Our research paper titled “Road Obstacle Detection based on Unknown Objectness Scores” has accepted for presentation at the International Conference on Robotics and Automation 2024 (ICRA2024)

Toyota Motor Corporation
Social System PF Development Div.
InfoTech-AS Data Analysis Team
Chihiro Noguchi
Toshiaki Ohgushi
Masao Yamanaka

Toyota Motor Corporation has developed AI technology that detects road obstacles from images. Our research paper, " Road Obstacle Detection based on Unknown Objectness Scores " was accepted at the ICRA2024 (International Conference on Robotics and Automation 2024. We will present our work at the main conference from May 14 to 16, 2024.

Backgrounds

To ensure a safe driving environment, technology that automatically detects obstacles on the road that could impede traffic is extremely important. In the context of autonomous driving, it is an indispensable technology for ensuring a minimum level of safety, but even in situations where humans are driving, detecting road obstacles can lead to the realization of various services through the sharing of obstacle information with other vehicles and providing information to road administrator. While autonomous vehicles often come equipped with expensive sensors and can detect obstacles with high accuracy by combining camera images with sensor data, in the case of ordinary vehicles, it is necessary to detect obstacles based solely on camera images, such as those from dashcams.

The biggest challenge in detecting road obstacles from images is the variety of potential obstacles. Various objects, including debris from vehicles, animal carcasses, and rockfalls, can appear on the road. It is not easy to train detection models to cover these diverse detection targets comprehensively. Additionally, encountering situations where obstacles are present on the road is extremely rare, making it difficult to collect sufficient training data. Therefore, many approaches, including the proposed method, introduce the concept of anomaly detection. By detecting unusual events that would not occur in normal driving environments, these approaches lead to detect road obstacles.

In detecting road obstacles, the most straightforward approach is considered to be the application of object detection methods. However, as mentioned above, the conventional supervised learning approach has the problem that it is difficult to cover a diverse range of detection targets. Additionally, there are barriers to introducing the concept of anomaly detection. Normally, the training datasets for object detection define labels for objects to be detected in advance and annotate objects in the images corresponding to each label. However, objects that are not predefined are naturally not annotated, so during model training, these are learned as background rather than objects. As a result, many road obstacles are learned as part of the background, making it difficult to separate the obstacles from the background area. Against this background, many existing methods are built on semantic segmentation techniques rather than object detection. By limiting the application to onboard camera footage, it is possible to assign predefined labels to almost all pixels in the image. Therefore, appropriate background labels (such as sky, road, etc.) are assigned to areas that correspond to the background, preventing areas with objects not predefined as labels from being learned as background. Furthermore, since the elements that can appear in images of vehicle driving scenes are limited, it is possible to cover most areas in the image with a relatively small number of labels. Thus, areas that do not fit any of the predefined labels can be regarded as containing anomalies.

However, in unsupervised learning that does not explicitly utilize label data of the detection targets as mentioned above, it is difficult to exhibit stable performance, especially in images containing complex structures like those from onboard cameras, presenting challenges from a practicality standpoint. Therefore, we have developed an obstacle detection method that can achieve more practical detection performance by introducing the concept of supervised learning-based object detection methods into anomaly detection techniques.

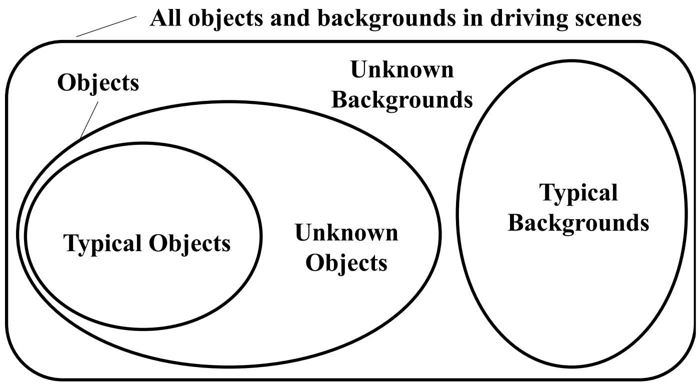


Figure 1　Venn diagram that illustrates unknown objectness scores (Eq. 1).

