

# Road Anomaly Segmentation Based on Pixel-wise Logit Variance with Iterative Background Highlighting

**Dongkun Lee**

Korea Advanced Institute of Science and Technology

**Han-Gyu Kim**

Naver (South Korea)

**Ho-Jin Choi** (✉ [hojinc@kaist.ac.kr](mailto:hojinc@kaist.ac.kr))

Korea Advanced Institute of Science and Technology

---

## Article

### Keywords:

**Posted Date:** July 21st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1845106/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Road Anomaly Segmentation Based on Pixel-wise Logit Variance with Iterative Background Highlighting

Dongkun Lee<sup>1</sup>, Han-Gyu Kim<sup>2</sup>, and Ho-Jin Choi<sup>1,\*</sup>

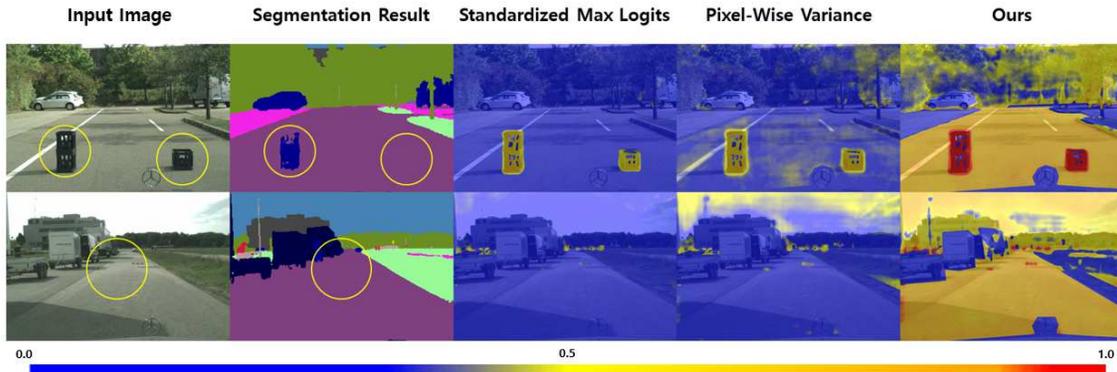
<sup>1</sup>Korea Advanced Institute of Science and Technology, School of Computing, Daejeon, 34141, Republic of Korea

<sup>2</sup>NAVER Corp, CLOVA, Gyeonggi-do, 13529, Republic of Korea

\*hojinc@kaist.ac.kr

## ABSTRACT

Anomaly segmentation on the urban landscape scene is an important task in autonomous driving. This process exploits a pre-trained semantic segmentation network to estimate anomalous regions. The anomaly segmentation approaches implemented with extra requirements such as out-of-domain data, extra network, or network retraining might increase computational cost or degradation of segmentation performance. In this research, to exploit information from the segmentation network for more robust anomaly segmentation, we propose the use of pixel-wise logit variance, which tends to be small for anomalies as network outputs even logits without confidence. Additionally, to detect anomalous objects on the background robustly, iterative background highlighting is also proposed, which is implemented by feeding the logits back into the linear classifier of the network. We achieved state-of-the-art performance among anomaly segmentation approaches without extra requirements, reaching relative average precision improvements of 21.7% on Fishyscapes Lost&Found and 17.4% on Fishyscapes Static compared to the state-of-the-art method.



**Figure 1.** Examples of our anomaly segmentation method. Yellow circle indicates location of anomalous object. When an image with anomalous object is used as input, there exist incorrectly classified pixels after semantic segmentation. Except for conventional standardized max logits, our approach also adopts pixel-wise logit variance, resulting in a better detection of anomalies.

## Introduction

Anomaly detection is an important task in image processing for identifying abnormal data or unseen defects that do not fit the normal data distribution. Previous studies on deep anomaly detection have been conducted in various domains, such as video analysis<sup>1-3</sup>. Anomaly segmentation, which is an advanced type of anomaly detection that specifies the anomaly region in the given image, is introduced for more sophisticated and safety-critical applications such as autonomous driving. A self-driving car that does not recognize an anomalous object may result in critical consequences such as roadkill or vehicle damage. Recent studies on anomaly segmentation of urban landscapes<sup>4-9</sup> show moderate performance on seen images. However, those models are easily confused by images containing unseen objects while real anomalies are more likely to be unseen in a practical situation.

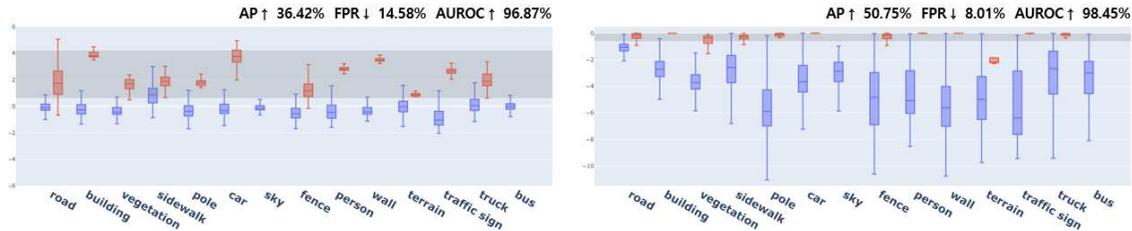
To resolve this problem, pixel-wise anomaly segmentation has been proposed to detect anomalous regions by exploiting

semantic segmentation networks.<sup>10,11</sup> use laborious human intervention such as extra data with anomaly labels or out-of-domain (OoD) data for pixel-wise anomaly detection. However, it is difficult to collect sufficient anomalous data. Besides, these approaches are strongly dependent on human intuition.<sup>12–16</sup> adopt extra network or retrain segmentation network for pixel-wise anomaly segmentation, which costs more computational resources.

