

Road Obstacle Detection based on Unknown Objectness Scores

Chihiro Noguchi, Toshiaki Ohgushi, and Masao Yamanaka

Abstract—The detection of unknown traffic obstacles is vital to ensure safe autonomous driving. The standard object-detection methods cannot identify unknown objects that are not included under predefined categories. This is because object-detection methods are trained to assign a background label to pixels corresponding to the presence of unknown objects. To address this problem, the pixel-wise anomaly-detection approach has attracted increased research attention. Anomaly-detection techniques, such as uncertainty estimation and perceptual difference from reconstructed images, make it possible to identify pixels of unknown objects as out-of-distribution (OoD) samples. However, when applied to images with many unknowns and complex components, such as driving scenes, these methods often exhibit unstable performance. The purpose of this study is to achieve stable performance for detecting unknown objects by incorporating the object-detection fashions into the pixel-wise anomaly detection methods. To achieve this goal, we adopt a semantic-segmentation network with a sigmoid head that simultaneously provides pixel-wise anomaly scores and objectness scores. Our experimental results show that the objectness scores play an important role in improving the detection performance. Based on these results, we propose a novel anomaly score by integrating these two scores, which we term as unknown objectness score. Quantitative evaluations show that the proposed method outperforms state-of-the-art methods when applied to the publicly available datasets.

I. INTRODUCTION

When driving, various types of obstacles can be encountered on the road, and detecting them from in-vehicle camera images is essential for safe autonomous driving. However, standard object-detection methods [1], [2], [3] fail to detect unknown objects that do not belong to any predefined object categories. It is also difficult to address this issue in an anomaly detection setting because unknown objects are learned to be identified as part of the image background. Since many road obstacles cannot be classified as the predefined object categories, they are basically identified as part of the background.

Against this background, recently proposed methods [4], [5], [6] are based on semantic-segmentation networks. Semantic segmentation [7], [8] aims to assign an appropriate label to almost every pixel, including the image background. This makes it easier to consider anomaly detection settings for detecting unknown regions because both predefined object and background pixels are included in the in-distribution; unknown regions can be detected as out-of-distribution (OoD) samples. This approach is known as pixel-wise anomaly detection [6]. While the pixel-wise anomaly-detection methods are effective for detecting unknown re-

Toyota Motor Corporation, Otemachi, Chiyoda-ku, Tokyo, 100-0004, Japan. Corresponding author is Chihiro Noguchi (chihiro.noguchi.aa@mail.toyota.co.jp)

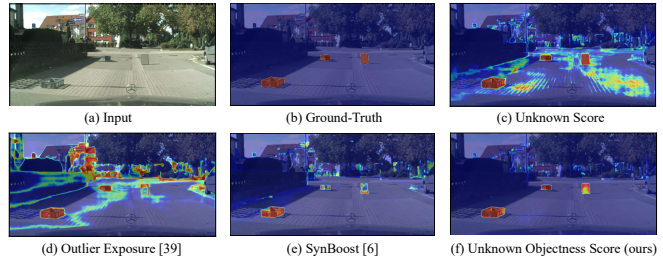


Fig. 1. Examples of road obstacle detection results using the proposed and representative methods. The proposed anomaly score—unknown objectness score—reduces the number of false-positive predictions, especially in the background regions, compared to the baseline scores (c) and the results of the existing methods (d,e).

gions in an image, their performance is often unstable when applied to road obstacle detection. These methods do not limit detection to unknown objects, but can detect any unknown regions in the image. In the context of the road obstacle detection, however, this also means an increase in false positive predictions, especially in the image background.

In this paper, we propose a novel road obstacle detection method that combines the principles of pixel-wise anomaly detection and object detection. Representative object-detection methods, such as Faster-RCNN [2] and YOLOv3 [1], calculate *objectness scores* that measure how much the corresponding local region in an image is object-like. The use of objectness scores is promising for stabilizing the performance of detecting unknown objects, as is the case with detecting predefined objects. However, in fact, these objectness scores are not compatible with the anomaly-detection settings, as mentioned earlier. Therefore, we aim to address this issue by integrating the objectness scores with pixel-wise anomaly scores defined from predicted probabilities in semantic segmentation.

The proposed method is based on a semantic-segmentation network utilizing a sigmoid head. This approach allows for the assignment of multiple classes to each pixel, enabling a more flexible design of class labels compared to a typical softmax head. Specifically, we merge predefined object classes, such as cars and traffic signals, into an *object class*, and define the objectness score as the predicted probability of the object class. The use of the sigmoid head allows the segmentation model to simultaneously learn the labels of both the object class and the predefined classes. Additionally, we define *unknown scores* as pixel-wise anomaly scores obtained from the outputs of the sigmoid head. By combining the objectness and unknown scores, we define *unknown objectness scores*.

