Supplementary Material For Self-supervised Likelihood Estimation with Energy Guidance for Anomaly Segmentation in Urban Scenes

Yuanpeng Tu[1*], Yuxi Li[2*], Boshen Zhang[2*], Liang Liu[2], Jiangning Zhang[2], Yabiao Wang[2†], Cai Rong Zhao[1†]

[1] Dept. of Electronic and Information Engineering, Tongji Univeristy, Shanghai
[2] YouTu Lab, Tencent, Shanghai

More Experimental Results

Results on Lost & Found. To make more comprehensive comparison, we also evaluate our method on the old Lost & Found dataset (Pinggera et al. 2016), where 13 challenging real-world scenes are included with 37 different anomalous instances, which is the first publicly available urban anomaly segmentation datasets. Specifically, the obstacles in this dataset vary significantly in size and material. And following the official protocol (Pinggera et al. 2016), 1,203 images collected from 112 video stereo sequences with a resolution of 2048×1024 are utilized as the test set. Similar to the setting on FS Lost & Found, the segmentation model of DeepLab series (Chen et al. 2018) with ResNet101 pretrained on Cityscapes is used and fixed without re-training on this dataset. Finally, the same OoD head and evaluation metrics are adopted as FS Lost & Found.

Tab. 2 shows the performance of the proposed SLEEG on the test set of Lost & Found. Notably, our method surpasses all previous approaches and achieves SOTA performance on AP. Specifically, though previous SOTA methods DenseHybrid (Grcić, Bevandić, and Šegvić 2022) and PEBAL (Tian et al. 2021) utilize auxiliary data and are re-trained with more complex WideResNet38 as backbone, our SLEEG still brings a relative improvement of 4.59% and 4.18% to them on AP respectively with ResNet101 as backbone and requires no auxiliary OoD samples or further re-training of the segmentation model. Furthermore, when compared with the SOTA method Energy (Liu et al. 2020a) that falls into the same category as SLEEG, our method surpasses it by a large gap of 16.51% on AP and meanwhile significantly reduces the false positive pixels from 15.69% to 4.6%. This results fully illustrate that SLEEG shows great robustness and effectiveness on localising various real-world unexpected objects. Moreover, SML (Jung et al. 2021) attempts to tackle this task by re-balance the class-wise discrepancy of inlier samples. However, since there are only two classes in Lost & Found, there exists large performance gap between the performance of SML on Fishyscapes and Lost & Found, implying that SML may not be suitable for real-world applications. By contrast, it is worth noting that our SLEEG is capable of achieving consistent performance boost on all these datasets, demonstrating its great potential in real world scenarios.

Results for OoD data replacement. Meanwhile, we also separately investigate our patch policy with other method (Grcić, Bevandić, and Šegvić 2022) by replacing the external OoD data used in original paper with our generated patches. The results can be illustrated in Tab. 3, where the results of Densehybrid+OoD data is borrowed from the original paper, and the other results are reimplemented by ourselves. It can be observed that although applying less data, our patch policy can achieve results close to those in the original paper, this further demonstrate that our scheme can be viewed as a potential replacement for OoD data relied by previous methods.

Parameter Sensitivity. Finally we investigate the influence of the balance factor λ on the FS Lost & Found validation set and FS Static validation set. As shown in Fig. 2, SLEEG performs best when the balance factor $\lambda = 0.5$. The results demonstrate that SLEEG achieves similar performance across all settings, implying SLEEG is not very sensitive to λ. Finally, we set $\lambda = 0.5$.

Qualitative Results

We also conduct further comparison with the SOTA method DenseHybrid on Lost & Found test set and FS Lost & Found validation set. As shown in Fig. 4, for the challenging small anomalous objects at a distance, SLEEG produces larger anomaly scores, allowing vehicles respond to prevent accidents in a more timely manner. Besides, for objects with diverse shapes, our SLEEG is capable of producing a more complete contour covering the entire anomalous object as shown in the third row of Fig. 4.