Technical Overview

In many object detection methods, a measure called objectness score, which represents the likelihood of being an object, is utilized. In deep learning-based methods, the objectness score is also a target of learning, aiming to acquire object-likeness that generalizes across a broader range of objects. In road obstacle detection, the detection targets are limited to objects on the road, so it is believed that detection performance can be improved by effectively utilizing the objectness score. Figure 1 shows a Venn diagram representing the basic concept of the proposed method. As mentioned before, when the application is limited to onboard camera footage, the elements that can appear in the image are limited and can be broadly categorized into two groups: objects (such as vehicles, people, benches) and background (roads, sky, sidewalks). Additionally, by defining elements that match predefined labels as typical elements and those that do not as unknown elements, objects included in labels such as vehicles and people can be considered typical objects, while road obstacles not included in the labels are considered unknown objects. Typical elements can be trained with defined labels for supervised learning, and objectness scores can also be learned by following the object detection methodology. Therefore, unknown objects can be detected as "neither backgrounds nor typical objects, but are object" (the blue area in Figure 1). Specifically, the unknown objectness score is defined as follows.

$$S_i = p_i^O \prod_{k=1}^{K} (1 - p_{ik}) \tag{1}$$

Here, S_i represents the unknown objectness score for pixel i. p_i^O and p_{ik} represent the objectness score and the predicted probability for label k corresponding to pixel i, respectively. K is the number of labels predefined. Areas with high unknown objectness scores are detected as regions containing unknown objects.

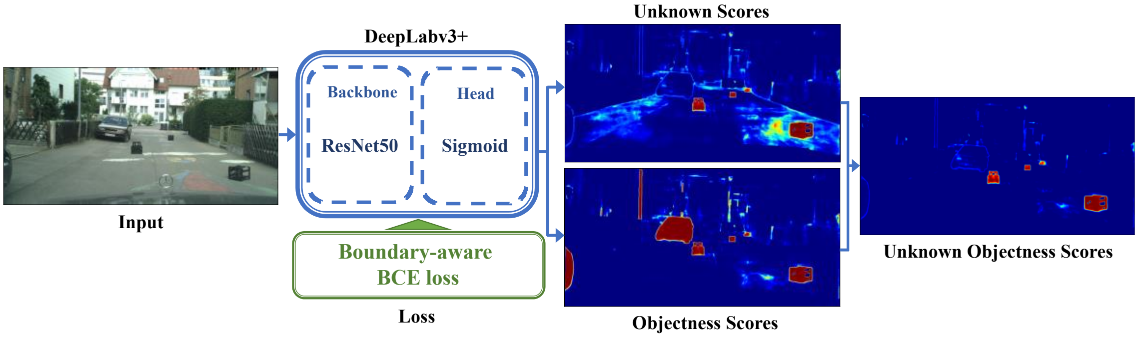


Figure 2　Overview of the proposed method.

Figure 2 shows the overview of the proposed method. The given image is input into a standard neural network for semantic segmentation. The proposed method uses, in part, a different configuration than the standard network, with a sigmoid function at the head. The sigmoid head is a configuration used for multi-label classification and can naturally represent states that do not correspond to any of the predefined labels. When attempting to do the same thing with the softmax function, which is standardly adopted, it is necessary to train in a way that increases the entropy of the prediction probability of pixels corresponding to unknown elements. However, this significantly reduces the performance of semantic segmentation. Additionally, adopting the sigmoid head has the advantage of being able to learn predefined labels and objectness scores simultaneously.

In Fig. 2 on the right, the results of visualizing the objectness score and the unknown score are shown. The unknown score represents the factors excluding the objectness score in Eq. 1. It is observed that using the unknown score alone results in many false positives primarily in the background regions. However, when the unknown score is multiplied by the objectness score to calculate the unknown objectness score, most of these are suppressed.