This design is particularly effective when only a limited number of images of road obstacles are available. In order to use these images for training our model, we assign the object class label to pixels corresponding to road obstacles. This approach eliminates the need to create new classes for road obstacles and also precludes the necessity to limit the detection targets to specific objects. This situation often arises in practical settings since encountering road obstacles in standard driving scenarios is infrequent. We experimentally show the effectiveness of the unknown objectness scores using the publicly available datasets (Fig. 1).

The major contributions of this study are as follows.

- A novel road obstacle detection method is proposed based on a novel anomaly score—unknown objectness score—defined by the outputs obtained from the sigmoid head.
- The proposed method exhibits better performance than state-of-the-art approaches, when applied to the publicly available datasets.
- The proposed method is based exclusively on a semantic-segmentation network; this facilitates fast detection without the need for additional modules.

II. RELATED WORK

A. Pixel-wise Anomaly Detection

Numerous pixel-wise anomaly detection methods have been proposed. These methods can be primarily classified into two approaches based on the concepts of autoencoder and uncertainty estimation.

1) *Autoencoder-based Approaches:* Autoencoder-based methods require the use of at least two modules—an encoder and a decoder. The encoder maps input images into a feature space and the decoder subsequently attempts to reconstruct the original image. If the test images contain an unknown region, the encoder and decoder cannot perform an accurate reconstruction. This facilitates the identification of the unknown region. The encoder and decoder modules are, in general, configured using networks such as the generative adversarial network (GAN) [9] and variational autoencoder (VAE) [10]. Previous studies have proposed several methods based on different autoencoder modules, including the RBM autoencoder [11], VAE [12], and GAN [13], [14], [15], [16].

Recent studies [17], [18], [5], [6] have proposed methods based on the adoption of a segmentation network as an encoder. The decoder in these methods attempts to reconstruct an input image from a predicted semantic label map [19], [20]. A segmentation network can fail to assign an accurate label map to an unknown region. This causes the subsequent decoder to generate a reconstructed image that features a large gap or an indeterminate spot in the unknown region. To quantify the gap between the input and reconstructed images, these methods calculate the perceptual difference [19]. The perceptual difference is obtained by calculating the difference between both feature maps from the middle layers of the ImageNet [21] pretrained VGG [22]. Furthermore, some existing studies [17], [6] introduced a discrepancy network and trained it to enhance this difference.

2) *Uncertainty-based Approaches:* Utilizing uncertainty estimation is another standard approach for performing pixel-wise anomaly detection. The predicted semantic labels in an unknown region can be expected to demonstrate high uncertainty. This is because the semantic-segmentation model is trained using dataset that does not contain the ground truth of such regions.

Bayesian neural networks [23], [24] based on Bernoulli’s approximation of the variational distribution are widely used for uncertainty estimation [25], [26]. This approach, which is closely related to the dropout technique [27], is referred to as the MC dropout. Although the MC dropout is applicable to pixel-wise uncertainty estimation [28], [29], [30], [31], [32], it encounters problems when applied to unknown object detection in images representing driving scenarios. Specifically, the high-uncertainty pixels estimated by the MC dropout often do not match the positions of unknown objects, especially in complex situations such as driving [17], [5]. Another uncertainty-estimation approach involves the simple utilization of the softmax entropy [33]. The softmax entropy can be easily calculated from the predicted probabilities generated from a segmentation network. To facilitate unknown object detection in driving scenarios, the use of the maximum logit instead of softmax entropy has been found to improve detection performance [34]. In addition, a recent study [4] has revealed that standardizing the maximum logit can further improve performance.

III. PROPOSED METHOD

Our proposed method builds on a semantic-segmentation network with a sigmoid head. This method leverages the properties of the sigmoid output to define the unknown objectness scores. Fig. 2 presents an overview of the proposed method. An input image is first fed to the segmentation network, the output of which provides unknown scores and objectness scores (Sec. III-A). Then, the unknown objectness scores are defined by combining these two scores (Sec. III-B). In addition, we introduce a loss function to reduce false positives in boundary regions (Sec. III-C).