Fig. 6 further visualizes the comparison between JEM and our SLEEG on Lost & Found test set. Notably, SLEEG can consistently perform better than JEM when detecting objects with various scales and distances. Beside JEM, Additionally, Fig. 6 and Fig. 5 describe the comparison of visualization results between the classic softmax entropy (Hendrycks and Gimpel 2017) and image re-synthesis (Lis et al. 2019a) methods and our SLEEG. It can be observed that these methods perform much worse than our SLEEG and are unable to effectively distinguish in-distribution objects from anomalous ones. And there even exists large background regions

[*] These authors contributed equally.
[†] Corresponding author.

Architecture	Method	FS LAF Val		FS Static Val		mIoU
		$\mathrm{FPR}_{95}\downarrow$	AP $\uparrow$	$\mathrm{FPR}_{95}\downarrow$	AP $\uparrow$	
OCRNet (Yuan and Wang 2020)	SML	18.28	39.96	15.07	47.90	80.66
	JEM	22.90	23.27	16.80	34.03	
	Void Class	16.65	46.62	17.74	29.30	
	SLEEG	**8.89**	**72.51**	**7.6**	**73.01**	
ISANet (Huang et al. 2019)	SML	18.67	28.76	14.86	32.15	80.51
	JEM	35.57	22.65	16.22	45.22	
	Void Class	29.17	43.53	17.18	24.61	
	SLEEG	**12.28**	**65.79**	**3.33**	**80.03**	
FCN (Shelhamer 2017)	SML	39.80	17.59	14.53	28.92	77.72
	JEM	39.36	21.83	13.47	30.51	
	Void Class	24.29	43.44	14.79	22.72	
	SLEEG	**21.01**	**63.11**	**5.31**	**61.29**	

Table 1: Comparison between SLEEG and other anomaly detection methods on FS validation sets with different segmentation models. ”Void Class” denotes training models with pixels that fall into void class as anomaly samples.

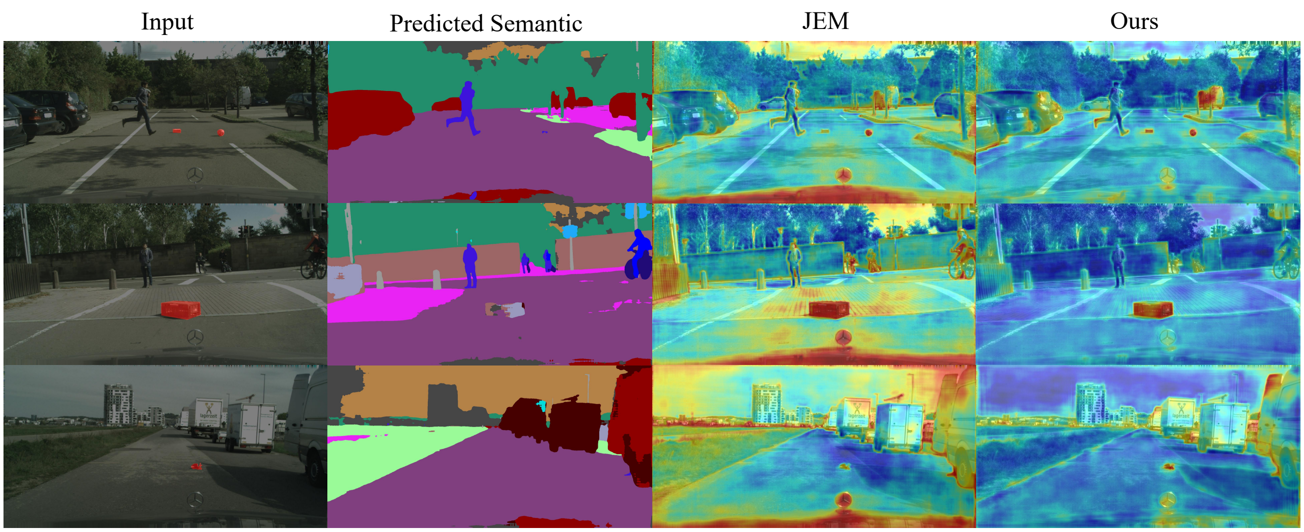

Figure 1: Visualization of on FS Lost & Found validation set. Compared with JEM, predictions from our SLEEG show anomaly maps with higher responses for anomalous instances and lower values for normal pixels.