For effective anomaly segmentation without any extra requirements such as extra data, network retraining or additional network, maximum softmax probability (MSP) and max logits are adopted with a simple assumption that MSP or max logits of anomaly regions are lower than those of normal regions<sup>17,18</sup>. Among those methods, standardized max logits (SML)<sup>19</sup> which standardizes max logits for robustness on out-of-domain objects and outperformed other logit-based anomaly segmentation methods.

However, SML has two limitations. The first limitation is that the performance of SML is highly dependent on the hyperparameters, which are the sample mean and variance of the logits of each class<sup>19</sup>. The second limitation is that the assumption of anomaly regions having low max logits is not always correct, because the segmentation model confused by anomalous objects may output high logits for all classes simultaneously.

For better utilizing the information included in the output of the segmentation network, we propose to use pixel-wise logit variance together instead of only using the max logits. For normal regions, segmentation network tends to output high logits for a specific class confidently, leading to a high pixel-wise logit variance; for anomaly regions, segmentation network is confused so that it may output relatively uniform logits for all classes, resulting in a small pixel-wise logit variance. Compared to SML, pixel-wise logit variance is more robust against the case that the segmentation network outputs uniformly high logits for all classes. Moreover, adopting pixel-wise logit variance may help reduce the degree of dependency on the hyperparameters of SML, as computing variance does not need any hyperparameters.



**Figure 2.** Box plot of anomaly score comparison between SML (left) and our method (right) on Fishyscapes Lost&Found validation dataset. We took up to 100,000 samples from each class. X-axis represents training classes sorted by the appearance frequency in training data. Y-axis represents the anomaly score (higher for anomaly). Red and blue represent anomalous pixels and in-distribution pixels, respectively. Gray region denotes false positive range at 95% true positive threshold ( $FPR_{95}$ ). The upper and lower bounds of each box refer to Q1 and Q3, respectively, and the upper and lower whiskers in the box plot are drawn in the 1.5 interquartile range. Samples out of whisker bounds and classes with zero pixels are omitted.

Additionally, we propose iterative background highlighting for further enhancement of the anomaly objects on the background, because meaningful anomalies are more likely to occur on background areas such as roads in the autonomous driving scenario. In order to briefly show this performance difference, we visualized the performance difference using 100,000 pixel samples per class using a boxplot as Figure 2. It is shown in the figure that our method shows a much higher anomaly segmentation performance on Fishyscapes validation dataset than the previous state-of-the-art. Additionally, we have significantly reduced false-positive pixels for the realistic scene input. Figure 1 compares our approach to SML with two examples, showing that our approach detects anomalous objects better than SML.

The main contributions of our work are summarized as follows:

- We propose to use pixel-wise logit variance and standardized maximum logits together for anomaly segmentation to reduce the dependency on the hyperparameters.
- Iterative background highlighting for better segmentation of anomalies on the background is proposed, which is important for practical use.
- We achieve a new state-of-the-art performance on Fishyscapes Lost&Found task among anomaly segmentation methods without any extra requirements with a large gap of **average precision improvement of 21.7% on Fishyscapes Lost&Found and 17.4% on Fishyscapes Static** (publicly available on Fishyscapes leaderboard website - <https://fishyscapes.com/results>).

The remainder of this paper is organized as follows. In **Related Work** section, the relevant background and related studies on anomaly segmentation described, **Proposed Method** introduces our novel anomaly segmentation method Pixel-wise Logit

Variance and Iterative Background Highlighting. [Results and Discussions](#) section presents detailed experimental design and results with discussions. Finally, In [Conclusion](#) section, concluding remarks and future applications are summarized.

## Related Work

This section describes the necessary background knowledge to understand the proposed method. We first introduce current research trends in anomaly detection and localization are discussed alongside their limitations regarding their core generation architecture. An overview of research on anomaly segmentation in urban landscape scene follows, along with a brief introduction to state-of-the-art logit-based anomaly detection, SML.

### Deep anomaly segmentation

Generally, supervised deep anomaly detection in image processing uses a set of deep neural networks that are trained on task-dependent datasets carefully created by a human expert for an intended task, such as anomaly classification<sup>20</sup> and novelty detection<sup>21</sup>. However, there are usually insufficient abnormal data in the real-world dataset for training a deep neural network in a supervised manner.

### Unsupervised anomaly detection

In order to overcome such data-insufficiency problem, unsupervised anomaly detection are developed. In unsupervised anomaly detection, neural model should automatically understanding the distribution of data based solely on the given normal data in the absence of any user-tagged input. The collection of a normal image dataset is relatively inexpensive. Its primary tasks are generative anomaly detection<sup>22,23</sup>, time series modeling<sup>24,25</sup>, anomaly classification<sup>26,27</sup>, novelty detection<sup>28</sup>, and anomaly segmentation<sup>29–33</sup>. However, previous approaches requires a considerable amount of anomaly data is required. It becomes more difficult to apply the unsupervised manner to anomaly segmentation where the model should distinguish abnormal region from the local image without supervision which is more complex than a two-class classification task.

### Anomaly Segmentation in Urban Landscape Scene

One of the most popular applications of anomaly segmentation is autonomous driving, where anomaly segmentation is conducted on urban landscape scene.<sup>10,11,13,14,16,34,35</sup> leverage the existing segmentation network to improve the performance of recognizing anomalous objects on urban landscape scene. Among those methods,<sup>10,14,35–37</sup> require extra training of the segmentation network and additional networks. Adopting additional network or retraining of segmentation network increases the time complexity while autonomous driving requires real-time abnormal segmentation as it is a safety-critical task, Meanwhile,<sup>11,16,34,36,37</sup> utilize OoD data. However, using OoD data is impractical as out-of-domain cases rarely occur in a practical environment.