A. Unknown Scores

State-of-the-art methods for performing pixel-wise anomaly detection in driving scenarios are often developed considering a semantic-segmentation network. In these networks, the labeled data facilitate a pixel-level semantic understanding, which is essential for detecting unknown object pixels. The proposed method is likewise built on a semantic-segmentation network. However, unlike most methods that adopt a softmax head in the output of the final layer, the proposed method adopts a sigmoid head.

The predicted probabilities obtained using the sigmoid head can be considered as a result of K binary classifiers, where K denotes the number of classes. In this setting, an unknown pixel can be detected as a pixel with low predicted probabilities for all predefined classes. Semantic-segmentation labels are usually assigned to nearly every

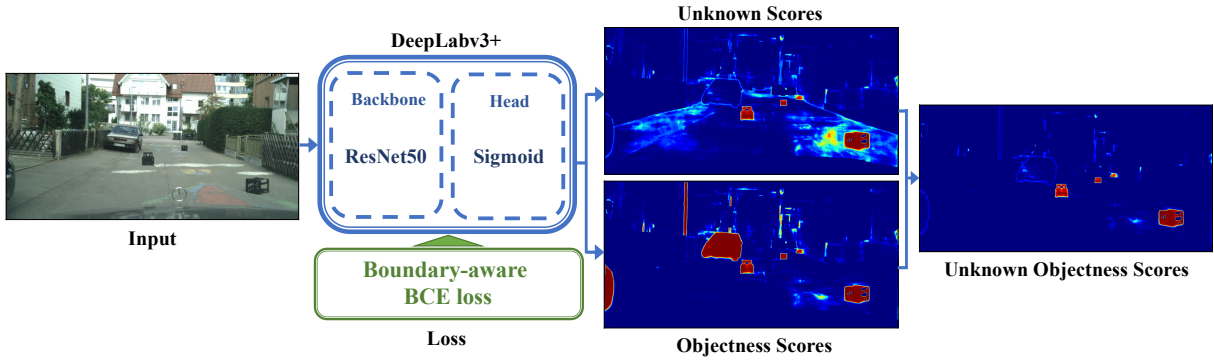


Fig. 2. Overview of the proposed method. The input image is fed to a semantic-segmentation network with a sigmoid head. Unknown objectness scores are obtained from the sigmoid outputs by combining unknown and objectness scores.

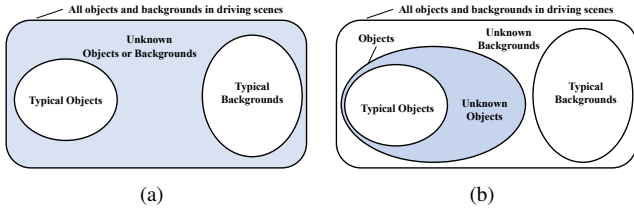


Fig. 3. Venn diagrams comparing between the unknown score (Eq. 1) and the unknown objectness score (Eq. 2). Typical objects and backgrounds indicate predefined semantic-segmentation classes, while unknown objects and backgrounds indicate ones that are not fall into the predefined classes. Road obstacles are classified into unknown objects.

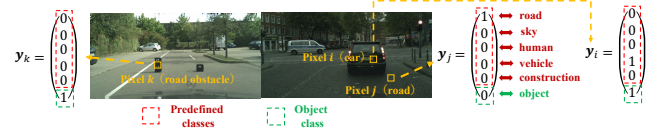


Fig. 4. Examples of labels for three typical types of pixels. This figure shows three pixels i, j, k corresponding to a car, a road and a road obstacle pixels, respectively. Pixel i is assigned to both the vehicle and object classes, while pixel j is assigned to the road class only, because cars are objects, and roads are not. Pixel k is assigned to the object class only, because road obstacles do not fall into any of the predefined classes, while they are objects.

pixel, thereby suggesting that the K classes cover all foreground objects and backgrounds that could typically appear in images captured while driving. Therefore, a pixel with low predicted probabilities for all predefined classes is atypical; that is, it represents an unknown pixel. This is illustrated in the Venn diagram in Fig. 3(a). Given the predicted probabilities of pixel i as p_{ik} ($k \in \{1, \dots, K\}$), the corresponding unknown score can be defined as

$$S_i = \prod_{k=1}^K (1 - p_{ik}). \quad (1)$$

A well-trained segmentation model can assign a high probability, that nearly equals one, to pixels corresponding to a predefined class, while the corresponding unknown score nearly equals zero. On the other hand, the predicted pixel probabilities in an unknown region remain small with correspondingly large unknown scores.

