2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

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

Dongkun Lee[1], Han-Gyu Kim[2] and Ho-Jin Choi[1]

***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. Anomaly segmentation approaches implemented with extra requirements such as out-of-domain data, extra network, or network retraining might increase the computational cost or degradation of segmentation performance. In this study, 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 iterative background highlighting is proposed to robustly detect anomalous objects on the background, 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 the Fishyscapes Lost&Found and 17.4% on the Fishyscapes Static compared to the state-of-the-art method. The code of this work is available at our Github repository (https://github.com/hagg30/LogitVar).**

I. 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 were conducted in various domains, such as video analysis [1], [2], [3]. Anomaly segmentation, which is an advanced type of anomaly detection that specifies the anomaly region in a given image, is introduced for sophisticated and safety-critical applications such as autonomous driving. In urban scene segmentation, anomalies are objects or areas that are not included in the dataset, which can cause critical safety issues. A self-driving car that does not recognize an anomalous object may result in severe accidents, such as roadkill or vehicle damage. Recent studies on anomaly segmentation of urban landscapes [4], [5], [6], [7], [8], [9] have shown moderate performance on seen images. However, the models are easily confused by images containing unseen objects, whereas 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. [10], [11] use laborious human intervention such as extra data with anomaly labels or out-of-domain (OoD) data which is outside of the scope of the training data for pixel-wise anomaly detection. However, it is difficult to collect sufficient anomalous data. In addition, the approaches strongly dependent on human intuition. [12], [13], [14], [15], [16] adopted an 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 networks, the maximum softmax probability (MSP) and max logits, which are the maximum values within the classes prior to the final softmax layer, were adopted with a simple assumption that the MSP or max logits of anomaly regions are lower than those of normal regions [17], [18]. Among these methods, standardized max logits (SML) [19] standardize max logits for robustness on OoD objects and outperform other logit-based anomaly segmentation methods.

However, the SML has two limitations. First, its performance highly depends on hyper-parameters, which include the sample mean and variance of the logits of each class [19]. Second, 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.

To better utilize the information included in the output of the segmentation network, we propose using pixel-wise logit variance (LOV) together instead of only using the max logits. For normal regions, the segmentation network outputs high logits for a specific class confidently, leading to a high pixel-wise LOV. For anomaly regions, the segmentation network is confused; therefore, it may output relatively uniform logits for all classes, resulting in a small pixel-wise LOV. Compared with SML, pixel-wise LOV is more robust when the segmentation network outputs uniformly high logits for all classes. Moreover, adopting pixel-wise LOV could reduce the degree of dependency on the hyper-parameters of SML, as computing variance does not require any hyper-parameters.

Additionally, we propose iterative background highlighting (IBH) to enhance anomaly objects on the background because meaningful anomalies are more likely to occur in background areas such as roads an autonomous driving scenario. To demonstrate the performance difference, we visualized the performance difference using 100,000 pixel samples per class using a box plot, as shown Figure 2. The figure shows a significantly higher anomaly segmentation performance on the Fishyscapes validation dataset than

*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]

[1]Dongkun Lee and Ho-Jin Choi are with School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Republic of Korea. {hagg30, hojinc}@kaist.ac.kr

[2]Han-Gyu Kim with NAVER Cloud, Gyeonggi-do, 13529, Republic of Korea. hangyu.kim@navercorp.com

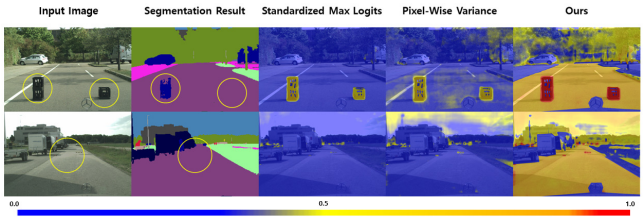

Fig. 1. Examples of our anomaly segmentation method. Yellow circle indicates location of anomalous object. When an image with anomalous object is used as input, there exist incorrectly classified pixels after semantic segmentation. Except for conventional standardized max logits, our approach also adopts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

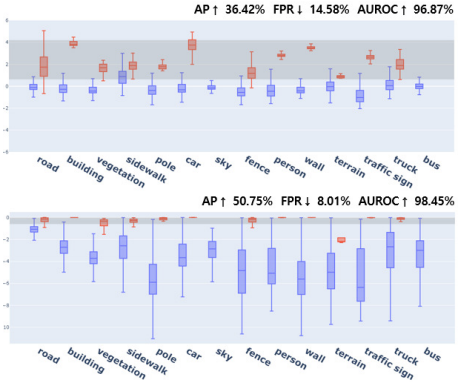

Fig. 2. Box plot of anomaly score comparison between SML (up) and our method (down) on Fishyscapes Lost&Found validation dataset. We took up to 100,000 samples from each class. X-axis represents training classes sorted by the appearance frequency in training data. Y-axis represents the anomaly score (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.

the previous state-of-the-art. Additionally, we significantly reduced false-positive pixels for realistic scene input. Figure 1 compares our approach to SML using two examples, indicating that our approach detects anomalous objects better than the SML.