### Standardized Maximum Logits

Standardized maximum logits (SML)<sup>19</sup> focuses on fast and effective anomaly segmentation without using additional data, segmentation network retraining, or extra network architecture. As the distributions of the logits of classes are different from each other, SML tries to project all distributions to the same scale by standardizing logits so that the logits of different classes can be fairly compared. The mean and variance of each class for standardizing logit distribution are computed using the logits of pixels of training data. The standardization step of SML may be described as follows:

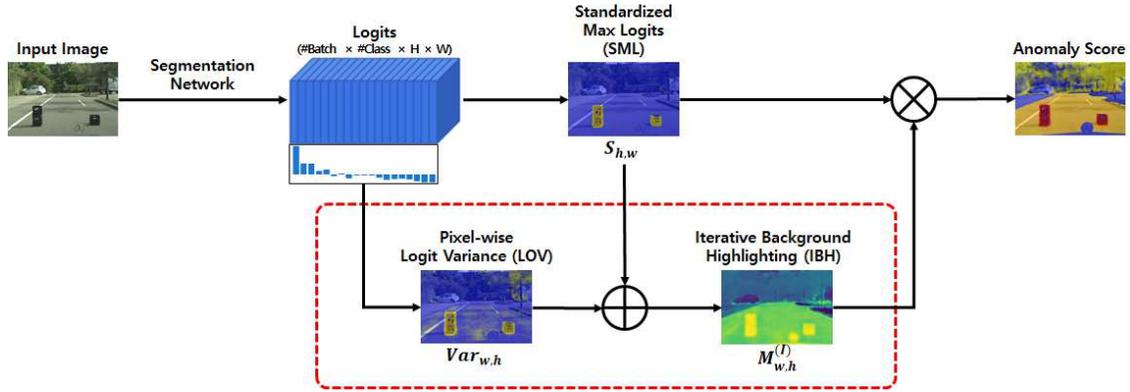
$$\mu_c = \frac{\sum_i \sum_{h,w} \mathbb{1}(\hat{Y}_{h,w}^{(i)} = c) \cdot L_{h,w}^{(i)}}{\sum_i \sum_{h,w} \mathbb{1}(\hat{Y}_{h,w}^{(i)} = c)}, \quad (1)$$

$$\sigma_c^2 = \frac{\sum_i \sum_{h,w} \mathbb{1}(\hat{Y}_{h,w}^{(i)} = c) \cdot (L_{h,w}^{(i)} - \mu_c)^2}{\sum_i \sum_{h,w} \mathbb{1}(\hat{Y}_{h,w}^{(i)} = c)}, \quad (2)$$

$$S_{h,w} = \frac{L_{h,w} - \mu_{\hat{Y}_{h,w}}}{\sigma_{\hat{Y}_{h,w}}}, \quad (3)$$

where  $\mu_c$  and  $\sigma_c^2$  are the mean and variance of class  $c$  respectively,  $i$  is the index of the training data,  $(h, w)$  represents the coordinate of a pixel,  $\hat{Y}_{h,w}$  is the predicted class,  $L_{h,w}$  is the logit,  $\mathbb{1}$  is the indicator function, and  $S_{h,w}$  denotes SML. In<sup>19</sup>, SML is used to differentiate normal and abnormal pixels. Pixels with large SML are classified as normal pixels and other pixels are classified as anomaly pixels.

## Proposed Method



**Figure 3.** Overview of the proposed approach. The logits are generated with the segmentation network. LOV and SML are obtained from the given logits. Base anomaly score is obtained with the summation of LOV and SML to complement each other. Subsequently, IBH highlights background pixels. The final anomaly score is computed by combining the base anomaly score and the highlighted background via the Hadamard product.

The conventional SML-based anomaly segmentation has two limitations. The first limitation is that SML uses sample mean  $\mu_c$  in Equation 1 and sample variance  $\sigma_c^2$  in Equation 2 as hyper-parameters in the inference stage. We have found out that the performance of final anomaly segmentation is highly related to these hyper-parameters, implying that the performance might be poor if the hyper-parameters does not fit the realistic data. Besides, if alternative segmentation network is used instead of the segmentation network in<sup>19</sup>, the hyper-parameters computed via Equations 1 and 2 does not always guarantee the best performance. The second limitation is that classifying pixels with small SML as anomaly class does not always stand. The confused segmentation network would likely to emit uniformly distributed logits, but it does not guarantee that the max logit is small because logits for all classes may evenly be large.

In order to overcome above-mentioned problems, we propose to use pixel-wise logit variance. Variance has several advantages compared to SML: computing variance does not require any hyper-parameter because variance might be computed only using the logits emitted by the segmentation network during the inference; variance is better criterion for finding out whether the logits are uniformly distributed compared to SML, where smaller variance implies more uniform distribution.

Instead of using variance or SML independently, we propose to ensemble variance and SML together in order to further improve the anomaly segmentation performance. Besides, we also propose iterative background highlighting in order to find out anomalies on background more accurately because in practical situation anomalies are more likely to occur on background. The overview of the proposed method is shown in Figure 3. Detailed information of each module is explained in following subsections.