B. Unknown Objectness Scores

The unknown score defined in Eq. 1 can be used to detect pixels in unknown regions, not just unknown object pixels. However, as with other pixel-wise anomaly-detection methods, this generic unknown score tends to be unstable when applied to road obstacle detection. Therefore, in this subsection, we extend the unknown score by incorporating the object detection fashions. Because road obstacles are naturally classified as objects, this strategy can be expected to be effective to achieve stable performance. We first define the

pixel-wise objectness score p_i^o , which measures how much pixel i is object-like. Pixels of road obstacles are expected to possess high objectness scores.

To define p_i^o , we make an *object class* by merging the foreground classes among classes predefined in semantic segmentation. p_i^o is defined by the predicted probability of the object class. By adopting the sigmoid head, it is possible to assign labels of both a predefined class and the object class to a single pixel at the same time. Figure 4 illustrates how to make such labels more concretely. Considering the objectness score p_i^o , the unknown score defined in Eq. 1 can be extended as follows.

$$S_i = p_i^o \prod_{k=1}^K (1 - p_{ik}). \quad (2)$$

This score is referred to as the unknown objectness score. As represented in Eq. 2 and the Venn diagram in Fig. 3(b), unknown objects can be detected as labels that are *neither backgrounds nor typical objects, but are objects*. As a result, we can focus on the more promising regions in an image where road obstacles could be located. This facilitates a reduction in the number of false-positive predictions.

In addition, the proposed method can use OoD data optionally for supervision. Specifically, if we obtain images that contain road obstacles and corresponding labels, we can use them for supervision by assigning the only object class to pixels corresponding to the road obstacles. Figure 4 shows label y_k of pixel k corresponding to a road obstacle

pixel. y_k has all 0 for elements of the predefined classes and 1 only for the element of object class. In practical scenarios, it is common to have a limited number of road obstacle images, yet there is a desire to improve detection performance using these images. However, introducing a new class for road obstacles can result in a class imbalance issue. In the proposed method, road obstacle labels are used to emphasize the difference between typical and unknown objects (Fig. 3). This approach obviates the need for a new class and concurrently enhances detection performance.

C. Boundary-aware Binary Cross-entropy Loss

The flexibility of the sigmoid head could lead to an inherent disadvantage when evaluating the boundary regions in an image. Whereas a segmentation classifier may confuse between at least two classes in boundary regions, a sigmoid classifier allows these regions to be assigned to two classes simultaneously. This increases the unknown scores of these regions and results in false-positive predictions. To overcome this disadvantage, we assign additional weights to the binary cross-entropy (BCE) loss incurred in the boundary regions. The loss function \mathcal{L}_n is expressed as

$$\mathcal{L}_n = -\frac{1}{N} \sum_{i=1}^N \sum_{k \in \mathcal{C}} f(y_{ik}, p_{ik}) - \frac{\lambda}{\sum_i \delta_i} \sum_{i=1}^N \delta_i \sum_{k \in \mathcal{C}} f(y_{ik}, p_{ik}), \quad (3)$$

where $f(y, p) = y \log p + (1 - y)(1 - \log p)$. In the above equation, the first term on the right of the equality indicates the standard BCE loss, and the second term indicates the additional boundary-aware loss. N denotes the number of pixels in image n , y_{ik} denotes the label, and \mathcal{C} denotes the set of classes that comprise the K predefined classes and the object class. Moreover, $\delta_i = 1$ when pixel i lies in the boundary regions; in all other cases, $\delta_i = 0$. The term λ controls the weights of the boundary-aware loss.

IV. EXPERIMENTS

A. Experimental Setup

Implementation Details. In this study, the experiments were performed using DeepLabv3+ [8] with the ResNet50 backbone [35]. The segmentation model was trained using the Cityscapes dataset [36]. The optimizer used was SGD with a momentum value of 0.9 and weight decay of 0.0001. The initial learning rate was 0.01, and the “poly” learning rate policy was used with the power set to 0.9. The value of λ was set to three.

Datasets. The performance of the proposed method was evaluated using three datasets comprising driving scene images that contain unknown road obstacles. We first used the LostAndFound dataset [37], which is widely used and contains temporally sequential images with annotations of small unknown objects. The test set used in this study comprises 1,203 images. Next, we used the Fishyscapes Validation set [38], which comprises 100 images. This set is a subset of the LoadAndFound dataset and was created by selecting a suitable set of sequential images capturing a driving scene.