The contributions of our study are summarized as follows:

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

II. RELATED WORK

A. Deep anomaly segmentation

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

B. Unsupervised anomaly detection

To overcome the data-insufficiency problem, unsupervised anomaly detection has been developed. In unsupervised anomaly detection, a neural model can understand the distribution of data solely based on the provided normal data without relying on any user-tagged input. The collection of a normal image dataset is relatively inexpensive. Its primary tasks are generative anomaly detection [22], [23], time-series modeling [24], [25], anomaly classification [26], [27], novelty detection [28], and anomaly segmentation [29], [30], [31], [32], [33]. However, previous approaches require a considerable amount of anomaly data. Therefore, it is difficult to apply the unsupervised method to anomaly segmentation, where the model should distinguish abnormal regions from the local image without supervision, which is more complex than the two-class classification task.

C. Anomaly segmentation in an urban landscape scene

One of the most popular applications of anomaly segmentation is autonomous driving, where anomaly segmentation is conducted on an urban landscape scene. [10], [11], [13], [14], [16], [34], [35] leverage the existing segmentation network to improve the recognition of anomalous objects on an urban landscape scene. Among these methods, [10], [14], [35], [36], [37] required extra training of the segmentation network and additional networks. Adopting an additional network or retraining the segmentation network increases the time complexity, whereas autonomous driving requires real-time abnormal segmentation as it is a safety-critical task. [11], [16], [34], [36], [37] utilized OoD data. However, using OoD data is impractical because OoD cases rarely occur in a practical environment.

D. Standardized maximum logits

Standardized maximum logits (SML) [19] 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 attempts to project all distributions to the same scale by standardizing logits; thus, 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 the pixels of

training data. The standardization step of SML may be described as follows:

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

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

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

where μ_c and σ_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; 1 is the indicator function; $S_{h,w}$ denotes SML; In [19], SML was used to differentiate normal and abnormal pixels. Pixels with a large SML are classified as normal pixels and other pixels are classified as anomalous.

III. PROPOSED METHOD

Conventional SML-based anomaly segmentation has two limitations. First, the SML uses the sample mean μ_c in Equation 1 and sample variance σ_c^2 in Equation 2 as the hyper-parameters in the inference stage. We established that the performance of final anomaly segmentation is highly related to the hyper-parameters, implying that the performance might be poor if the hyper-parameters do not fit the realistic data. The most extreme and intuitive case where this can be discovered is when sample points are not available for sample extraction. In such cases, the class-specific sample mean/variance in the SML is undefined, leading to immediate malfunction. In addition, if an alternative segmentation network is used instead of the segmentation network in [19], the hyper-parameters computed according to Equations 1 and 2 do not always guarantee the best performance. Second, classifying pixels with a small SML as an anomaly class does not always stand. The confused segmentation network is likely to emit uniformly distributed logits; however, it does not guarantee that the maximum logit is small because logits for all classes may evenly be large.

To overcome the above-mentioned problems, we propose use of pixel-wise LOV. Variance has several advantages compared to SML: computing variance does not require any hyper-parameter because it can be computed using only the logits emitted by the segmentation network during the inference; variance is a better criterion for determining whether the logits are uniformly distributed compared to SML, where a smaller variance implies a uniform distribution.

Instead of using variance or the SML independently, we propose the combination of variance and SML to improve the anomaly segmentation performance further. In addition, we propose IBH to detect anomalies on background more accurately because anomalies are more likely to occur in the background practical situation. Our method can be applied to any model that has a final linear classifier layer; thus it can be used for most urban landscape segmentation models and datasets where the background area is distinguished from areas other than the background on the network. An overview of the proposed method is shown in Figure 3. Detailed information of each module is explained in the subsections below.