### Pixel-wise Logit Variance

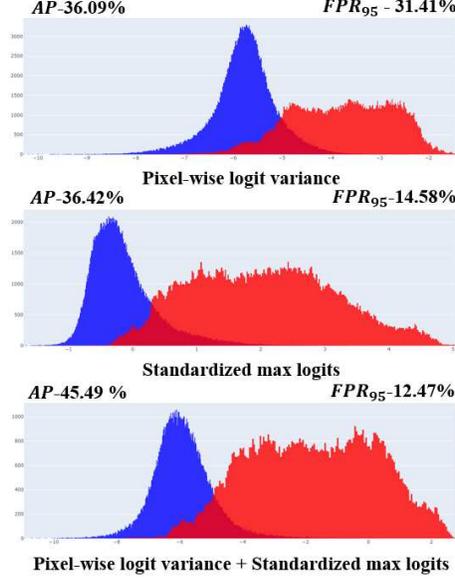
Pixel-wise logit variance (LOV) is the variance calculated using the logits of a pixel output by the segmentation network, which might be described by the following equation:

$$Var_{w,h} = \frac{\sum_c (L_{w,h}(c) - \mu_{w,h})^2}{C}, \quad (4)$$

where  $Var_{w,h}$  is the logit variance of pixel  $(w, h)$ ,  $\mu_{w,h}$  is the mean of the logits of the pixel,  $L_{w,h}(c)$  is the logit of class  $c$ , and  $C$  is the number of classes, respectively. As shown in the equation, the pixel-wise logit variance does not need any hyper-parameters. We adopt the logit variance for anomaly segmentation with an assumption that the variance should be small for anomaly pixels compared to that of normal pixels because segmentation network tends to emit more uniformly distributed logits for an anomaly pixel.

In our proposed method, we use an ensemble of SML and logit variance to achieve better performance. Among various ensemble methods, we discovered that simply using the summation of SML and logit variance shows the best performance. Additionally, we used iterative boundary suppression and dilated smoothing, the post-processing algorithms used in<sup>19</sup> with the same hyper-parameters for fair comparison. Thus, the base anomaly score may be computed with following equation:

$$A_{h,w} = P(Var_{w,h} + S_{w,h}), \quad (5)$$



**Figure 4.** Histogram of pixel-wise logit variance, standardized max logit and summation of standardized logits and class-wise variance (without iterative background highlighting) in Fishyscapes Lost&Found. X-axis denotes pixel-wise anomaly score. Red (anomaly) and blue (normal) are obtained from 400,000 (200,000 each) randomly chosen pixels from 100 images. Summation of max logits and pixel-wise logit variance clearly reduces the  $FPR_{95}$  and boosts AP score.

where  $A_{w,h}$  is the base anomaly score,  $P$  is post-processing function of iterative boundary suppression and dilated smoothing, and  $Var_{w,h}$  and  $S_{h,w}$  denote LOV and SML, respectively. The performance variation according to the ensemble methods is further illustrated in Figure 4.

### Iterative Background Highlighting

In the autonomous driving scenario, detecting anomaly objects in the background such as road is more important, because the background is the most frequent type of class to meet and anomaly objects on the background usually endanger safety. To enhance the performance of anomaly segmentation on background, we propose iterative background highlighting (IBH). The main idea of this approach is to exploit the last fully connected layer of the segmentation network for background highlighting.

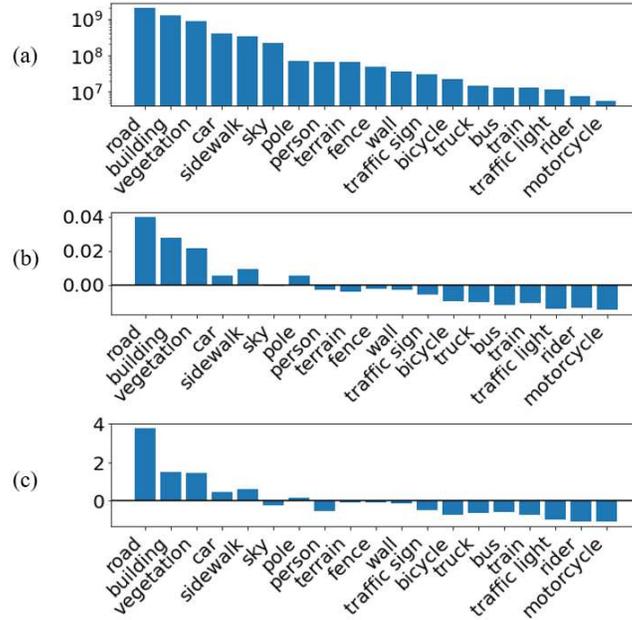
---

#### Algorithm 1 Iterative Background Highlighting

---

- 1: Initialize  $M_{w,h}^{(0)} \leftarrow A_{w,h}$
  - 2: Initialize  $F_{w,h}^{(0)}$  with the input of linear classifier
  - 3: Set  $i = 0$
  - 4: **while**  $i < I$  **do**
  - 5:   Blending:  $F_{w,h}^{(i+1)} \leftarrow (1 - M_{w,h}^{(i)}) \cdot F_{w,h}^{(i)} + M_{w,h}^{(i)} \cdot \max F_{w,h}^{(i)}$
  - 6:   Forward of linear classifier:  $L_{w,h}^{(i+1)} \leftarrow W_{last} F_{w,h}^{(i+1)}$
  - 7:   Mask generation:  $M_{w,h}^{(i+1)} \leftarrow \frac{L_{w,h}^{(i+1)} - \min L_{w,h}^{(i+1)}}{\max L_{w,h}^{(i+1)} - \min L_{w,h}^{(i+1)}}$
  - 8: **end while**
  - 9: Final anomaly score:  $\hat{A}_{w,h} \leftarrow A_{h,w} \cdot (1 - M_{w,h}^{(I)})$
- 

As background class frequently appears in training data, segmentation network has high prior for background class during the training process, resulting in emitting relatively high logits on average during the inference stage, which is shown in Figure 5-(a) and Figure 5-(b). Thus, an average of the weights for the background class is likely to be larger than those of other classes, as shown in Figure 5-(c), implying that a larger input to a linear classifier will probably result in larger logit for background classes. We adopted an adequately large value, which is the maximum of all input of linear classifier, to target pixels that would result in enhancement of the logits of the targets.