The images in this dataset contain additional annotations for pixels assigned to classes not considered in the Cityscapes dataset. Furthermore, we used the road anomaly dataset [17], which comprises online images depicting a variety of unknown objects, such as animals and rocks. This dataset contains 60 images depicting a variety of driving scenes, which are not limited solely to urban cityscapes. The images in this dataset contain objects of various sizes and roads under various conditions. Identifying unknown objects from this dataset is, therefore, challenging.

OoD Data. The state-of-the-art method proposed in [6] uses the “void class” in the Cityscapes dataset as OoD data for model training. The “void class” is typically used to mask objects not assigned to any predefined class. Therefore, they can be considered pixels of unknown objects. Following this procedure, the “void class” was used as the OoD data in the comparison between methods using OoD data. [39], [6], [40].

Baselines. Existing methods can be classified as those that use OoD data for supervision [39], [6], [40] and those that do not [33], [34], [5], [4]. To evaluate the performance of the methods proposed in [5], [6], [4], we used the source code implemented by the respective authors. The remaining methods were evaluated using the segmentation model and backbone identical to those considered in this study.

Evaluation Metrics. We used three metrics widely adopted in existing research concerning obstacle detection studies. The first is the area under the ROC curve (AUROC), which plots the true-positive rates against the false-positive rates at various threshold settings. The second corresponds to average precision (AP), i.e., the area under the precision–recall curve. The third metric corresponds to the false-positive rate at 95% true-positive rate (FPR95).

Evaluation using the LostAndFound dataset. The test images in LostAndFound contain not only road obstacles but also other unknown objects. However, this dataset provides annotations for road obstacles exclusively. In other words, regardless of the identification of pixels corresponding to an unknown object, if those pixels do not indicate a road obstacle, they are considered false positives. Therefore, additional processing is required to mask unknown objects other than road obstacles, thereby facilitating appropriate evaluation. Following existing studies [17], [5], we utilized the ground-truth annotations to mask all pixels except those corresponding to the road and obstacles thereupon.

B. Evaluation Results

Table I presents the experimental results obtained by applying the proposed method to the LostAndFound Test, Fishyscapes Validation, and Road Anomaly datasets. The performances of the candidate methods are first compared considering the setting without OoD data. Among the five methods presented in the top row of Table I, the proposed method outperforms the other approaches when applied to the LostAndFound Test and Road Anomaly datasets. SML [4] demonstrates better performance than our method on the Fishyscapes Validation dataset, but its performance is inferior

TABLE I. Performance comparison between proposed and existing methods.

Methods	OoD Data	LostAndFound Test			Fishyscapes Validation			Road Anomaly		
		FPR95↓	AP↑	AUROC↑	FPR95↓	AP↑	AUROC↑	FPR95↓	AP↑	AUROC↑
Softmax Entropy [33]		19.45	39.58	95.40	30.55	7.92	91.38	70.93	17.06	69.35
Max Logit [34]		16.44	53.06	96.91	30.74	12.59	93.89	68.03	18.64	72.82
Ohgushi et al. (2020) [5]		14.06	50.68	97.16	36.86	6.07	90.83	67.64	20.25	71.92
SML [4]		35.51	39.65	92.87	14.53	36.56	96.34	50.91	25.32	81.37
Ours (w/o OoD data)		3.92	81.50	98.94	27.92	25.88	93.81	44.15	48.41	89.21
Outlier Exposure [39]	✓	15.76	70.21	97.80	36.94	22.64	93.45	67.83	19.71	70.61
Outlier Head [40]	✓	13.92	73.24	97.61	32.18	24.32	94.04	71.41	24.30	73.45
SynBoost [6]	✓	22.04	78.64	96.63	46.43	48.11	94.72	66.15	35.52	81.16
Ours (w/ OoD data)	✓	1.17	87.74	99.52	24.63	48.96	95.10	45.37	49.07	88.78

TABLE II. Comparison between the results of the unknown scores (US) defined in Eq. 1 and the unknown objectness scores (UOS) defined in Eq. 2.

Methods	LostAndFound Test			Fishyscapes Validation			Road Anomaly		
	FPR95↓	AP↑	AUROC↑	FPR95↓	AP↑	AUROC↑	FPR95↓	AP↑	AUROC↑
US (w/o OoD data)	29.49	33.96	93.83	43.92	32.19	91.53	72.73	23.26	70.64
UOS (w/o OoD data)	3.92	81.50	98.94	27.92	25.88	93.81	44.15	48.41	89.21
US (w/ OoD data)	18.18	69.86	96.99	37.25	49.38	93.94	68.88	33.74	77.88
UOS (w/ OoD data)	1.17	87.74	99.52	24.63	48.96	95.10	45.37	49.07	88.78

to that of our method by a large margin on the LostAndFound Test set. The performance of the methods presented at the bottom of Table I were compared under the OoD data setting. The proposed method demonstrates the best performance for all the datasets.