The pixel-wise LOV is calculated using the logits of a pixel output by the segmentation network, which can be described according to 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 LOV 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. As shown in the equation, the pixel-wise LOV does not require any hyper-parameters. We adopted the LOV for anomaly segmentation with the assumption that the variance should be small for anomaly pixels compared to that of normal pixels because the segmentation network tends to emit more uniformly distributed logits for an anomaly pixel.

In the proposed method, we used an ensemble of SML and LOV to achieve better performance. Among various ensemble methods, we discovered that using the summation of SML and LOV shows the best performance. Additionally, we used iterative boundary suppression and dilated smoothing, which are the post-processing algorithms used in [19] with the same hyper-parameters for fair comparison. Thus, the base anomaly score may be computed according to the following equation:

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

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

A. Iterative background highlighting

In an autonomous driving scenario, detecting anomalous objects in the background, such as roads is important, because the background is the most frequent type of class to meet, and anomalous objects on the background usually endanger safety. To enhance the performance of anomaly segmentation on the background, we propose IBH. The approach exploits the last fully connected layer of a segmentation network to highlight the background.

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

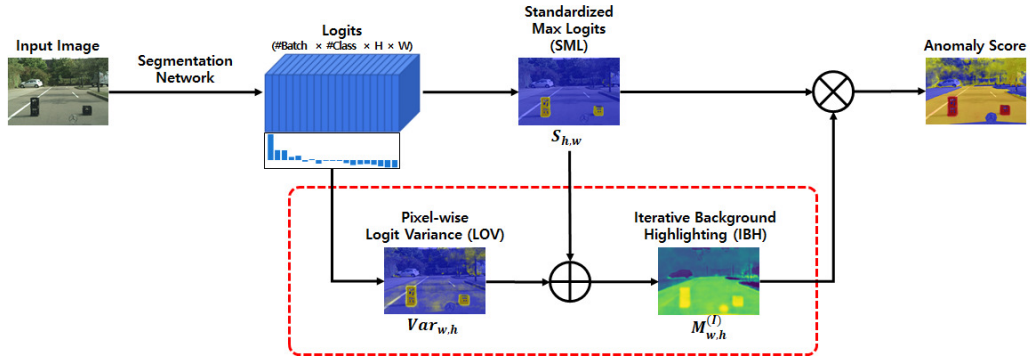

Fig. 3. Overview of the proposed approach. The logits are generated with the segmentation network. LOV and SML are obtained from the given logits. Base anomaly score is obtained with the summation of LOV and SML to complement each other. Subsequently, IBH highlights background pixels. The final anomaly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

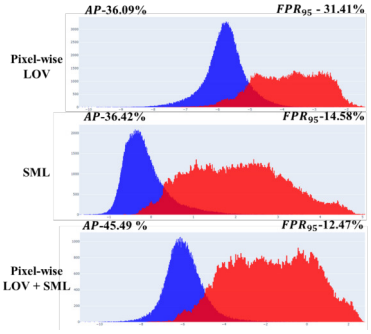

Fig. 4. Histogram of pixel-wise logit variance, standardized max logit and summation of standardized logits and class-wise variance (without iterative background highlighting) in Fishyscapes Lost&Found. X-axis denotes pixel-wise anomaly score. Red (anomaly) and blue (normal) are obtained from 400,000 (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.

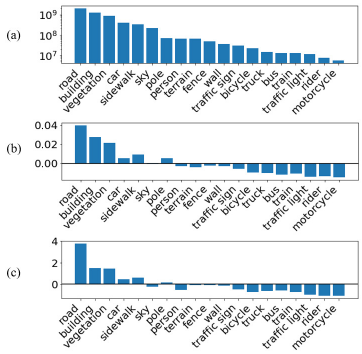

Fig. 5. Illustration for showing the correlation between logit and weight of linear classifier. The classes are sorted by appearance frequency in the training dataset. Each figure indicates: (a) Total number of appearances for each class, (b) Mean of the weight values of each linear classifier, and (c) Mean of the logit values for each class. To clarify, all class share a linear classifier, and the "Mean of the weight values of each linear classifier" refers to the average of weight values for each linear classifier class among channel depth, which is similar to most of standard segmentation networks.

Algorithm 1 Iterative Background Highlighting

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

of the linear classifier, to target pixels that could enhance the logits of the targets.

We designed a background highlighting algorithm as Algorithm 1. We used anomaly score for the initial blending mask to include the confident anomalies in the background highlighting target. During the iterative steps, the blending mask becomes larger because of the blending step, where pixels with a large blending mask, that belong to the background obtain a larger input for the linear classifier. Such iterations were is executed I times, for which we adopted 3 in our approach. Finally, the blending mask was multiplied with the base anomaly score to obtain the final anomaly score. The variations in the performance according to I is further provided in Table I. We report the performance of over 100 images on FS Lost&Found validation dataset. The highest AP score was observed when $I = 3$.