**Figure 5.** Illustration for showing the correlation between logit and weight of linear classifier. The classes are sorted by appearance frequency in the training dataset. Each figure indicates: (a) Total number of appearances for each class, (b) Mean of the weight values of each linear classifier, and (c) Mean of the logit values for each class.

**Table 1.** Effect of each blending method. There was no difference between adding and averaging because the post-processing algorithm works regardless of scaling, and both Max and Min were inferior to the summation operation.

Method	FS Lost&Found		FS Static	
	AP $\uparrow$	FPR $_{95}\downarrow$	AP $\uparrow$	FPR $_{95}\downarrow$
Min	0.81	28.37	4.47	<b>27.56</b>
Max	47.74	16.50	45.91	51.42
Average/Add	<b>50.75</b>	<b>8.01</b>	<b>54.06</b>	39.07

With such phenomenon, we design background highlighting algorithm as Algorithm 1. We use anomaly score for initial blending mask because we want to include the confident anomalies into the background highlighting target. During the iterative steps, the blending mask gets larger because of the blending step, where pixels with large blending mask, which belong to background, obtains larger input for linear classifier. Such iteration is executed for  $I$  times, for which we adopted 3 in our approach. Finally, the blending mask is multiplied with the base anomaly score to obtain the final anomaly score. The variations of the performance according to blending method and  $I$  are further provided in Table 1 and Table 2, respectively.

**Table 2.** Quantitative results according to to number of iterations  $I$ . We report performance over 100 images on Fishyscapes Lost&Found validation dataset. The highest AP score was observed when  $I = 3$ .

Method	AP $\uparrow$	FPR $_{95}\downarrow$	AUROC $\uparrow$
$I = 0$	45.49	12.47	97.34
$I = 1$	48.94	7.87	98.36
$I = 2$	50.44	<b>7.57</b>	<b>98.49</b>
$I = 3$	<b>50.75</b>	8.01	98.45
$I = 4$	49.71	8.90	98.31

**Table 3.** Anomaly segmentation performance reported on Fishyscapes leaderboard. The best performances are highlighted in bold. Our method achieved the new state-of-the-art performance among approaches without any extra requirements in Fishyscapes leaderboard.

Method	Method Requirements			FS Lost&Found		FS Static		mIoU
	Seg. Net. Retrain	Extra Net.	OoD Data	AP $\uparrow$	FPR $_{95}\downarrow$	AP $\uparrow$	FPR $_{95}\downarrow$	
Density - Single-layer NLL <sup>13</sup>	✗	✓	✗	3.01	32.90	40.86	21.29	80.30
Density - Minimum NLL <sup>13</sup>	✗	✓	✗	4.25	47.15	62.14	17.43	80.30
Density - Logistic Regression <sup>13</sup>	✗	✓	✓	4.65	24.36	57.16	13.39	80.30
Image Resynthesis <sup>14</sup>	✗	✓	✗	5.70	48.05	29.60	27.13	81.40
Bayesian Deeplab <sup>35</sup>	✓	✗	✗	9.81	38.46	48.70	15.50	73.80
OoD Training - Void Class <sup>13</sup>	✓	✗	✓	10.29	22.11	45.00	19.40	70.40
Discriminative Outlier Detection Head <sup>37</sup>	✓	✓	✓	31.31	19.02	<b>96.76</b>	<b>0.29</b>	79.57
Dirichlet Deeplab <sup>36</sup>	✓	✗	✓	34.28	47.43	31.3	84.60	70.50
PEBAL <sup>34</sup>	✗	✗	✓	44.17	<b>7.58</b>	92.38	1.73	-
SynBoost <sup>16</sup>	✗	✗	✓	<b>44.47</b>	18.7	71.00	17.17	81.4
MSP <sup>18</sup>	✗	✗	✗	1.77	44.85	12.88	39.83	80.30
Entropy <sup>18</sup>	✗	✗	✗	2.93	44.83	15.41	39.75	80.30
kNN Embedding - Density <sup>13</sup>	✗	✗	✗	3.55	30.02	44.03	20.25	80.30
Standardized Max Logits <sup>19</sup>	✗	✗	✗	31.05	21.52	53.11	<b>19.64</b>	80.33
<b>Ours</b>	✗	✗	✗	<b>37.81</b>	<b>18.58</b>	<b>62.39</b>	45.65	80.33

## Results and Discussions

### Experimental Setting

#### Dataset

We used two popular benchmark datasets for anomaly detection on the urban landscape scene, which are Fishyscapes Lost&Found (FS Lost&Found) and Fishyscapes Static (FS Static)<sup>13</sup>. FS Lost&Found is a real-world dataset of road images with anomalous objects visible in the front sight of the vehicle. It contains 100 validation images and 275 undisclosed test images. FS Static is a synthetic dataset with blended anomalous objects on The PASCAL Visual Object Classes (PASCAL VOC) images. It has 50 validation images and 1000 undisclosed test images.

#### Implementation Details

For a more direct comparison with SML, which is previous state-of-the-art method, we mainly adopted pre-trained model of DeepLabv3+/ResNet101 architecture of official SML implementation<sup>19</sup> as the segmentation network. Besides, we also implemented our method on various segmentation networks from the mmsegmentation toolbox<sup>1</sup> to show the general superiority of our method.