C. Effectiveness of Objectness Scores

Table II shows the comparison between the results of the unknown scores defined in Eq. 1 and those of the unknown objectness scores defined in Eq. 2. We can see that most of the evaluation metrics are improved by considering the objectness scores, both with and without using OoD data. In particular, FPR95 values are significantly reduced due to the reduction of false positive predictions in the background regions. On the other hand, when pixels corresponding to road obstacles have low objectness scores, it can lead to reduced true positive predictions. In fact, this is the reason why AP values of the unknown objectness scores are slightly lower than those of the unknown scores for the evaluation using Fishyscapes Validation.

D. Grouping of Original Classes

Although labelled data with 19 classes are generally used in the Cityscapes dataset, these classes can be grouped into seven large categories—“road,” “flat(w/o road),” “human,” “vehicle,” “construction,” “object,” and “background” [36]. Table III presents the performance of the proposed approach when applied to the Fishyscapes Validation set between considering the grouped 7 and the original 19 classes under the with and without OoD data settings. As can be seen, the proposed method demonstrates better performance when trained using 7 rather than 19 classes. This is because some of the 19 classes are rare classes that do not appear frequently in the training data. Pixels corresponding to such classes are often ambiguously predicted under other similar classes. This indicates that such pixels are more likely to achieve higher unknown scores and yield false-positive results. Therefore,

TABLE III. Comparison between training results obtained considering the original 19 Cityscapes and 7 grouped classes.

Methods	OoD data	Fishyscapes Validation		
		FPR95 ↓	AP ↑	AUROC ↑
Ours w/ 19 classes		27.92	25.88	93.81
Ours w/ 7 classes		25.49	32.43	93.82
Ours w/ 19 classes	✓	24.63	48.96	95.10
Ours w/ 7 classes	✓	9.85	66.11	98.26

TABLE IV. Comparison of Cityscapes mIoU values.

Methods	OoD data	Cityscapes mIoU
Softmax Entropy [33]		77.74
Outlier Exposure [39]	✓	68.83
Outlier Head [40]	✓	77.27
Ours	✓	76.85

incorporating rare classes under a large category can facilitate a reduction in the prediction ambiguity among similar classes and improve performance.

E. Comparison to Softmax Heads

When segmentation models are trained with a softmax head, uncertainty scores, such as entropy [33] or max logit [34] can be used instead of the unknown scores defined in Eq. 1. However, when trained with OoD data, the softmax head is possible to degrade the segmentation performance. Outlier Exposure [39] is a technique to yield low maximum softmax values for OoD data by minimizing the cross-entropy loss between model predictions and uniform distribution. As shown in Tab. IV, the mIoU of Outlier Exposure is lower than the others. The predictions performed using a softmax head are based on the assumption that each pixel is assigned to exactly one class. That is, it is not suitable for representing ambiguous states. On the other hand, predictions performed using a sigmoid head can be more suitable, because the sigmoid head allows each pixel to be assigned to multiple classes simultaneously or not to any classes. In fact, we can see that the performance degradation of the proposed method