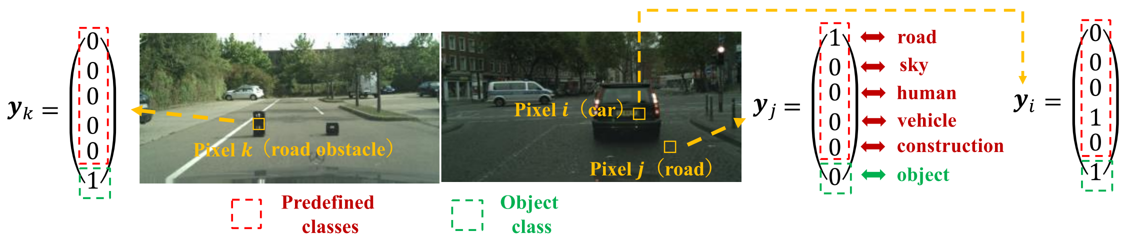
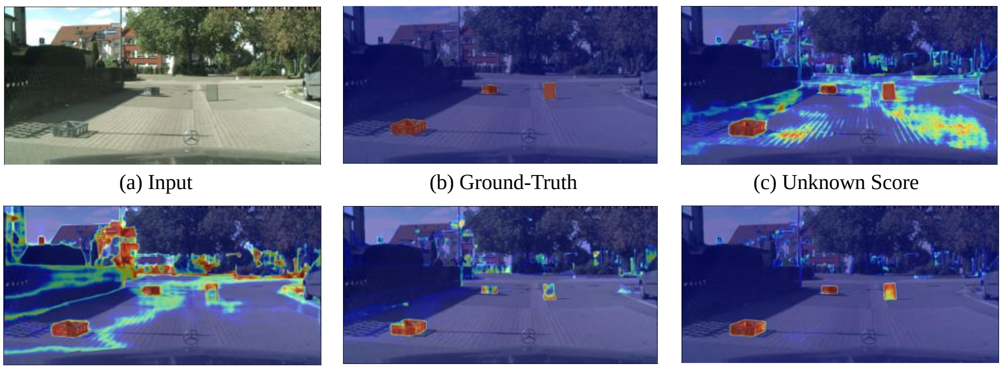


Figure 3　Label assignment in the proposed method.

As mentioned at the beginning, the frequency of encountering road obstacles during normal driving is extremely low, making it difficult to collect a sufficient number of datasets as training data. However, it is possible to obtain a small amount of data, and it is practically important that this small amount of training data can be used to improve detection performance. The proposed method allows for the use of these small amounts of training data to enhance the detection performance of objectness scores. In the image on the left of Fig. 3, there are two road obstacles. The labels assigned to these pixels are defined as follows: all predefined labels are 0, and labels corresponding to the objectness score are set to 1. This allows for the effective use of a small amount of training data to improve detection performance without the need to create a new class every time a new obstacle is discovered.

$$\mathcal{L}_n = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k\in C} f(y_{ik}, p_{ik}) - \frac{\lambda}{\sum_i \delta_i}\sum_{i=i}^{N}\delta_i \sum_{k\in C} f(y_{ik}, p_{ik}) \qquad (2)$$

Equation 2 presents the cost function for optimizing the proposed model. $\sum_{k\in C} f(y_{ik}, p_{ik})$ represents the binary cross-entropy between the label y_{ik} and the predicted probability p_{ik} at pixel i. Moreover, C denotes the set of labels combining predefined labels and object classes. However, optimization using standard binary cross-entropy often results in a number of false positives in the boundary regions between different labels. This is because the network struggles to decide between at least two labels in boundary areas, leading to an increase in unknown scores in these regions. Therefore, the proposed method introduces an additional penalty (the second term on the right-hand side) for boundary regions, ensuring that the network learns to output predicted probabilities with low entropy even in boundary areas.



(a) Input (b) Ground-Truth (c) Unknown Score

(d) Outlier Exposure [39] (e) SynBoost [6] (f) Unknown Objectness Score (ours)

Figure 4 Qualitative comparison with existing methods.

Figure 4 shows the results of a qualitative evaluation comparing the proposed method with existing methods. It can be seen that the three objects located near the center of the image are relatively well detected by the proposed method (f). Compared to existing methods, it is confirmed that the detection of false positives is significantly reduced, mainly in the background area.

Summary and Future Works

In this study, we proposed a new method for detecting road obstacles by incorporating two concepts: object detection and anomaly detection. By utilizing objectness scores that can be learned through supervised learning, we successfully significantly improved detection performance, especially in reducing the detection of false positives in background areas. However, the introduction of objectness scores means that if the proposed model fails to predict high objectness scores for road obstacles, it could lead to decreased detection performance. Therefore, learning more generalized objectness scores is considered key to further improving detection performance.