#### Evaluation Metric

For performance evaluation, metrics provided by the Fishyscapes benchmark are adopted, which are average precision (AP) and false positive rate at 95% true positive rate (FPR $_{95}$ ). Fishyscapes benchmark suggests AP as the primary metric, as it is invariant under data imbalance and threshold. FPR $_{95}$  is also a meaningful metric because anomaly segmentation is generally used for safety-critical applications.

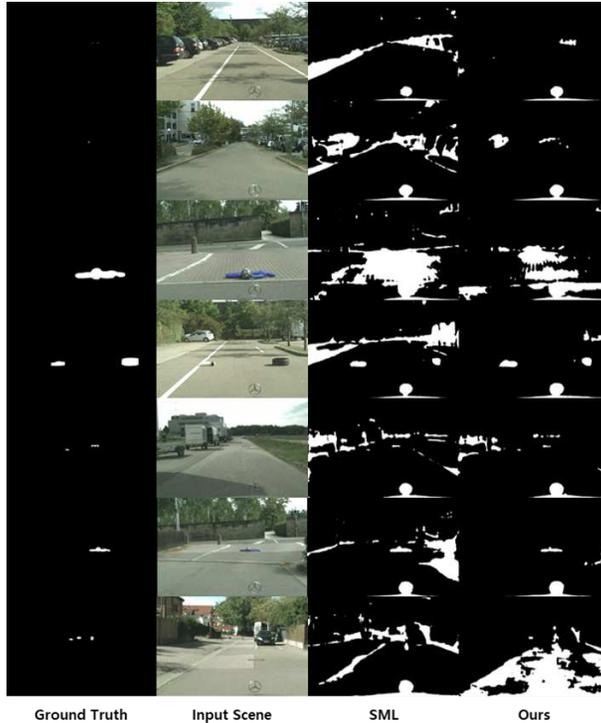
## Results

### Comparison on Fishyscapes Leaderboard

To show the superiority of our proposed method, we compared the performance of ours with the performances of anomaly segmentation methods on the Fishyscapes benchmark leaderboard until July 9th 2022. Those methods which have neither papers with detailed explanations nor publicly available source codes are not included in our comparison. The performance comparison result is shown in Table 3, which includes both AP and FPR $_{95}$  for benchmark datasets FS Lost&Found and FS Static. Moreover, this table includes the brief information for requirements of compared methods—whether to use retraining, extra network, or OoD data. As our method avoids extra requirements, MSP, entropy, kNN embedding and standardized max logits are main comparison targets. It is shown in Table 3 that ours outperforms any of the main comparison targets in AP score

<sup>1</sup><https://github.com/open-mmlab/msegmentation>

which is primary comparison metric. Besides, our method also outperforms in  $FPR_{95}$  for FS Lost&Found dataset. Compared with SML, which is the baseline study for us, performance significantly improved by 6.76 AP (21.7%) for FS Lost&Found and by 9.28 AP (17.4%) for FS Static. Additionally, we also compared our method to the methods with extra requirements, who certainly outperform our method because of extra benefits. However, our methods show competitive generalization ability across multiple datasets, ranking 3rd for FS Lost&Found and 4th for FS Static in AP score, and our method does not require any retraining, which helps preserve segmentation accuracy.



**Figure 6.** Anomaly detected with  $TPR_{95}$  on Fishyscapes Lost&Found validation dataset. White region indicates anomaly.

### Qualitative Analysis

We compare the results of SML and our approach on the validation dataset in Figure 6 for a qualitative comparison. As shown in the figure, our method usually yields a much smaller false positive area on real data. Our method generally showed significantly fewer false positives region on real-world dataset compared to SML. However, occasionally our method did not work properly when the road surface has an exceptional pattern such as a sidewalk block or dirty surface.

**Table 4.** Comparison of inference time. We used NVIDIA GeForce A100 for evaluation. We report mean inference time over 1000 trials on Fishyscapes Lost&Found validation dataset. When only the LOV was used, the performance degradation was negligible.

Method	Infer. Time (s)
Full Framework	2.90
w/o IBH	1.97
w/o LOV	2.88
w/o Max logits	2.88
w/o IBH & Max logits	1.95
SML	1.93

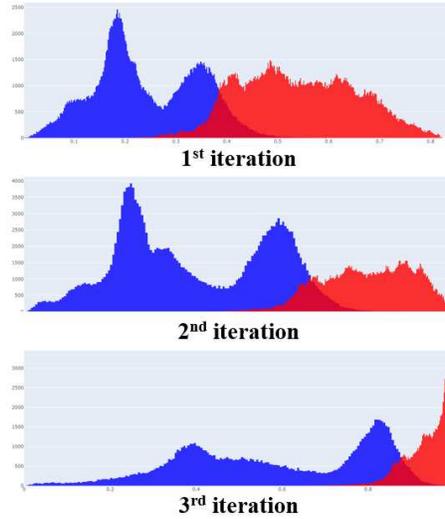
### Inference Time Analysis

Our method has the advantage of fast inference compared to methods requiring extra segmentation network. An analysis of the inference time of our method is additionally described in Table 4. Moreover, our approach can be applied to domains without

OoD data.

**Table 5.** Ablation study of SML, LOV and IBH. While LOV is inferior to SML, there is a clear performance improvement with ensemble of LOV and SML. IBH results in a performance improvement only for LOV-included methods.

