SUPPLEMENTARY MATERIAL
Image-Consistent Detection of Road Anomalies as Unpredictable Patches

Tomáš Vojíř and Jiří Matas
Czech Technical University in Prague, FEE
Technická 2, Prague, Czech Republic
{vojirtom,matas}@fel.cvut.cz

1. Additional Qualitative Results

A supplementary video *wacv2023_supp_ID220.mp4* shows the qualitative results on evaluated datasets plus videos containing more challenging scenarios. Specifically, we evaluate the proposed DaCUP method on RA [9], RO [8], RO21A [2], the bad weather images from [2] which are not included in the official *SegmentMeIfYouCan* benchmark evaluation (for details see explanation in [2]) and several Youtube dashcam videos from-the-wild.

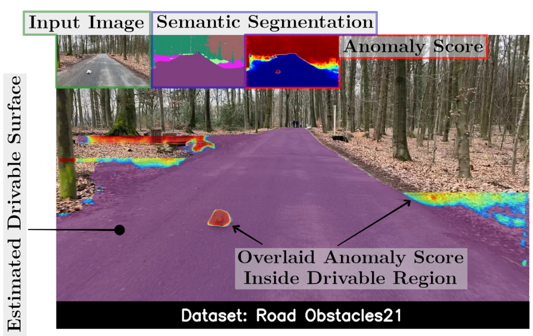

Figure 1. Illustration of the DaCUP output in the supplementary video. For detailed explanation see the text in Section 1.

The output of the proposed method is visualized (See Fig 1) as follows: The three small top left images display the input image, the estimated semantic segmentation and the anomaly scores.

To simulate real-world application, we estimate the drivable surface which is overlaid on top of the main image (purple color). The drivable surface is estimated from the semantic segmentation by first finding a road vanishing row as a topmost row with at least 5% row pixels classified as road. For each row below the vanishing row, all pixels between the leftmost and rightmost road classified pixels are considered as drivable surface. Lastly, for each drivable surface pixel, we overlay the color encoded (jet color scheme) anomaly score if it is larger than a threshold (set to 0.2). This type of visualization simulates the real-world application where the proposed method gets the drivable surface from external signal and has to estimate anomaly score for every location.

2. Results for Individual Datasets

This section provides additional results for all individual datasets used in the main paper. The evaluation is carried out on commonly available anomaly detection datasets: Lost-and-Found (LaF) [10], Road Anomaly (RA) [9], Road Obstacles (RO) [8], Fishyscapes (FS) [1] and a novel dataset from *SegmentMeIfYouCan* benchmark, Road Obstacles 21 [2], which is composed of RO and new (222) and validation (30) images, denoted as RO21A. Tables 1, 2 and 3 show results for the individual dataset and experiments as described in the main paper.

	Training Data	*LaFRAROFS*										*Obstacle Track+*					
		LaF		FS		RA		RO		Average		RO		RO21A		Average	
		$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$
baseline	CityScapes	78.0	4.1	78.3	4.0	94.4	9.2	84.0	0.4	83.7	4.4	84.0	0.4	28.3	52.1	56.2	26.3
w/ emb. space	CityScapes	84.2	2.3	90.4	1.4	92.7	8.7	85.0	0.2	88.1	3.2	85.0	0.2	11.8	11.6	48.4	5.9
baseline	S(CityScapes,BDD100k)	80.3	4.8	83.6	5.1	95.8	6.7	81.9	1.2	85.4	4.5	81.9	1.2	40.7	18.7	61.3	10.0
w/ emb. space	S(CityScapes,BDD100k)	80.1	3.6	87.0	1.7	93.6	7.9	89.6	0.1	87.6	3.3	89.6	0.1	36.4	7.7	63.0	3.9
baseline	CityScapes,BDD100k	75.8	5.9	82.2	5.1	95.6	6.8	80.3	1.2	83.5	4.8	80.3	1.2	24.0	44.6	52.2	22.9
w/ emb. space	CityScapes,BDD100k	85.2	3.3	89.9	1.8	95.8	6.3	93.8	0.1	91.2	2.9	93.8	0.1	63.2	5.3	78.5	2.7

Table 1. Embedding bottleneck - dependence on training with data varying in size and road appearance diversity. $S(\cdot)$ indicates sub-sampling of the datasets to the size of CityScapes and to have roughly equal number of images from each dataset.

baseline	embedding space	embedding channels	inpainting	*LaFRAROFS*										*Obstacle Track+*					
				LaF		FS		RA		RO		Average		RO		RO21A		Average	
				$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$
✓				80.3	4.8	83.6	5.1	95.8	6.7	81.9	1.2	85.4	4.5	81.9	1.2	40.7	18.7	61.3	10.0
✓	✓			80.1	3.6	87.0	1.7	93.6	7.9	89.6	0.1	87.6	3.3	89.6	0.1	36.4	7.7	63.0	3.9
✓	✓	✓		86.0	1.9	90.6	1.2	94.9	6.3	92.2	0.2	90.9	2.4	92.2	0.2	68.8	5.3	80.5	2.7
✓	✓	✓		0.6	0.3	1.1	0.3	0.5	0.6	1.3	0.1	0.9	0.3	1.3	0.1	3.9	1.4	2.5	0.7
✓	✓	✓	✓	84.5	2.6	89.7	1.4	96.2	5.5	94.3	0.1	91.2	2.4	94.3	0.1	77.9	2.9	86.1	1.5

Table 2. Ablation study of the methods building blocks. The standard deviation in performance for multiple training runs to established a significance of result differences for the third row are shown in the red marked row.