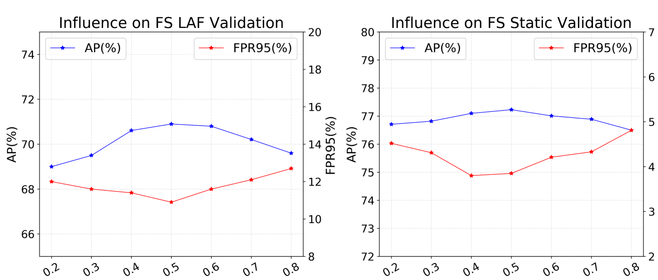

Figure 2: Investigation on the influence on AP and false positive rate with varied λ value on FS Lost & Found validation set (left) and FS Static validation set (right).

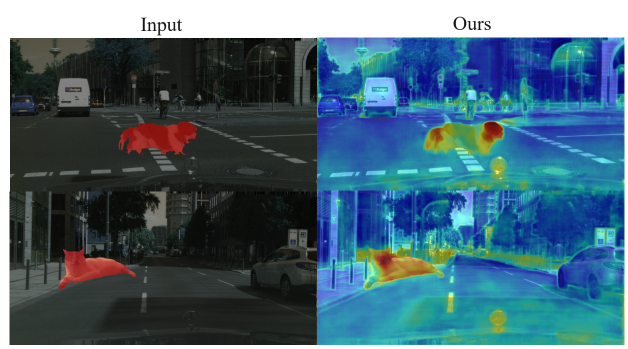

Figure 3: Visualization results of some failure cases from SLEEG on FS Static validation set.

that are assigned with higher values than anomalous objects. By contrast, our SLEEG generates anomaly maps where unexpected instances are intensively assigned with higher scores, resulting in much fewer false positive pixels as well.

Limitations

As shown in the Table 1 of the manuscript and Fig. 3, our SLEEG performs inferior to SOTA methods that utilize

Methods	OoD Data	Re-training	AUC ↑	AP ↑	FPR_{95} ↓
Meta-OoD (Chan, Rottmann, and Gottschalk 2021)	✓	✓	97.95	71.23	5.95
SynBoost† (Di Biase et al. 2021)	✓	✗	98.38	70.43	4.89
Deep Gambler† (Liu et al. 2019)	✓	✓	98.67	72.73	3.81
PEBAL† (Di Biase et al. 2021)	✓	✓	**99.76**	<u>78.29</u>	**0.81**
DenseHybrid† (Grcić, Bevandić, and Šegvić 2022)	✓	✓	<u>99.37</u>	**78.70**	<u>2.10</u>
MSP (Hendrycks et al. 2019a)	✗	✗	85.49	38.20	18.56
Mahalanobis (Lee et al. 2018a)	✗	✗	79.53	42.56	24.51
Max Logit (Hendrycks and Gimpel 2016)	✗	✗	<u>94.52</u>	65.45	<u>15.56</u>
Entropy (Hendrycks and Gimpel 2016)	✗	✗	86.52	50.66	16.95
SML† (Jung et al. 2021)	✗	✗	88.05	25.89	44.48
Energy (Liu et al. 2020a)	✗	✗	94.45	<u>66.37</u>	15.69
SLEEG	✗	✗	**98.59**	**82.88**	**4.60**

Table 2: Comparison on **Lost & Found** testing set. All methods use the same segmentation models. † indicates training with WideResNet38 backbone. **Bold** values and <u>underlined</u> values represent the best and second best results.

Method	FS LAF Validation		
	FPR_{95} ↓	AP ↑	AUROC ↑
DenseHybrid + OoD data (Grcić, Bevandić, and Šegvić 2022)	6.1	63.8	-
DenseHybrid + Convex patch	22.2	56.2	96.6
DenseHybrid + Refined convex patch	18.6	60.5	97.0
SLEEG + Refined convex patch	10.9	70.9	98.3

Table 3: Investigation when combining other state-of-the-art method DenseHybrid (Grcić, Bevandić, and Šegvić 2022) with our strategy of patch generation on FS LAF validation set.

auxiliary OoD data on the FS Static dataset. This mainly accounts for that the synthetic anomalous objects in FS Static are similar to the instances in the auxiliary dataset that these methods use, i.e., COCO (Lin et al. 2014) and ADE20K (Zhou et al. 2019). Therefore, training with such OoD samples allows models foresee the anomalous objects, resulting in their good performance. Moreover, training in this manner also results in a large gap between these methods and ours in performance on the FS Lost & Found and Lost & Found datasets since anomalous objects in the FS Lost & Found dataset are more realistic. Finally, though our SLEEG performs inferior on FS Static dataset, since the anomaly segmentation task targets at real-world automatic driving scenarios, the ability of tackling realistic unexpected samples is more essential for our SLEEG. Therefore, we focus on utilizing the inherent spatial context to improve the ability of tackling with the distribution discrepancy of normal and anomalies without requiring extra OoD data.

