

Supplementary Material:

Unsupervised Anomaly Detection from Time-of-Flight Depth Images

Pascal Schneider, Jason Rambach, Bruno Mirbach, Didier Stricker
German Research Center for Artificial Intelligence (DFKI)
Trippstadter Str. 122, 67663 Kaiserslautern, Germany

firstname.lastname@dfki.de

1. Contents

This document provides the following supplementary material for the main paper:

- a list of anomaly types assigned to the three anomaly categories we used in the evaluation
- evaluation of inference time performance
- a more fine-grained evaluation of the anomaly detection performance on each single anomaly type

2. Anomaly Categories

Aggressive Behaviour	Medical Issue	Left-behind Object
COF-ARG	COLLAPSE	STP-LBO
CROSS-ARG	CRAWL	WALK-LBO
PASS-SO	FLOOR	PASS-1LBO
RUN-THO	STAGGER	PASS-2LBO
WALK-THO		RUN-LBO
CROSS-SO		COF-LBO

Table 1. List of anomaly types from the TIMo dataset [1] assigned to each of the categories that were used for evaluation in the main paper.

Table 1 lists the anomaly types we assigned to the three coarse categories that were used in the evaluations in the main paper. A short description of each of the anomaly types in the TIMo dataset [1] can be found on the dataset website¹. Note that a few anomaly types that occur in the tilted view data did not fit well into any of the three categories we defined and also do not form a coherent category combined. We therefore left those anomaly types out.

¹<https://vizta-tof.k1.dfki.de/building-data-format/>

The fine-grained evaluation in Section 4 does cover these anomaly types, though.

3. Runtime Performance

		MSE	F-MSE	W-MSE
Tilted View (288×320)	R-CAE	120	120	119
	P-CAE	47	46	46
	R-ViT-AE	139	137	136
	P-ViT-AE	59	58	58
	P-ConvLSTM	39	39	39
Top-Down V. (512×512)	R-CAE	48	48	48
	P-CAE	20	20	20
	R-ViT-AE	82	80	80
	P-ViT-AE	84	82	82
	P-ConvLSTM	41	40	40

Table 2. Inference time performance in terms of frames per second. Computation of the foreground mask is not considered, because pre-computed masks were used.

Table 2 shows the frames per second that the *Nvidia GeForce GTX Titan X* is able to process during inference time. Note that this evaluation uses pre-computed foreground masks because the implementation was not yet adapted for on-line operation. The performance of the networks which use either the F-MSE or the W-MSE loss function can be expected to be slightly lower when running on-line. However, this has the equal on networks using the same loss function, so the difference in performance across network architectures remains roughly the same.

The performance is usually lower on the top-down view data due to the larger image size compared to the tilted view data. However, there are a exceptions to this for the P-ViT-AE and the P-Conv-LSTM. We suspect that the image size of 512×512 facilitates automatic internal optimizations in the processing pipeline, but did not investigate this further.

4. Detailed Detection Performance

Table 3 and Table 4 show the detection performance w.r.t. each anomaly type from the TIMo dataset [1]. The first row shows the overall performance again. Note that due to imbalanced occurrences of the different anomaly types in the dataset, the overall performance is not equal to the average performance across anomaly types.

References

- [1] Pascal Schneider, Yuriy Anisimov, Raisul Islam, Bruno Mirbach, Jason Rambach, Frédéric Grandidier, and Didier Stricker. Timo – a dataset for indoor building monitoring with a time-of-flight camera, 2021. 1, 2

		OVERALL	COF-ARG	COF-LBO	COF-NONS	COLLAPSE	CRAWL	CROSS-ARG	FLOOR	PASS-2LBO	PASS-SO
MSE	R-CAE	66.4	75.8	69.5	70.0	44.0	38.6	83.8	34.0	70.3	83.0
	P-CAE	71.4	75.2	73.6	65.7	56.4	42.7	78.4	53.4	74.5	82.8
	P-ConvLSTM	62.8	36.8	60.7	55.2	44.5	52.0	70.6	33.8	61.5	83.7
	R-ViT-AE	64.9	24.6	49.9	47.2	59.7	57.2	79.0	36.5	81.9	83.1
	P-ViT-AE	65.1	23.7	50.3	45.7	59.1	56.1	78.3	37.3	81.6	83.5
F-MSE	R-CAE	68.5	97.4	67.2	87.4	54.5	39.7	87.1	51.0	81.3	85.4
	P-CAE	68.6	96.2	65.6	86.2	53.5	37.8	87.0	50.5	82.9	84.5
	P-ConvLSTM	67.5	93.3	64.3	79.5	53.5	40.1	79.5	49.4	78.1	86.0
	R-ViT-AE	71.2	90.4	64.2	84.7	58.4	47.8	80.4	63.6	84.9	82.0
	P-ViT-AE	71.2	90.6	64.4	85.8	62.9	49.0	78.0	64.3	86.4	81.8
W-MSE	R-CAE	70.0	92.9	71.9	81.3	48.4	35.8	90.5	38.3	77.5	87.2
	P-CAE	78.1	90.5	75.5	83.1	61.7	47.5	87.9	65.6	85.8	86.3
	P-ConvLSTM	67.1	62.5	64.0	64.5	48.8	50.2	76.9	38.3	69.4	88.2
	R-ViT-AE	71.7	79.4	61.1	72.6	63.7	51.1	82.6	49.5	87.1	84.5
	P-ViT-AE	70.5	83.6	61.0	70.7	58.7	44.6	79.4	48.1	83.7	84.7