Method	FS Lost&Found		FS Static	
	AP↑	FPR <sub>95</sub> ↓	AP↑	FPR <sub>95</sub> ↓
<b>Full Framework</b>	<b>50.75</b>	<b>8.01</b>	<b>54.06</b>	39.07
w/o IBH	45.49	12.47	52.05	19.17
w/o LOV	0.81	28.37	4.47	27.56
w/o Max logits	47.74	16.50	45.92	51.42
w/o IBH & Max logits	36.09	31.40	43.07	48.68
w/o IBH & LOV (SML; baseline)	36.42	14.58	48.60	<b>16.79</b>



**Figure 7.** Histogram of mask value for each iteration in Fishyscapes Lost&Found. X-axis denotes mask value  $M_{w,h}^{(i+1)}$ . Red (anomaly) and blue (normal) are obtained from 400,000 (200,000 each) randomly chosen pixels from 100 images. The histogram shows how anomaly score of normal pixels outside of the highlighted background area are separated from the anomaly pixels

### Ablation Study

We have also conducted the ablation study to find out how much each part of our method contributes to the anomaly segmentation performance, whose result is shown in Table 5. It is shown in the table that applying LOV to SML does help improve the performance. Besides, combined with IBH, LOV shows superior performance even without SML. The best performance is achieved when SML, LOV and IBH are used together. IBH separates the values of the pixel region with a large prior over iterations from the values in the other region. The change in mask value with each iteration is further illustrated in Figure 7.

### Framework Generalization

In order to show the superiority of our proposed method on various segmentation networks, we have conducted anomaly segmentation experiments on other segmentation networks from mmsegmentation toolbox instead of DeepLabv3+ with ResNet101 backbone network that provided by<sup>19</sup>, whose results are shown in Table 6. Although our proposed method does not show significant improvement for FPR<sub>95</sub> for FS Static dataset, the proposed method generally shows better AP for both dataset and better FPR<sub>95</sub> for FS Lost&Found dataset, implying that the proposed method works robustly along with various types of segmentation networks. In particular, for the two networks that has highest IoU score, ISANet<sup>39</sup> and OCRNet<sup>38</sup> our methods showed the superior performance than SML.

**Table 6.** Anomaly segmentation results with various segmentation networks on Fishyscapes validation dataset.

Segmentation Architecture	Method	FS Lost&Found		FS Static		mIoU
		AP $\uparrow$	FPR $_{95}\downarrow$	AP $\uparrow$	FPR $_{95}\downarrow$	
OCRNet <sup>38</sup>	SML	39.96	18.28	47.90	15.07	81.35
	LOV	48.40	16.61	47.67	21.67	
	Sum	<b>51.89</b>	12.65	<b>53.55</b>	12.05	
	SML + IBH	0.76	23.20	3.87	23.61	
	LOV + IBH	51.48	9.96	46.23	22.71	
	Sum + IBH	44.11	<b>7.82</b>	53.54	<b>10.30</b>	
ISANet <sup>39</sup>	SML	18.67	28.76	32.15	14.86	80.81
	LOV	35.41	47.47	29.60	37.97	
	Sum	32.21	32.08	36.18	16.60	
	SML + IBH	0.94	28.02	5.74	18.62	
	LOV + IBH	<b>40.18</b>	38.35	26.31	28.83	
	Sum + IBH	30.40	<b>14.42</b>	<b>36.46</b>	<b>11.76</b>	
DeepLabV3+ <sup>40</sup>	SML	7.11	28.62	30.53	18.21	80.52
	LOV	24.97	71.72	45.96	58.04	
	Sum	16.76	35.18	45.00	<b>18.07</b>	
	SML + IBH	1.16	<b>27.35</b>	7.66	20.53	
	LOV + IBH	<b>28.54</b>	50.79	<b>46.78</b>	64.54	
	Sum + IBH	13.23	34.68	22.89	49.08	
DeepLabV3 <sup>41</sup>	SML	24.58	31.45	39.05	<b>15.63</b>	80.10
	LOV	24.55	59.14	30.31	67.71	
	Sum	<b>30.57</b>	37.09	<b>39.82</b>	22.91	
	SML + IBH	0.66	31.34	4.04	30.99	
	LOV + IBH	27.15	38.85	29.86	63.84	
	Sum + IBH	25.87	<b>30.28</b>	38.23	33.54	

## Conclusion

We proposed a simple method called pixel-wise logit variance (LOV) and iterative background highlighting (IBH) to aid in the unexpected behavior of segmentation networks triggered by anomalous objects on the road. This approach does not require external datasets, additional training, or external network. We strengthened our method of recognizing pixel-wise logit output variance and iterative background highlighting based on two intuitions: that the network’s latent uncertainty is expressed through the variance value of the output value, and high logit implies more prior information. The experiment result shows that our approach achieves a new state-of-the-art performance in the Fishyscapes Lost&Found and Fishyscapes Static benchmark. Additionally, extensive experiments on various segmentation networks also demonstrate the superiority of our method over the previous state-of-the-art method. In future work, we will try to improve performance on detecting small anomalous objects which is difficult for logit-based methods.

## References

1. Nguyen, T.-N. & Meunier, J. Anomaly detection in video sequence with appearance-motion correspondence. In *Proceedings of the IEEE International Conference on Computer Vision*, 1273–1283 (2019).
2. Luo, W. *et al.* Video anomaly detection with sparse coding inspired deep neural networks. *IEEE Transactions on Pattern Analysis Mach. Intell.* (2019).
3. Li, N. & Chang, F. Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder. *Neurocomputing* **369**, 92–105 (2019).
4. Choi, S. *et al.* Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11580–11590 (2021).
5. Choi, S., Kim, J. T. & Choo, J. Cars can’t fly up in the sky: Improving urban-scene segmentation via height-driven attention networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9373–9383 (2020).
6. Zhu, Y. *et al.* Improving semantic segmentation via video propagation and label relaxation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8856–8865 (2019).