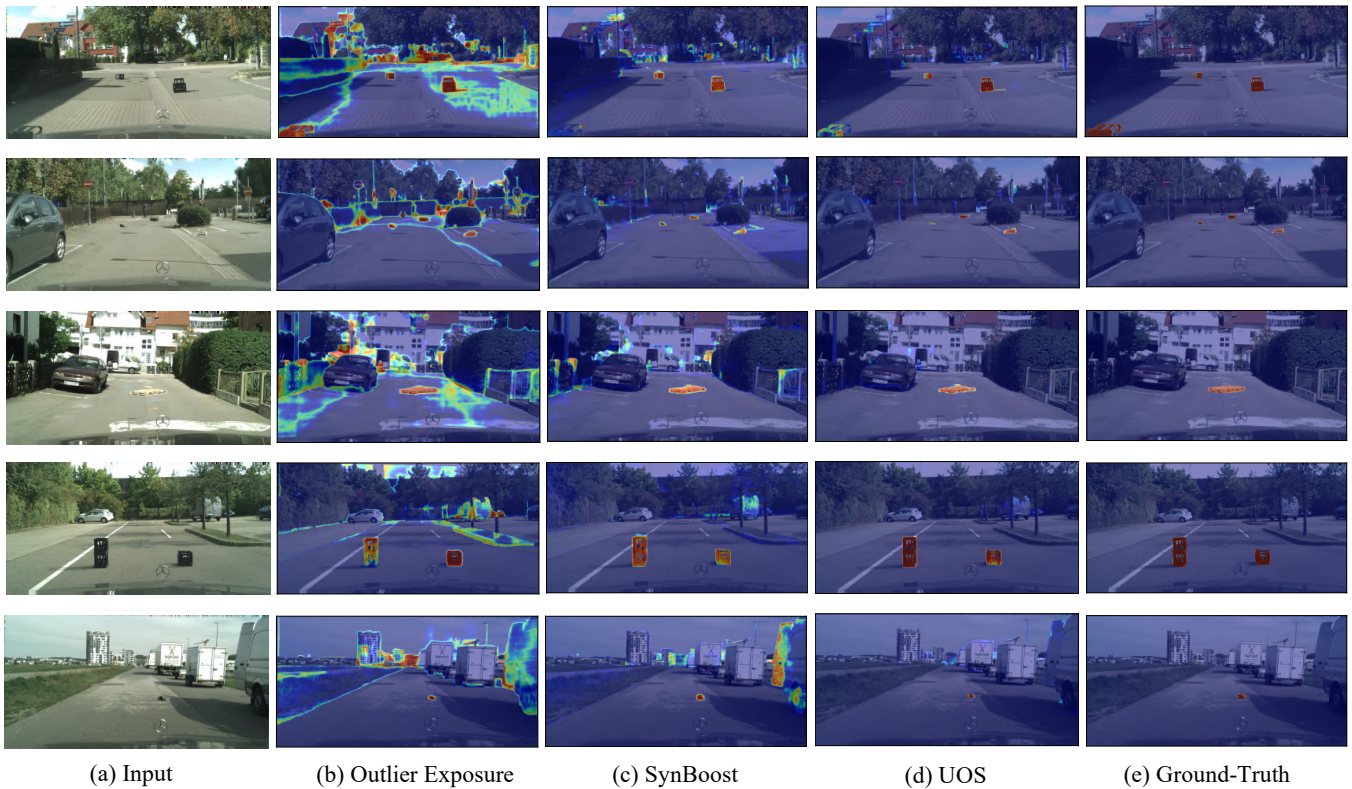


Fig. 5. Qualitative results for Outlier Exposure, Synboost, and the unknown objectness score (UOS).

TABLE V. Comparison between computational times per image required by different candidate models measured over 100 repetitions of the experiment.

Methods	Computational time per image (ms)
SynBoost [6]	190.97 ± 18.37
Ohgushi et al. (2020) [5]	125.10 ± 4.39
SML [4]	20.79 ± 2.20
Ours	15.57 ± 2.19

is mitigated compared to Outlier Exposure.

F. Comparison of Computational Costs

Table V compares the computational time per image required by the state-of-the-art methods [5], [4], [6]. All experiments were performed on a workstation equipped with Tesla V100 GPUs. To facilitate a fair comparison, the computational times were measured using the same segmentation model as ours. The methods based on the perceptual difference [5], [6] require the use of extra networks, hence, incurring additional computational time compared to other approaches. The method proposed in [4] also does not require any extra networks, and thus its computational time is close to the proposed method. However, its computational time is longer than the proposed method due to the additional post-processing required.

G. Qualitative Evaluation

Figure 5 presents the qualitative results for Outlier Exposure, SynBoost, and the unknown objectness score. It is

evident that the use of the unknown object score significantly reduces the number of false negative samples compared to the other two methods. This reduction can be attributed to the fact that the unknown objectness score explicitly considers objectness, which substantially lowers the anomaly scores in the background regions.

V. LIMITATIONS

The performance of the proposed method depends on the objectness scores. This means that the method’s ability to detect road obstacles may be compromised if it cannot recognize the pixels of these obstacles as objects. Therefore, devising strategies to grasp more universal object concepts stands as a crucial direction for future research.

VI. CONCLUSIONS

In this paper, we presented a novel method for identifying unknown objects in images of driving scenarios. The method is based on a semantic-segmentation network with a sigmoid head that allows us to obtain objectness scores and predicted probabilities of predefined classes simultaneously. By incorporating the objectness score into the unknown scores, our method is capable of stably detecting unknown objects in driving scenes. We demonstrated the effectiveness of our approach by evaluating it on publicly available datasets and showing that it outperforms state-of-the-art methods. Our results indicate that the proposed method has the potential to enhance the safety of autonomous driving systems by enabling the detection of unknown road obstacles.