TABLE I
QUANTITATIVE RESULTS ACCORDING TO TO NUMBER OF ITERATIONS I.

Method	AP ↑	FPR_{95}↓	AUROC ↑
$I = 0$	45.49	12.47	97.34
$I = 1$	48.94	7.87	98.36
$I = 2$	50.44	**7.57**	**98.49**
I = 3	**50.75**	8.01	98.45
$I = 4$	49.71	8.90	98.31

IV. RESULTS AND DISCUSSIONS

A. Experimental setting

1) Dataset: We used two popular benchmark datasets for anomaly detection on the urban landscape scene: Fishyscapes

Lost&Found (FS Lost&Found) and Fishyscapes Static (FS Static) [13]. FS Lost&Found is a real-world dataset of road images with anomalous objects visible in the front sight of a vehicle. It contained 100 validation and 275 undisclosed test images. FS Static is a synthetic dataset with blended anomalous objects on PASCAL Visual Object Classes (PASCAL VOC) images. It contained 50 validation and 1000 undisclosed test images.

2) Implementation details: For a more direct comparison with SML, which is a previous state-of-the-art method, we mainly adopted a pre-trained model of DeepLabv3+/ResNet101 architecture of the official SML implementation [19] as the segmentation network. We also implemented our method on various segmentation networks using the mmsegmentation toolbox[1] to demonstrate the general superiority of our method.

3) Evaluation metric: The metrics provided by the FS benchmark including average precision (AP) and false positive rate at 95% true positive rate (FPR_{95}) are adopted for performance evaluation. FS benchmark suggests AP as the primary metric, because it is invariant under data imbalance and threshold. FPR_{95} is a meaningful metric for safety-critical applications.

B. Results

1) Comparison on the FS leaderboard: We compared our performance with that of anomaly segmentation methods on the Fishyscapes benchmark leaderboard until July 9, 2022. The methods with neither papers with detailed explanations nor publicly available source codes are not included in our comparison. The performance comparison result is summarized in Table II, which includes both AP and FPR_{95} for benchmark datasets FS Lost&Found and FS Static, respectively. Moreover, this table includes brief information for requirements of the compared methods–whether to use retraining, extra networks, or OoD data. As our method avoids extra requirements, MSP, entropy, kNN embedding, and standardized max logits are the main comparison targets. It is shown in Table II that ours outperforms any of the main comparison targets in the AP score which is the primary comparison metric. In addition, our method outperforms in FPR_{95} for the FS Lost&Found dataset. Compared with SML, which is our baseline study, the performance significantly improved by 6.76 AP (21.7%) for the FS Lost&Found and 9.28 AP (17.4%) for FS Static. Additionally, we compared our method to methods with extra requirements, which outperform our method because of extra benefits. However, our methods show competitive generalization ability across multiple datasets, ranking 3rd for the FS Lost&Found and 4th for the FS Static in AP score, and our method does not require any retraining, which preserves the segmentation accuracy.

2) 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 yields a significantly smaller false positive area on the real data. Our method generally showed significantly fewer false positive regions on real-world datasets than SML. However, our method did not work properly when the road surface had an exceptional pattern such as a sidewalk block or dirty surface.

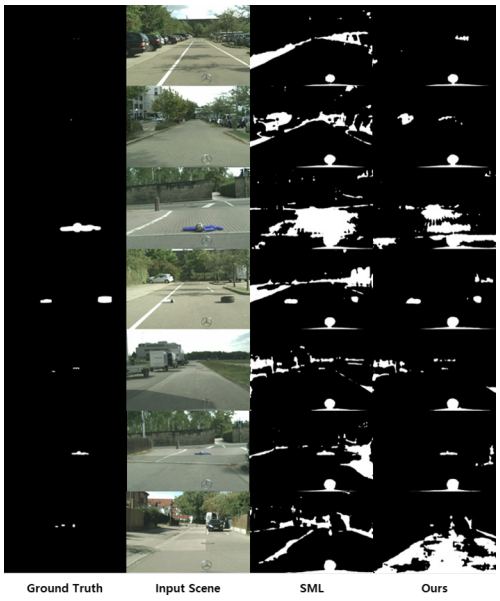

Fig. 6. Anomaly detected with TPR_{95} on Fishyscapes Lost&Found validation dataset. White region indicates anomaly.