References

Bevandic, P.; Kreso, I.; Orsic, M.; and Segvic, S. ???? Simultaneous Semantic Segmentation and Outlier Detection in Presence of Domain Shift. In *41st DAGM German Conference, DAGM GCPR 2019*.

Bevandic, P.; Kreso, I.; Orsic, M.; and Segvic, S. 2021. Dense outlier detection and open-set recognition based on training with noisy negative images. *CoRR*, abs/2101.09193.

Biase, G. D.; Blum, H.; Siegwart, R.; and Cadena, C. 2021. Pixel-Wise Anomaly Detection in Complex Driving Scenes. In *Computer Vision and Pattern Recognition, CVPR*.

Blum, H.; Sarlin, P.-E.; Nieto, J.; Siegwart, R.; and Cadena, C. 2021. The Fishyscapes Benchmark: Measuring Blind Spots in Semantic Segmentation. *International Journal of Computer Vision*, 129(11): 3119–3135.

Cen, J.; Yun, P.; Cai, J.; Wang, M. Y.; and Liu, M. 2021. Deep Metric Learning for Open World Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 15333–15342.

Chan, R.; Lis, K.; Uhlemeyer, S.; Blum, H.; Honari, S.; Siegwart, R.; Salzmann, M.; Fua, P.; and Rottmann, M. 2021. SegmentMeIfYouCan: A Benchmark for Anomaly Segmentation. *CoRR*, abs/2104.14812.

Chan, R.; Rottmann, M.; and Gottschalk, H. 2021. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 5128–5137.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proc. European Conference on Computer Vision (ECCV)*, 801–818.

Di Biase, G.; Blum, H.; Siegwart, R.; and Cadena, C. 2021. Pixel-wise anomaly detection in complex driving scenes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 16918–16927.

Table 4: Performance evaluation on the SMIYC benchmark (Chan et al. 2021).

Method	OoD Data	Re-training	AnomalyTrack FPR$_{95}$ ↓	AP ↑
SynBoost (Biase et al. 2021)	✓	✗	61.9	56.4
JSRNet (Vojir et al. 2021)	✗	✓	43.9	33.6
Void Classifier (Blum et al. 2021)	✓	✓	63.5	36.6
DenseHybrid (Grcić, Bevandić, and Šegvić 2022)	✓	✓	**9.8**	**78.0**
PEBAL (Tian et al. 2021)	✓	✓	40.8	49.1
Image Resyn. (Lis et al. 2019b)	✗	✗	**25.9**	52.3
Embed. Dens. (Blum et al. 2021)	✗	✗	70.8	37.5
ODIN (Liang, Li, and Srikant)	✗	✗	71.7	33.1
MC Dropout (Kendall and Gal 2017)	✗	✗	69.5	28.9
Max softmax (Hendrycks and Gimpel)	✗	✗	72.1	28.0
Mahalanobis (Lee et al. 2018b)	✗	✗	87.0	20.0
SLEEG (ours)	✗	✗	36.4	**52.7**

