Supplementary for Pixel-wise Energy-biased Abstention Learning for Anomaly Segmentation on Complex Urban Driving Scenes

Yu Tian[1⋆], Yuyuan Liu[1⋆], Guansong Pang[2], Fengbei Liu[1], Yuanhong Chen[1], and Gustavo Carneiro[1]

[1] Australian Institute for Machine Learning, University of Adelaide
[2] Singapore Management University

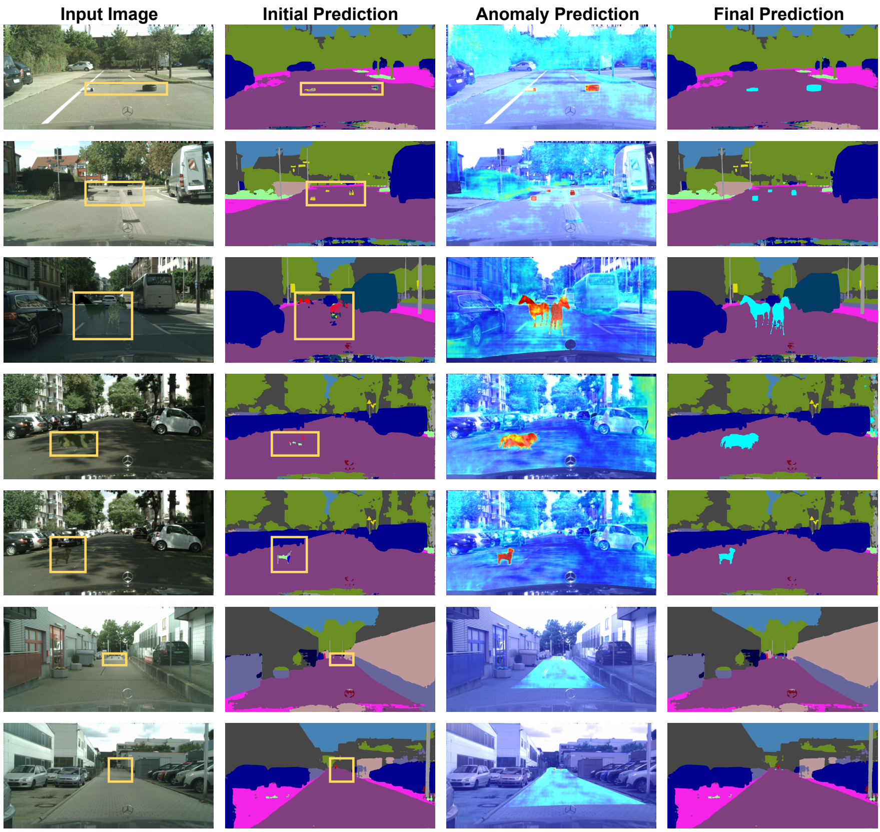

Fig. 1: From the **input image** (anomaly highlighted with a yellow box), the **initial prediction** shows the original segmentation results with anomalies classified as a one of the pre-defined inlier classes. **Anomaly predictions** from **our** method show an anomaly map with high scores (in yellow and red) for anomalous pixels. In our **final prediction**, anomalous pixels are coloured in cyan.

⋆ First two authors contributed equally to this work. GP is the corresponding author.

1 Qualitative results

In Figure 1, we show some additional qualitative results. Our approach can effectively detect small and distant objects (rows 6 and 7) and objects with different scales (rows 1 to 5).

2 More AUC results

In Tables 1 and 2, we show the AUC results in addition to the AP and FPR results in Tables 6 and 7 of the main paper. We achieve consistently SOTA AUC performance regardless of the selection of outlier classes or the number of outlier training samples.

Class Per.	FS LF - AUC	FS Static - AUC
1%	97.59 ±0.39	98.37 ±0.56
5%	98.17 ±0.45	98.25 ±0.71
10%	98.47 ±0.39	99.59 ±0.25
25%	98.39 ±0.28	99.52 ±0.17
50%	98.63 ±0.07	99.54 ±0.08
75%	98.71 ±0.05	99.59 ±0.03

Table 1: AUC testing results (mean results over six random seeds) of our approach on Fishyscapes benchmark w.r.t. different **diversity of OE classes**.

Train Size	FS LF - AUC	FS Static - AUC
5%	98.13 ±0.12	99.16 ±0.09
10%	98.35 ±0.15	99.57 ±0.07
25%	98.36 ±0.06	99.51 ±0.06
50%	98.69 ±0.05	99.37 ±0.07

Table 2: AUC testing results (mean results over six random seeds) of our approach on Fishyscapes benchmark w.r.t. different **amount of OE training samples**.

3 Hyper-parameters Selection

For testing, we note a small performance gap with $\lambda \in \{0.1, 0.01\}$ on LF test set, with AP=78.29 for $\lambda = 0.01$ and AP=77.15 for $\lambda = 0.1$. For the EBM margin, PEBAL reaches AP$\in [76.9, 78.3]$ and FPR$\in [0.8, 1.3]$ for $m_{in} \in [-12, -22]$ and $m_{out} \in [-2, -8]$ for different values of m_{in} and m_{out} on LF test set.

