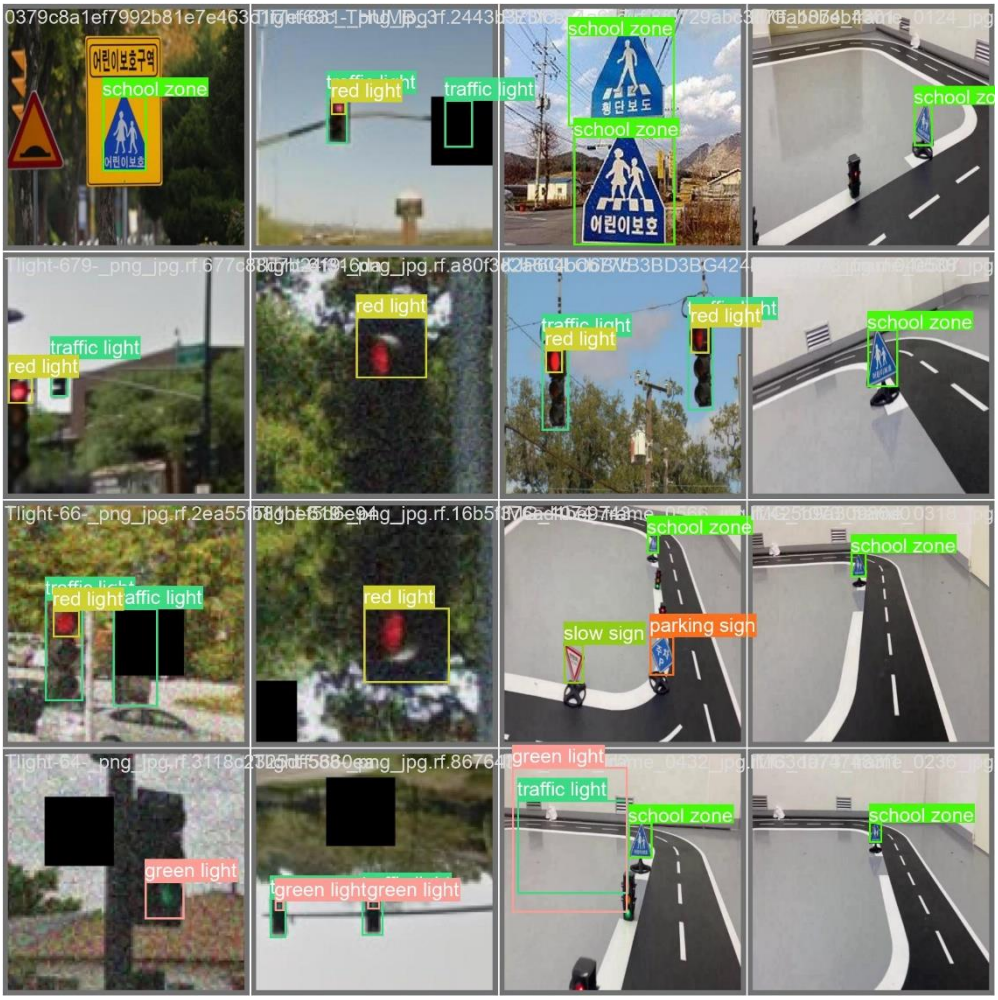


**Figure 7. Example of Detection Results**

## V. CONCLUSION

In this study, we analyzed various deep learning models to improve the performance of object detection to prevent various dangerous accidents and events in front of schools. A YOLOv10-based detection model was proposed, and its effectiveness was verified using real-world big data. The used dataset is categorized into 9 types of classes, and 1213 images were trained.

The experimental results showed a performance with precision of 0.86, mAP50 of 0.85, and processing time of 5.8ms with only 2.7 million parameters. This is probably feasible in real time with low-cost portable devices. This

model will help to make child safety systems more widespread and feasible, which will greatly contribute to safe driving in school zones.

As future work, we would like to improve this project. We would like to prepare a high quality training dataset with more diverse objects and use data augmentation techniques to improve the generalization performance of the mode, and investigate lightweighting the model so that it can perform well on lower-end hardware.

## ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2022R1F1A1063659)

## REFERENCES

[1] Sang Joon Park, Jong Chan Lee, Dea-Sic Jang and Gi Sung Lee, "A study and design of monitor ing module for schoolzone safety," Journal of the Korea Academia-Industrial cooperation Society, vol. 12, no. 4, pp.1940-1946, 2011, DOI http://dx.doi.org/10.5762/ KAIS. 2011.12.4.1940

[2] Kwan-Joong Kim, "A Scheme of Database Design and Management in School-zone System," Journal of The Korea Society of Computer and Information, vol.18, no.5, May 2013 , http://dx.doi.org/10.9708/ jksci. 2013.18.5.061

[3] Moon-Soo Park, Dea-Woo Park, "The Improvement of the LIDAR System of the School Zone Applying Artificial Intelligence," Journal of the Korea Institute of Information and Communication Engineering, vol.26, no.8, pp. 1248-1254, Aug. 2022.

[4] Choong-Yuen Cho, Hong-Kyu Yim, Min-Jae Lee, "Development of ICT-based road safety integrated facilities for pedestrian crossing," Journal of the Korea Academia-Industrial cooperation Society, vol. 18, no. 12, pp. 93-99, 2017, https://doi.org/10.5762/KAIS. 2017.18.12.93

[5] Duo Zhao, Jin-Ju Lee, Hong Kwan Seon. "A Study on Ways to Improve Child-friendly Environmental Design through Analysis of School Zone Cases in Shenzhen, China," vol.24, no.4, pp. 186-197, 2024, doi: https://doi.org/10.5392/JKCA.2024.24.04.186

[6] Kim, Ah-Ram, Kim, Donghyeon, Byun, Yo-Seph, Lee, Seong-Won, "Object Detection of Concrete Structure Using Deep Learning and Image Processing Method in Geotechnical Engineering," Journal Of The Korean Geotechnical Society, vol.34, no.12, pp. 145-154, December 2018.

[7] Kim, Y., "Development of Object Recognition System for Concrete Structure Using Image Processing Method," J. of Korean Institute of Information Technology, vol.14, no.10, pp.163-168, 2016.

[8] Park, H.S., "Performance Analysis of the Tunnel Inspection System Using High Speed Camera," J. of Korean Institute of Information Technology, vol.11, no.4, pp.1-6, 2013.

[9] Jongwoo Ha, Kyongwon Park, Minsoo Kim, "A Development of Road Object Detection System Using Deep Learning-based Segmentation and Object Detection," The Journal of Society for e-Business Studies, vol.26, no.1, pp.93-106, February 2021, https://doi.org/10.7838/jsebs.2021.26.1.0

[10] M. Kasian and H. Kilavo., "A comparative study on performance of SVM and CNN in Tan-zania sign language translation using image recognition", Applied Artificial Intelligence, vol.36, no.1, pp.452-466, Nov. 2021, DOI: 10.1080/08839514.2021.2005297.

[11] S. Byun, "Composite Context-based Offloading Scheme for IoT Fault Diagnosis," Journal of Next-generation Convergence Technology Association, vol.7, no.5, pp.762-773, May 2023, DOI: 10.33097/JNCTA.2023.07.05.762.

[12] S. D. Nguyen, T. S. Tran, V.P. Tran, H. J. Lee, Md. J. Piran, and V. P. Le., "Deep Learning-Based Object Detection: A Survey," International Journal of Pavement Research and Technology, vol. 16, pp.943-967, Apr. 2022, DOI:10.1007/s42947-022-00172-z.

[13] Cho, S., Kim, B., and Lee, Y.I., "Image-Based Concrete Object and Spalling Detection using Deep Learning," J. of the Korean Society of Civil Engineers, vol.66, no.8, pp.92-97, 2018.

[14] Roboflow,https://universe.roboflow.com/object-detection-lmq58/infrared-model.

[15] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Journal of Machines and Tooling, vol.11, no.7, pp.677:1-25, Jun. 2023, DOI: 10.3390/machines11070677.

[16] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, Guiguang Ding, "YOLOv10: Real-Time End-to-End Object Detection". Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR 2024).