# Convolutional Neural Network-based Jaywalking Data Generation and Classification

Jaeseo Park, Yunsoo Lee, Jun Ho Heo, and Suk-Ju Kang
Department of Electronic Engineering
Sogang University
Seoul, Republic of Korea
geeryu12@gmail.com, profitshore@gmail.com, jayceheo92@gmail.com, sjkang@sogang.ac.kr

*Abstract*— **In this paper, we propose a novel system to generate jaywalking images. To synthesize a pedestrian on the road and label the binary case such as jaywalk or normal-walk, the pre-trained Convolutional Neural Network (CNN) is used to segment the drivable area from the large-scale dataset. The proposed system automatically generates a jaywalker based on existing pedestrian objects in the image. The proposed system performs three main steps. First, we train the existing network with both black box image dataset and object dataset to segment road areas and pedestrians. Second, the generator synthesizes jaywalkers randomly within the road segmentation masks. Third, a CNN classifier is trained using the generated synthetic dataset and performs the inference from natural jaywalking images. The experiment results show that the jaywalking classifier trained with both generated synthetic dataset and the untouched natural dataset has a high accuracy of 0.96, which is 0.08 higher than the accuracy using only the untouched natural dataset on the same model.**

***Keywords; Convolutional Neural Network; Deep Learning***

## I. INTRODUCTION

Recently, intelligent systems that collect and analyze road traffic safety information have been widely studied. The data is automatically collected using closed-circuit televisions (CCTV). Due to the quantitative and qualitative shortage of pre-existing datasets, there are difficulties in developing the study further.

The commonly used road view datasets, UCSD Ped1 and Ped2 [1], consist of 18,560 grayscale images with low resolutionAnother dataset, Shanghai Tech Campus [2], consists of extracted images from only 12 videos, and therefore, lacks data diversity. Further, and importantly, the dataset introduced above does not include jaywalker data. Therefore, we need a new system that can efficiently generate the jaywalking dataset using the dataset in a richer field. In this paper, we propose a method of generating the jaywalking images dataset using a vehicle black box dataset.

## II. PROPOSED SYSTEM

The proposed generator system consists of three stages. First, the drivable area is detected in the base road image by Mask R-CNN [4] learned from the black box dataset. This detector is called a drivable area detector. Second, Mask R-CNN pre-trained with MS-COCO2014 [5] detects and segments person objects. Third, the segmented mask of a person is copied, transformed, and pasted into the inner area of the mask.

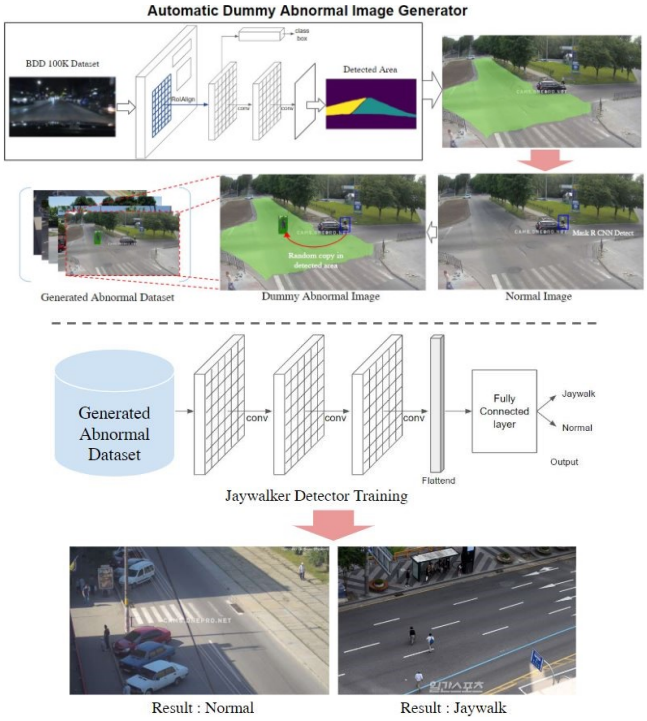


Figure 1. The overall architecture of a deep learning-based Jaywalker Image generator for traffic and pedestrian motion analysis (Top) Architecture of Automatic dummy jaywalker image generator (Bottom) Experiment jaywalk classifier by generated Jaywalker Dataset

The overall architecture is shown in Fig. 1, (Top) represents a dataset generator, and the (Bottom) represents a structure evaluated by the untouched natural dataset.

### A. Dataset Generation

From the Mask R-CNN trained with the black box dataset [3], we only segment the mask of drivable areas called $D$. Vehicles and peoples are excluded from $D$. In addition, we get a mask of person $O_k$ through Mask R-CNN model trained with MS-COCO dataset. With these masks, we generate synthetic images by combining the mask of person and the drivable area mask . The location of the mask of the person is selected randomly.