TABLE III
ABLATION STUDY OF SML, LOV AND IBH.

Method	FS Lost&Found		FS Static	
	AP↑	**FPR_{95}↓**	**AP↑**	**FPR_{95}↓**
Full Framework	**50.75**	**8.01**	**54.06**	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	**16.79**

3) Ablation study: We conducted an ablation study to determine the contribution of each component of our method to the anomaly segmentation performance; the results are presented in Table III. The table shows that applying LOV to SML improves the performance. In addition, the LOV combined with IBH, 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 those of the other region.

4) Framework generalization: To show the superiority of our proposed method on various segmentation networks, we conducted anomaly segmentation experiments on other segmentation networks from the mmsegmentation toolbox instead of DeepLabv3+ with the ResNet101 backbone

[1]https://github.com/open-mmlab/mmsegmentation

TABLE II

ANOMALY SEGMENTATION PERFORMANCE REPORTED ON FISHYSCAPES LEADERBOARD. THE BEST PERFORMANCES ARE HIGHLIGHTED IN BOLD.

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

TABLE IV

ANOMALY SEGMENTATION RESULTS WITH VARIOUS SEGMENTATION NETWORKS ON FISHYSCAPES VALIDATION DATASET.

Segmentation Architecture	**Method**	FS Lost&Found		FS Static		**mIoU**
		AP↑	**FPR$_{95}$↓**	**AP↑**	**FPR$_{95}$↓**	
OCRNet [38]	SML	39.96	18.28	47.90	15.07	81.35
	LOV	48.40	16.61	47.67	21.67	
	Sum	**51.89**	12.65	**53.55**	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	**7.82**	53.54	**10.30**	
ISANet [39]	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	**40.18**	38.35	26.31	28.83	
	Sum + IBH	30.40	**14.42**	**36.46**	**11.76**	
DeepLabV3+ [40]	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	**18.07**	
	SML + IBH	1.16	**27.35**	7.66	20.53	
	LOV + IBH	**28.54**	50.79	**46.78**	64.54	
	Sum + IBH	13.23	34.68	22.89	49.08	
DeepLabV3 [41]	SML	24.58	31.45	39.05	**15.63**	80.10
	LOV	24.55	59.14	30.31	67.71	
	Sum	**30.57**	37.09	**39.82**	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	**30.28**	38.23	33.54	

network that provided by [19]; the results are presented in Table IV. Although the proposed method does not show significant improvement for $\texttt{FPR}_{95}$ for the FS Static dataset, it generally shows better AP for both datasets and better $\texttt{FPR}_{95}$ for the FS Lost&Found dataset, This indicates that the proposed method works robustly along with various types of segmentation networks. In particular, for the two networks that have the highest mIoU score, ISANet [39] and OCRNet [38], our methods exhibited superior performance than the SML.

V. CONCLUSION

Herein, we proposed a simple method pixel-wise LOV and IBH to aid in the unexpected behavior of segmentation networks triggered by anomalous objects on the road. This approach does not require an external datasets, additional training, or external network. We strengthened our method on two intuitions: the network's latent uncertainty is expressed through the variance value of the output value and a high logit implies more prior information. The experimental results indicate that our approach achieves a new state-of-the-art performance in the FS Lost&Found and FS Static benchmark. Additionally, extensive experiments on various segmentation networks demonstrate the superiority of our method over previous state-of-the-art methods. Our future work will focus on developing a general model that replaces both SML and conventional anomaly segmentation in urban landscape scenes.

ACKNOWLEDGMENT

We thank Hermann Blum and the FishyScapes team for providing the baseline and leaderboard platform.

REFERENCES

[1] T.-N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1273–1283.

[2] W. Luo, W. Liu, D. Lian, J. Tang, L. Duan, X. Peng, and S. Gao, "Video anomaly detection with sparse coding inspired deep neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[3] N. Li and F. Chang, "Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder," *Neurocomputing*, vol. 369, pp. 92–105, 2019.

[4] S. Choi, S. Jung, H. Yun, J. T. Kim, S. Kim, and J. Choo, "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*, 2021, pp. 11 580–11 590.

[5] S. Choi, J. T. Kim, and J. Choo, "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*, 2020, pp. 9373–9383.

[6] Y. Zhu, K. Sapra, F. A. Reda, K. J. Shih, S. Newsam, A. Tao, and B. Catanzaro, "Improving semantic segmentation via video propagation and label relaxation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8856–8865.