	LaFRAROFS										*Obstacle Track+*					
	LaF		FS		RA		RO		Average		RO		RO21A		Average	
	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$	$\overline{\text{AP}}\uparrow$	$\overline{\text{FPR}}_{95}\downarrow$
Maximum Softmax [5]	27.0	35.6	34.0	8.2	44.4	58.5	13.3	13.2	29.7	28.9	13.3	13.2	10.2	29.1	11.8	21.2
+ inpainted	43.5	28.6	54.2	6.8	51.6	58.4	59.8	12.9	52.3	26.7	59.8	12.9	22.8	26.8	41.3	19.8
Mahalanobis [6]	38.2	26.7	64.7	6.1	30.2	91.7	45.6	6.0	44.7	32.6	45.6	6.0	17.5	29.9	31.6	17.9
+ inpainted	49.5	7.2	69.1	3.1	41.3	56.7	61.0	1.0	55.2	17.0	61.0	1.0	30.7	17.5	45.8	9.2
JSRNet [11]	78.0	4.1	78.3	4.0	94.4	9.2	84.0	0.4	83.7	4.4	84.0	0.4	28.3	52.1	56.2	26.3
+ inpainted	83.2	1.7	84.2	2.0	93.2	7.6	87.7	0.2	87.1	2.9	87.7	0.2	43.8	43.9	65.7	22.0
ODIN [7]	54.9	23.3	68.5	10.5	45.9	59.8	33.6	9.9	50.7	25.9	33.6	9.9	7.1	27.6	20.4	18.7
+ inpainted	69.1	4.6	78.8	3.4	61.9	38.3	74.9	1.0	71.2	11.8	74.9	1.0	31.8	13.2	53.4	7.1
Image Resynthesis [9]	62.9	43.1	66.7	3.1	76.4	48.1	59.2	5.5	66.3	25.0	59.2	5.5	50.6	14.1	54.9	9.8
+ inpainted	62.4	43.1	68.1	3.1	76.6	48.1	59.2	5.5	66.6	25.0	59.2	5.5	56.6	14.1	57.9	9.8
SynBoost [4]	77.8	6.8	92.5	0.7	63.7	52.3	76.2	1.7	77.5	15.4	76.2	1.7	60.7	5.1	68.4	3.4
+ inpainted	80.2	3.3	90.6	0.5	69.7	52.4	82.2	0.9	80.7	14.3	82.2	0.9	68.8	3.0	75.5	2.0
Maximized Entropy [3]	75.6	9.4	77.2	10.1	96.2	6.0	96.3	0.1	86.3	6.4	96.3	0.1	76.6	3.8	86.4	1.9
+ inpainted	75.0	9.2	78.1	10.1	93.2	5.5	95.3	0.0	85.4	6.2	95.3	0.0	77.4	1.9	86.3	1.0
DaCUP (**Ours**) w/o	86.0 ±0.6	1.9±0.3	90.6 ±1.1	1.2±0.3	94.9 ±0.5	6.3±0.6	92.2 ±1.3	0.2±0.1	90.9 ±0.9	2.4±0.3	92.2 ±1.3	0.2±0.1	68.8 ±3.9	5.3±1.4	80.5 ±2.5	2.7±0.7
+ inpainted	86.2 ±0.7	1.8±0.3	90.7 ±0.9	1.2±0.5	93.4 ±0.9	5.9±0.3	92.6 ±0.4	0.1±0.0	90.7 ±0.6	2.3±0.2	92.6 ±0.4	0.1±0.0	69.3 ±2.8	3.1±0.2	80.9 ±1.6	1.6±0.1
+ inpainted trained	84.5	2.6	89.7	1.4	96.2	5.5	94.3	0.1	91.2	2.4	94.3	0.1	77.9	2.9	86.1	1.5

Table 3. Inpainting module impact on state-of-the-art methods.

References

[1] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. Fishyscapes: A Benchmark for Safe Semantic Segmentation in Autonomous Driving. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 2403–2412, 2019.

[2] Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Siegwart, Pascal Fua, Mathieu Salzmann, and Matthias Rottmann. SegmentMeIfYouCan: A Benchmark for Anomaly Segmentation, 2021.

[3] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy Maximization and Meta Classification for Out-of-Distribution Detection in Semantic Segmentation. In *Int. Conf. Comput. Vis.*, pages 5128–5137, October 2021.

[4] Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. Pixel-Wise Anomaly Detection in Complex Driving Scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 16918–16927, June 2021.

[5] Dan Hendrycks and Kevin Gimpel. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. *Proceedings of International Conference on Learning Representations*, 2017.

[6] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Adv. Neural Inform. Process. Syst.*, NIPS'18, page 7167–7177, 2018.

[7] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *Proceedings of International Conference on Learning Representations*, 2018.

[8] Krzysztof Lis, Sina Honari, Pascal Fua, and Mathieu Salzmann. Detecting Road Obstacles by Erasing Them, 2021.

[9] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the Unexpected via Image Resynthesis. In *Int. Conf. Comput. Vis.*, October 2019.

[10] 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 *International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[11] Tomas Vojir, Tomáš Šipka, Rahaf Aljundi, Nikolay Chumerin, Daniel Olmeda Reino, and Jiri Matas. Road Anomaly Detection by Partial Image Reconstruction With Segmentation Coupling. In *Int. Conf. Comput. Vis.*, pages 15651–15660, October 2021.