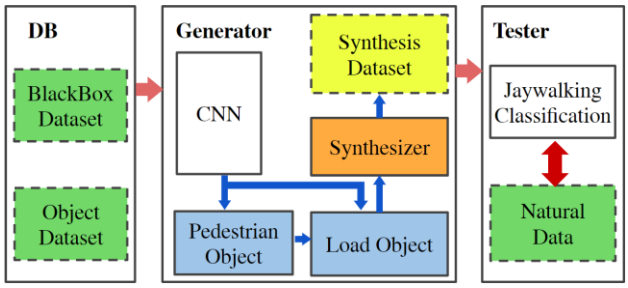


Figure 2. The overall architecture of proposed system

In this case, we should consider the case when multiple person objects are in a one image. To clarify this, we set up a process to select objects to copy. $O_k$ to be copied should not be located in drivable area $D$, and $O_k$ with the lowest $k$ is selected. (1) illustrates this process. Then, we randomly flip the selected mask to increase the amount of data.

$$\text{If } O_k\ (|k| > 1) : \min k\ (O_k \not\subset D) \tag{1}$$

### B. Jaywalk Classifier

Because there is no public dataset of jaywalking, the purpose of the generated synthetic dataset is to make the jaywalker classifier work reasonably on the natural input images. This overall system flow is shown in Fig .2. The classifier, GoogLeNet [6] is trained with binary labeled(jaywalk/normal-walk) images. This classifier contains the architecture model trained on ImageNet [7] data. And we retrained this classifier with a new top layer that can recognize jaywalk classes of images. Finally, this classifier determines whether each test image contains a jaywalking event.

## III. Experimental Results

### A. Dataset

The dataset for the jaywalker image generator consists of 70,000 imges, some of both the BDD 100K dataset and the MS-COCO dataset. Both datasets were fed into Mask R-CNN.

Figure 3. (Left) Drivable Area Detection (Right) Generated Image

A total of 500 jaywalking images were generated through the proposed method. And the number of normal-walk images is a total of 400, half of them are original images used in the generator and the rest are new ones. The normal-walk dataset consists of 200 original images used in the generator and 200 street images of different views not used in the generator. We collected 337 images of natural jaywalking through a Google image crawler or capturing CCTV images followed by proper selection step. These natural jaywalking data were used as the test set.

### B. Experiment Configuration

The experiment was conducted using TensorFlow library, on Intel ® Xeon ® CPU E5-2690 v4 @ 2.60GHz, TITAN Xp. As shown in Fig. 3 (Left), the vehicle driving area is segmented using the drivable area detector. The actual synthesized jaywalker image is shown in Fig. 3 (right) with the original segmented mask of a person. The proposed method was evaluated with the real images collected in advance.

TABLE I. Comparison Trained Model by Generated Data

| | Confusion Matrix | | | |
|---|---|---|---|---|
| | ***TP*** | ***TN*** | ***FN*** | ***ACC*** |
| Natural | **328** | 320 | 81 | 0.878 |
| Natural+Ours | 327 | **387** | **14** | **0.967** |

The evaluation method is described as follows.

$$ACC(Accuracy) = \frac{(True\ Positive + True\ Negative)}{Positive + Negative} \tag{2}$$

In order to compare the accuracy, we obtained 2 results, one is from a model trained with natural data only and the other is a model from both natural and synthetic datasets. Table. 1 shows the different results of them. The True Positive(TP) scores of jaywalking detection have little gap while normal-walk is 5.7 times higher for the natural data only at the False Negative(FN) scores.

Based on these scores, the calculated accuracy of using both datasets is 8% higher than the one from using the natural dataset only. As a result, the generated synthetic dataset can improve the performance of the jaywalking classifier.

## IV. Conclusion

In this paper, we proposed a novel system that automatically generates jaywalking scenes. Also, a jaywalking classifier trained with generated data is suggested and evaluated. And the better results are shown when using both the untouched dataset and the generated synthetic dataset, having an 8% improved accuracy.

## Acknowledgment

This research was supported by a grant(19PQWO-B153369-01) from Smart road lighting platform development and empirical study on test-bed Program funded by Ministry of the Interior and Safety of Korean government, and the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-0-01421) supervised by the IITP(Institute for Information & communications Technology Promotion.

## References

[1] http://www.svcl.ucsd.edu/projects/anomaly/

[2] Liu, Wen, et al. "Future frame prediction for anomaly detection–a new baseline." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[3] Yu, Fisher, et al. "BDD100K: A diverse driving video database with scalable annotation tooling." arXiv preprint arXiv:1805.04687 (2018).

[4] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[5] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.

[6] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[7] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.