[7] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 593–602.

[8] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3146–3154.

[9] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Denseaspp for semantic segmentation in street scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*, 2018, pp. 3684–3692.

[10] P. Bevandić, I. Krešo, M. Oršić, and S. Šegvić, "Dense outlier detection and open-set recognition based on training with noisy negative images," *arXiv preprint arXiv:2101.09193*, 2021.

[11] R. Chan, M. Rottmann, and H. Gottschalk, "Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5128–5137.

[12] M. Grcić, P. Bevandić, and S. Šegvić, "Dense open-set recognition with synthetic outliers generated by real nvp," *arXiv preprint arXiv:2011.11094*, 2020.

[13] H. Blum, P.-E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena, "The fishyscapes benchmark: Measuring blind spots in semantic segmentation," *arXiv preprint arXiv:1904.03215*, 2019.

[14] 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.

[15] 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 *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1099–1106.

[16] 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*, 2021, pp. 16 918–16 927.

[17] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *arXiv preprint arXiv:1610.02136*, 2016.

[18] 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.

[19] 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.

[20] S. M. Erfani, M. Baktashmotlagh, M. Moshtaghi, V. Nguyen, C. Leckie, J. Bailey, and K. Ramamohanarao, "From shared subspaces to shared landmarks: A robust multi-source classification approach," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[21] V. Jumutc and J. A. Suykens, "Multi-class supervised novelty detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 12, pp. 2510–2523, 2014.

[22] W. Lawson, E. Bekele, and K. Sullivan, "Finding anomalies with generative adversarial networks for a patrolbot," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 12–13.

[23] A. Mishra, S. Krishna Reddy, A. Mittal, and H. A. Murthy, "A generative model for zero shot learning using conditional variational autoencoders," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 2188–2196.

[24] P. Dasigi and E. Hovy, "Modeling newswire events using neural networks for anomaly detection," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 1414–1422.

[25] P. Filonov, F. Kitashov, and A. Lavrentyev, "RNN-based early cyber-attack detection for the tennessee eastman process," *arXiv preprint arXiv:1709.02232*, 2017.

[26] E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, M. Prabhat, and C. Pal, "ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events," in *Advances in Neural Information Processing Systems*, 2017, pp. 3402–3413.

[27] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5450–5463, 2019.

[28] M. Kliger and S. Fleishman, "Novelty detection with GAN," *arXiv preprint arXiv:1802.10560*, 2018.

[29] S. Zhao, J. Song, and S. Ermon, "Towards deeper understanding of variational autoencoding models," *arXiv preprint arXiv:1702.08658*, 2017.

[30] D. Dehaene, O. Frigo, S. Combrexelle, and P. Eline, "Iterative energy-based projection on a normal data manifold for anomaly localization," in *International Conference on Learning Representations*, 2019.

[31] L. Zhou, W. Deng, and X. Wu, "Unsupervised anomaly localization using VAE and beta-VAE," *arXiv preprint arXiv:2005.10686*, 2020.

[32] W. Liu, R. Li, M. Zheng, S. Karanam, Z. Wu, B. Bhanu, R. J. Radke, and O. Camps, "Towards visually explaining variational autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8642–8651.

[33] D. Dehaene and P. Eline, "Anomaly localization by modeling perceptual features," *arXiv preprint arXiv:2008.05369*, 2020.

[34] Y. Tian, Y. Liu, G. Pang, F. Liu, Y. Chen, and G. Carneiro, "Pixel-wise energy-biased abstention learning for anomaly segmentation on complex urban driving scenes," *arXiv preprint arXiv:2111.12264*, 2021.

[35] J. Mukhoti and Y. Gal, "Evaluating bayesian deep learning methods for semantic segmentation," *arXiv preprint arXiv:1811.12709*, 2018.

[36] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks," *arXiv preprint arXiv:1802.10501*, 2018.

[37] P. Bevandić, I. Krešo, M. Oršić, and S. Šegvić, "Simultaneous semantic segmentation and outlier detection in presence of domain shift," in *German Conference on Pattern Recognition*. Springer, 2019, pp. 33–47.

[38] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 173–190.

[39] L. Huang, Y. Yuan, J. Guo, C. Zhang, X. Chen, and J. Wang, "Interlaced sparse self-attention for semantic segmentation," *arXiv preprint arXiv:1907.12273*, 2019.

[40] 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 (ECCV)*, 2018, pp. 801–818.

[41] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.