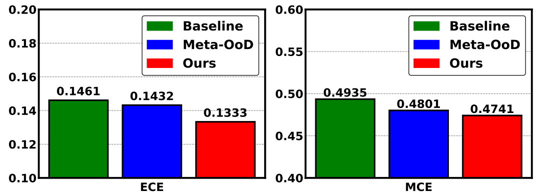

Fig. 2: Confidence calibration performances between WideResnet38 baseline, Meta-OoD [2], and our approach.

4 Training Details on Cityscapes

Following [1,2], we use the same DeepLabv3+ [3] with WideResnet38 (90.3 mIoU on Cityscapes Val) trained by Nvidia [10] as one of the backbones of our segmentation model. As mentioned in [10], the model is firstly pre-trained on Mapillary Vista dataset [9], and then fine-tuned on Cityscapes train set with their proposed label relaxation loss and sdc-aug label propagation. Their model uses a different {cv2: monchengladbach, strasbourg, stuttgart} validation split than the standard split {cv0: munster, lindau, frankfurt}. Please refer to their paper for more details. For DeepLabv3+ [3] with Resnet101 backbone (80.3 mIoU on Cityscapes Val) from [7], the authors trained their model with the standard cv0 train/validation split using default formulations in [3]. All those checkpoints are downloaded from their official Github pages.

5 Results Based on Different DeepLabv3+ Checkpoint

In this section, we show the results of another DeepLabv3+ [3] with WideResnet38 trained by Nvidia [10] using the Cityscapes **{cv0: munster, lindau, frankfurt}** standard train/val split. The checkpoint is downloaded from the their official Github page [10], with a 81.8% mIoU on Cityscapes validation set. This model was firstly pre-trained on Mapillary Vista dataset [9] and then fine-tuned on Cityscapes but without their label relaxation loss and sdc-aug label propagation. As shown in Tab. 3, our model outperforms the previous methods by a large margin on all three benchmarks, regardless of the backbones, the segmentation accuracy and the Cityscapes train/val splits. Notably, our method surpasses the previous SOTA SML by 40%, 50% and 20% of AP on three datasets, respectively. We also achieve best AUC and FPR results on all datasets.

Confidence Calibration. In Fig. 2, we show that our model can also improve the calibration of the segmentation confidence. This figure shows that we improve the ECE and MCE [4] scores by a small margin, showing another benefit of using our PEBAL approach.

Table 3: Anomaly segmentation results on **Fishyscapes validation sets** (LostAndFound and Static), and the **Road Anomaly testing set**, with **WideResnet38** backbone under **cv0** standard train/val split.

Methods	FS LostAndFound			FS Static			Road Anomaly		
	AUC ↑	AP ↑	FPR_{95} ↓	AUC ↑	AP ↑	FPR_{95} ↓	AUC ↑	AP ↑	FPR_{95} ↓
MSP [5]	89.26	11.84	32.55	89.26	11.84	32.55	72.37	20.23	67.98
Max Logit [5]	93.14	12.78	38.15	93.27	18.89	25.49	76.39	23.46	64.55
Entropy [6]	89.01	8.79	47.81	90.28	15.19	31.71	73.70	22.13	67.42
Energy [8]	93.45	14.29	37.71	93.52	19.22	25.02	76.76	23.48	64.04
SML [7]	96.03	21.71	20.09	95.79	32.04	15.81	74.45	22.16	68.59
Ours	**98.52**	**64.43**	**6.56**	**99.33**	**86.01**	**2.63**	**88.85**	**44.41**	**37.98**

References

1. Chan, R., Lis, K., Uhlemeyer, S., Blum, H., Honari, S., Siegwart, R., Salzmann, M., Fua, P., Rottmann, M.: Segmentmeifyoucan: A benchmark for anomaly segmentation. NeurIPS (2021) 3
2. 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. pp. 5128–5137 (2021) 3
3. 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). pp. 801–818 (2018) 3
4. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International Conference on Machine Learning. pp. 1321–1330. PMLR (2017) 3
5. Hendrycks, D., Basart, S., Mazeika, M., Mostajabi, M., Steinhardt, J., Song, D.: Scaling out-of-distribution detection for real-world settings. arXiv preprint arXiv:1911.11132 (2019) 4
6. Hendrycks, D., Gimpel, K.: A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint arXiv:1610.02136 (2016) 4
7. 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. pp. 15425–15434 (2021) 3, 4
8. Liu, W., Wang, X., Owens, J.D., Li, Y.: Energy-based out-of-distribution detection. arXiv preprint arXiv:2010.03759 (2020) 4
9. Neuhold, G., Ollmann, T., Rota Bulo, S., Kontschieder, P.: The mapillary vistas dataset for semantic understanding of street scenes. In: Proceedings of the IEEE international conference on computer vision. pp. 4990–4999 (2017) 3
10. Zhu, Y., Sapra, K., Reda, F.A., Shih, K.J., Newsam, S., Tao, A., Catanzaro, B.: Improving semantic segmentation via video propagation and label relaxation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8856–8865 (2019) 3