REFERENCES

- [1] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [3] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 213–229.
- [4] S. Jung, J. Lee, D. Gwak, S. Choi, and J. Choo, "Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 425–15 434.
- [5] T. Ohgushi, K. Horiguchi, and M. Yamanaka, "Road obstacle detection method based on an autoencoder with semantic segmentation," in *Proceedings of the Asian Conference on Computer Vision*, 2020.
- [6] G. Di Biase, H. Blum, R. Siegwart, and C. Cadena, "Pixel-wise anomaly detection in complex driving scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2021, pp. 16918–16927.
- [7] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proceedings of the European Conference on Computer Vision*, 2018.
- [8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 833–851.
- [9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the International Conference on Neural Information Processing Systems*, 2014, p. 2672–2680.
- [10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the International Conference on Learning Representations*, 2014.
- [11] C. Creusot and A. Munawar, "Real-time small obstacle detection on highways using compressive rbm road reconstruction," in *IEEE Intelligent Vehicles Symposium*, 2015, pp. 162–167.
- [12] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, "Deep auto-encoding models for unsupervised anomaly segmentation in brain mr images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2019, pp. 161–169.
- [13] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Information Processing in Medical Imaging*, 2017, pp. 146–157.
- [14] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *Medical Image Analysis*, vol. 54, pp. 30–44, 2019.
- [15] H. Zenati, C.-S. Foo, B. Lecouat, G. Manek, and V. Chandrasekhar, "Efficient gan-based anomaly detection," *ArXiv*, vol. abs/1802.06222, 2018.
- [16] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4666–4679, 2020.
- [17] K. Lis, K. Nakka, P. Fua, and M. Salzmann, "Detecting the unexpected via image resynthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2152–2161.
- [18] Y. Xia, Y. Zhang, F. Liu, W. Shen, and A. L. Yuille, "Synthesize then compare: Detecting failures and anomalies for semantic segmentation," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 145–161.
- [19] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *Proceedings of the IEEE International Conference on Computer Vision*, Oct 2017.
- [20] X. Liu, G. Yin, J. Shao, X. Wang, and h. Li, "Learning to predict layout-to-image conditional convolutions for semantic image synthesis," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations*, 2015.
- [23] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," in *Proceedings of the International Conference on Learning Representations Workshops*, 2016.
- [24] —, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of the International Conference on Machine Learning*, vol. 48, New York, New York, USA, 20–22 Jun 2016, pp. 1050–1059.
- [25] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [26] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [28] V. B. Alex Kendall and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," in *Proceedings of the British Machine Vision Conference*, 2017, pp. 57.1–57.12.
- [29] P.-Y. Huang, W.-T. Hsu, C.-Y. Chiu, T.-F. Wu, and M. Sun, "Efficient uncertainty estimation for semantic segmentation in videos," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 520–535.
- [30] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 680–688.
- [31] J. Mukhoti and Y. Gal, "Evaluating bayesian deep learning methods for semantic segmentation," *arXiv preprint arXiv:1811.12709*, 2018.
- [32] F. K. Gustafsson, M. Danelljan, and T. B. Schon, "Evaluating scalable bayesian deep learning methods for robust computer vision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 2020.
- [33] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proceedings of the International Conference on Learning Representations*, 2017.
- [34] D. Hendrycks, S. Basart, M. Mazeika, M. Mostajabi, J. Steinhardt, and D. Song, "Scaling out-of-distribution detection for real-world settings," *arXiv preprint arXiv:1911.11132*, 2019.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [36] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M.ENZWEILER, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [37] P. Pinggera, S. Ramos, S. Gehrig, U. Franke, C. Rother, and R. Mester, "Lost and found: Detecting small road hazards for self-driving vehicles," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016, p. 1099–1106.
- [38] H. Blum, P.-E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena, "The fishyscapes benchmark: Measuring blind spots in semantic segmentation," *International Journal of Computer Vision*, 2021.
- [39] D. Hendrycks, M. Mazeika, and T. Dieterich, "Deep anomaly detection with outlier exposure," in *Proceedings of the International Conference on Learning Representations*, 2019.
- [40] P. Bevandić, I. Krešo, M. Oršić, and S. Šegvić, "Simultaneous semantic segmentation and outlier detection in presence of domain shift," *Pattern Recognition*, pp. 33–47, 2019.