7. Zhu, Z., Xu, M., Bai, S., Huang, T. & Bai, X. Asymmetric non-local neural networks for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 593–602 (2019).
8. Fu, J. *et al.* Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3146–3154 (2019).
9. Yang, M., Yu, K., Zhang, C., Li, Z. & Yang, K. Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*, 3684–3692 (2018).
10. Bevandić, P., Krešo, I., Oršić, M. & Šegvić, S. Dense outlier detection and open-set recognition based on training with noisy negative images. *arXiv preprint arXiv:2101.09193* (2021).
11. Chan, R., Rottmann, M. & Gottschalk, H. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5128–5137 (2021).
12. Grcić, M., Bevandić, P. & Šegvić, S. Dense open-set recognition with synthetic outliers generated by real nvp. *arXiv preprint arXiv:2011.11094* (2020).
13. Blum, H., Sarlin, P.-E., Nieto, J., Siegwart, R. & Cadena, C. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *arXiv preprint arXiv:1904.03215* (2019).
14. Lis, K., Nakka, K., Fua, P. & Salzmann, M. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2152–2161 (2019).
15. Pinggera, P. *et al.* Lost and found: detecting small road hazards for self-driving vehicles. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1099–1106 (IEEE, 2016).
16. Di Biase, G., Blum, H., Siegwart, R. & Cadena, C. Pixel-wise anomaly detection in complex driving scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16918–16927 (2021).
17. Hendrycks, D. & Gimpel, K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136* (2016).
18. Hendrycks, D. *et al.* Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132* (2019).
19. Jung, S., Lee, J., Gwak, D., Choi, S. & Choo, J. 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*, 15425–15434 (2021).
20. Erfani, S. M. *et al.* From shared subspaces to shared landmarks: A robust multi-source classification approach. In *Thirty-First AAAI Conference on Artificial Intelligence* (2017).
21. Jumutc, V. & Suykens, J. A. Multi-class supervised novelty detection. *IEEE transactions on pattern analysis machine intelligence* **36**, 2510–2523 (2014).
22. Lawson, W., Bekele, E. & Sullivan, K. Finding anomalies with generative adversarial networks for a patrolbot. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 12–13 (2017).
23. Mishra, A., Krishna Reddy, S., Mittal, A. & Murthy, H. A. A generative model for zero shot learning using conditional variational autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2188–2196 (2018).
24. Dasigi, P. & Hovy, E. Modeling newswire events using neural networks for anomaly detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1414–1422 (2014).
25. Filonov, P., Kitashov, F. & Lavrentyev, A. RNN-based early cyber-attack detection for the tennessee eastman process. *arXiv preprint arXiv:1709.02232* (2017).
26. Racah, E. *et al.* ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. In *Advances in Neural Information Processing Systems*, 3402–3413 (2017).
27. Perera, P. & Patel, V. M. Learning deep features for one-class classification. *IEEE Transactions on Image Process.* **28**, 5450–5463 (2019).
28. Kliger, M. & Fleishman, S. Novelty detection with GAN. *arXiv preprint arXiv:1802.10560* (2018).
29. Zhao, S., Song, J. & Ermon, S. Towards deeper understanding of variational autoencoding models. *arXiv preprint arXiv:1702.08658* (2017).

30. Dehaene, D., Frigo, O., Combrexelle, S. & Eline, P. Iterative energy-based projection on a normal data manifold for anomaly localization. In *International Conference on Learning Representations* (2019).
31. Zhou, L., Deng, W. & Wu, X. Unsupervised anomaly localization using VAE and beta-VAE. *arXiv preprint arXiv:2005.10686* (2020).
32. Liu, W. *et al.* Towards visually explaining variational autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8642–8651 (2020).
33. Dehaene, D. & Eline, P. Anomaly localization by modeling perceptual features. *arXiv preprint arXiv:2008.05369* (2020).
34. Tian, Y. *et al.* Pixel-wise energy-biased abstention learning for anomaly segmentation on complex urban driving scenes. *arXiv preprint arXiv:2111.12264* (2021).
35. Mukhoti, J. & Gal, Y. Evaluating bayesian deep learning methods for semantic segmentation. *arXiv preprint arXiv:1811.12709* (2018).
36. Malinin, A. & Gales, M. Predictive uncertainty estimation via prior networks. *arXiv preprint arXiv:1802.10501* (2018).
37. Bevandić, P., Krešo, I., Oršić, M. & Šegvić, S. Simultaneous semantic segmentation and outlier detection in presence of domain shift. In *German Conference on Pattern Recognition*, 33–47 (Springer, 2019).
38. Yuan, Y., Chen, X. & Wang, J. Object-contextual representations for semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, 173–190 (Springer, 2020).
39. Huang, L. *et al.* Interlaced sparse self-attention for semantic segmentation. *arXiv preprint arXiv:1907.12273* (2019).
40. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 801–818 (2018).
41. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).

## Acknowledgements

This research was supported and funded by the Korean National Police Agency. [Project Name: XR Counter-Terrorism Education and Training Test Bed Establishment / Project Number: PR08-04-000-21]

## Author contributions statement

D.L. conceived and conducted the experiments, D.L. and H.K. analysed the results, D.L. and H.K. wrote the main manuscript text, D.L. prepared all figures, H.C. supervised and advised. All authors reviewed the manuscript.

## Data availability

Fishyscapes Lost&Found and Fishyscapes Static benchmark datasets that support the findings of this study are publicly available from The Fishyscapes Benchmark website <https://fishyscapes.com/dataset>. The source code for reproducibility and raw data files generated and/or analysed during the current study are available in the github repository, <https://github.com/hagg30/LogitVar>.