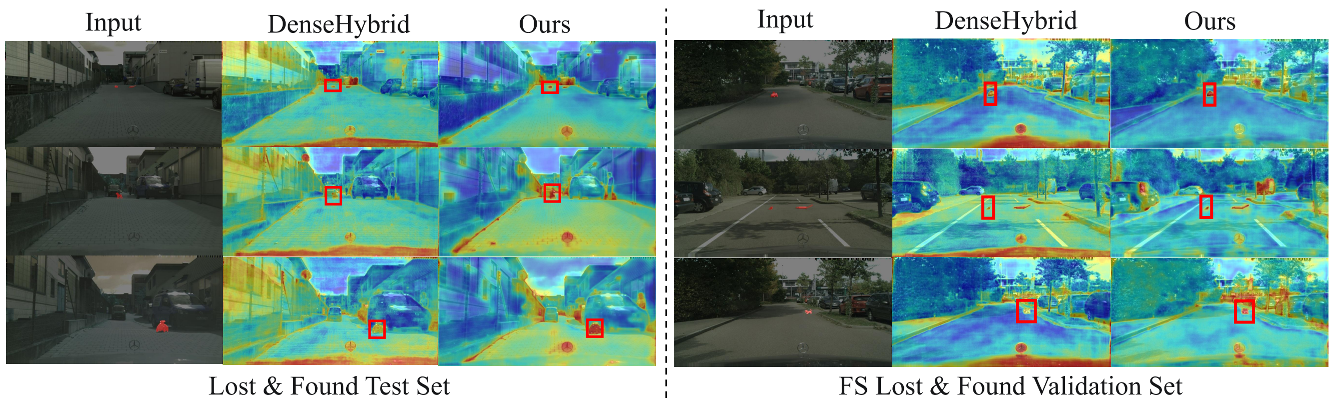

Figure 4: Comparison of visualization results with Densehybrid on Lost & Found test set and FS Lost & Found validation set. The proposed SLEEG is capable of localizing small anomalous objects more accurately. For example, SLEEG generates higher anomaly scores for the instances in the red box of the second and third rows.

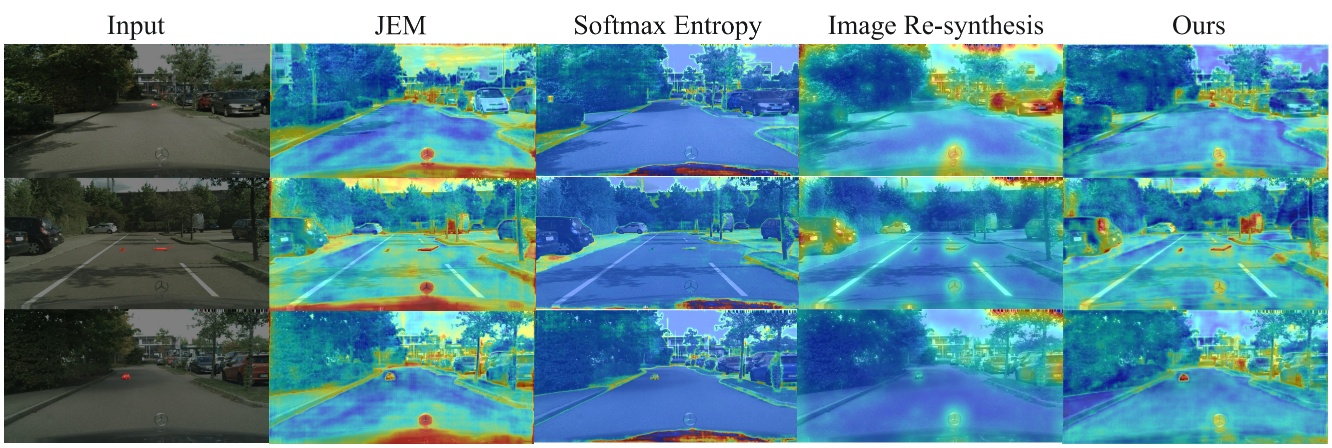

Figure 5: Comparison of visualization results with JEM, Softmax Entropy and Image Re-synthesis on FS Lost & Found validation set.

Franchi, G.; Bursuc, A.; Aldea, E.; Dubuisson, S.; and Bloch, I. 2020a. One Versus all for deep Neural Network Incertitude (OVNNI) quantification. *CoRR*, abs/2006.00954.

Franchi, G.; Bursuc, A.; Aldea, E.; Dubuisson, S.; and Bloch, I. 2020b. TRADI: Tracking Deep Neural Network Weight Distributions. In *16th European Conference on Computer Vision, ECCV*.

Grcić, M.; Bevandić, P.; and Šegvić, S. 2022. DenseHybrid: Hybrid Anomaly Detection for Dense Open-set Recognition. *arXiv preprint arXiv:2207.02606*.

Table 5: Performance evaluation on StreetHazards (Hendrycks et al. 2019b). SLEEG achieves competitive anomaly detection performance.

Method	OoD Data	Re-training	Anomaly detection		
			FPR$_{95}$ ↓	AUROC ↑	AP ↑
Energy (Liu et al. 2020b)	✓	✓	18.2	93.0	12.9
Outlier Exposure (Hendrycks, Mazeika, and Dietterich)	✓	✓	17.7	94.0	14.6
OOD-Head (Bevandic et al.)	✓	✓	56.2	88.8	19.7
OH-MSP (Bevandic et al. 2021)	✓	✓	30.9	89.7	18.8
DenseHybrid (Grcić, Bevandić, and Šegvić 2022)	✓	✓	13.0	95.6	30.2
CoroCL (Liu et al. 2023)	✓	✓	**8.2**	**97.2**	**31.2**
SynthCP (Xia et al. 2020)	✗	✗	28.4	88.5	9.3
Dropout (Kendall and Gal 2017)(Xia et al. 2020)	✗	✗	79.4	69.9	7.5
TRADI (Franchi et al. 2020b)	✗	✗	25.3	89.2	7.2
OVNNI (Franchi et al. 2020a)	✗	✗	22.2	91.2	12.6
SO+H (Grcić, Bevandić, and Šegvić)	✗	✗	25.2	91.7	12.7
DML (Cen et al. 2021)	✗	✗	17.3	93.7	14.7
MSP (Hendrycks and Gimpel)	✗	✗	27.9	90.1	7.5
ML (Hendrycks et al. 2019c)	✗	✗	22.5	92.4	11.6
ODIN (Liang, Li, and Srikant)	✗	✗	28.7	90.0	7.0
ReAct (Sun, Guo, and Li 2021)	✗	✗	21.2	92.3	10.9
SLEEG (ours)	✗	✗	**13.3**	**95.7**	**27.6**

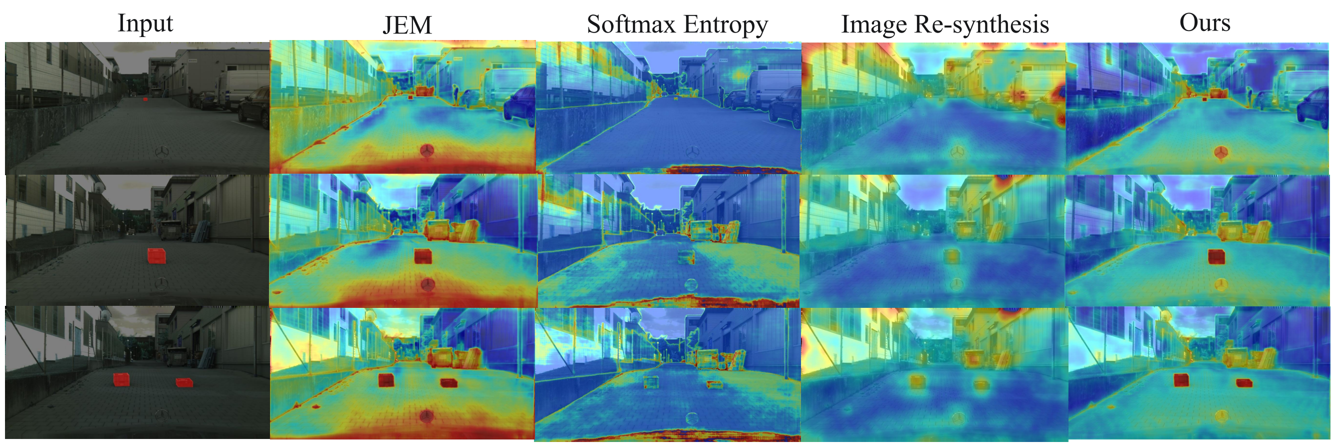

Figure 6: Comparison of visualization results with JEM, Softmax Entropy and Image Re-synthesis on Lost & Found test set.

Grcić, M.; Bevandić, P.; and Šegvić, S. ???? Dense Open-set Recognition with Synthetic Outliers Generated by Real NVP. In *16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2021*.

Hendrycks, D.; Basart, S.; Mazeika, M.; Mostajabi, M.; Steinhardt, J.; and Song, D. 2019a. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*.

Hendrycks, D.; Basart, S.; Mazeika, M.; Mostajabi, M.; Steinhardt, J.; and Song, D. 2019b. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*.

Hendrycks, D.; Basart, S.; Mazeika, M.; Mostajabi, M.; Steinhardt, J.; and Song, D. 2019c. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*.

Hendrycks, D.; and Gimpel, K. ???? A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *5th International Conference on Learning Representations, ICLR 2017*.

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

Hendrycks, D.; and Gimpel, K. 2017. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. *Proc. International Conference on Learning Representations (ICLR)*.

Hendrycks, D.; Mazeika, M.; and Dietterich, T. G. ???? Deep Anomaly Detection with Outlier Exposure. In *7th In-*

ternational Conference on Learning Representations, ICLR 2019.

Huang, L.; Yuan, Y.; Guo, J.; Zhang, C.; Chen, X.; and Wang, J. 2019. Interlaced Sparse Self-Attention for Semantic Segmentation. *arXiv preprint arXiv:1907.12273*.

Jung, S.; Lee, J.; Gwak, D.; Choi, S.; and Choo, J. 2021. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 15425–15434.

Kendall, A.; and Gal, Y. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In *Neural Information Processing Systems*.

Lee, K.; Lee, K.; Lee, H.; and Shin, J. 2018a. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 31.

Lee, K.; Lee, K.; Lee, H.; and Shin, J. 2018b. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. In *Neural Information Processing Systems, NeurIPS*.

Liang, S.; Li, Y.; and Srikant, R. ???? Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks. In *6th International Conference on Learning Representations, ICLR 2018*.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Proc. European Conference on Computer Vision (ECCV)*, 740–755. Springer.

Lis, K.; Nakka, K.; Fua, P.; and Salzmann, M. 2019a. Detecting the Unexpected via Image Resynthesis. In *Proc. IEEE International Conference on Computer Vision (ICCV)*.

Lis, K.; Nakka, K. K.; Fua, P.; and Salzmann, M. 2019b. Detecting the Unexpected via Image Resynthesis. In *International Conference on Computer Vision, ICCV*.

Liu, W.; Wang, X.; Owens, J.; and Li, Y. 2020a. Energy-based out-of-distribution detection. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 33: 21464–21475.

Liu, W.; Wang, X.; Owens, J. D.; and Li, Y. 2020b. Energy-based Out-of-distribution Detection. In *NeurIPS*.

Liu, Y.; Ding, C.; Tian, Y.; Pang, G.; Belagiannis, V.; Reid, I.; and Carneiro, G. 2023. Residual Pattern Learning for Pixel-wise Out-of-Distribution Detection in Semantic Segmentation. arXiv:2211.14512.

Liu, Z.; Wang, Z.; Liang, P. P.; Salakhutdinov, R. R.; Morency, L.-P.; and Ueda, M. 2019. Deep gamblers: Learning to abstain with portfolio theory. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 32.

Pinggera, P.; Ramos, S.; Gehrig, S.; Franke, U.; Rother, C.; and Mester, R. 2016. 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.

Shelhamer. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(4): 640–651.

Sun, Y.; Guo, C.; and Li, Y. 2021. ReAct: Out-of-distribution Detection With Rectified Activations. In *NeurIPS*.

Tian, Y.; Liu, Y.; Pang, G.; Liu, F.; Chen, Y.; and Carneiro, G. 2021. Pixel-wise Energy-biased Abstention Learning for Anomaly Segmentation on Complex Urban Driving Scenes. *arXiv preprint arXiv:2111.12264*.

Vojir, T.; Šipka, T.; Aljundi, R.; Chumerin, N.; Reino, D. O.; and Matas, J. 2021. Road Anomaly Detection by Partial Image Reconstruction With Segmentation Coupling. In *International Conference on Computer Vision, ICCV*.

Xia, Y.; Zhang, Y.; Liu, F.; Shen, W.; and Yuille, A. L. 2020. Synthesize Then Compare: Detecting Failures and Anomalies for Semantic Segmentation. In *16th European Conference on Computer Vision, ECCV*.

Yuan, Y.; and Wang, J. 2020. Object-Contextual Representations for Semantic Segmentation.

Zhou, B.; Zhao, H.; Puig, X.; Xiao, T.; Fidler, S.; Barriuso, A.; and Torralba, A. 2019. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 127(3): 302–321.