		RUN-LBO	RUN-THO	RUN-WEAP	SPORT	STAGGER	STP-LBO	UMB	WALK-LBO	WALK-THO	WALK-WEAP
MSE	R-CAE	53.4	79.7	68.3	46.1	75.3	75.3	53.2	64.6	61.6	58.3
	P-CAE	64.1	84.4	69.3	47.8	87.2	81.7	55.6	73.3	75.5	51.6
	P-ConvLSTM	73.4	81.5	84.1	61.3	73.1	67.2	63.0	61.5	71.8	66.3
	R-ViT-AE	61.5	82.9	63.3	62.8	59.4	87.0	52.2	78.9	72.1	69.3
	P-ViT-AE	66.2	83.8	66.7	63.1	61.9	86.6	54.7	78.2	73.0	69.4
F-MSE	R-CAE	63.9	54.9	45.5	61.1	63.8	55.3	64.9	58.3	67.2	57.6
	P-CAE	66.2	57.0	49.4	53.9	67.3	57.5	66.0	61.8	68.7	57.5
	P-ConvLSTM	70.9	63.4	64.2	63.6	66.0	52.3	65.4	57.1	70.8	62.3
	R-ViT-AE	67.8	59.7	51.2	75.5	71.2	64.1	62.8	66.9	69.3	60.0
	P-ViT-AE	71.0	58.2	52.4	71.0	71.0	64.0	60.5	65.3	70.8	58.7
W-MSE	R-CAE	58.9	75.1	64.4	54.5	75.3	70.4	61.2	64.3	65.7	60.3
	P-CAE	71.9	84.2	69.6	66.5	86.2	83.2	58.1	79.1	80.6	52.1
	P-ConvLSTM	75.2	79.5	81.6	71.5	73.1	63.5	66.6	61.4	74.5	67.2
	R-ViT-AE	65.6	74.1	57.5	65.9	68.0	80.3	59.7	74.2	75.7	65.6
	P-ViT-AE	67.5	72.7	63.9	61.1	70.1	77.5	64.7	71.4	76.2	63.9

Table 3. Evaluation of performance on the tilted view data. Table cells are color-coded according to the AUC.

		OVERALL	COLLAPSE	CRAWL	CROSS-ARG	CROSS-SO	FLOOR	PASS-1LBO	PASS-2LBO	PASS-SO	RUN-LBO	STAGGER	STP-LBO	WALK-LBO
MSE	R-CAE	56.4	45.7	59.0	55.9	65.9	60.2	58.8	56.9	63.3	56.0	54.7	56.2	52.7
	P-CAE	52.1	75.4	60.8	21.7	44.9	42.8	48.4	45.5	41.0	57.0	41.9	54.2	55.9
	P-ConvLSTM	55.3	41.4	63.9	65.7	81.4	54.9	56.3	55.7	75.4	57.3	55.6	52.2	51.5
	R-ViT-AE	57.0	52.5	63.7	56.3	65.8	61.3	59.8	57.8	62.5	56.2	53.8	55.3	51.5
	P-ViT-AE	57.3	53.2	64.7	59.1	69.9	60.3	58.9	57.8	66.5	57.7	54.6	54.5	52.1
F-MSE	R-CAE	73.2	87.6	70.8	84.0	88.9	83.6	83.1	83.7	87.9	62.0	60.4	66.8	70.4
	P-CAE	66.0	70.3	63.1	94.6	92.5	78.7	74.9	78.5	92.0	54.0	67.3	56.5	59.3
	P-ConvLSTM	65.6	61.3	66.6	90.7	93.3	74.9	76.1	78.4	90.8	57.1	65.6	55.9	59.9
	R-ViT-AE	65.3	68.4	65.3	90.7	90.9	76.0	74.1	79.1	91.7	53.0	63.9	56.5	59.0
	P-ViT-AE	63.0	60.1	63.5	92.5	92.5	72.0	71.8	76.6	92.6	52.2	68.1	53.1	56.6
W-MSE	R-CAE	63.9	61.5	65.6	70.5	79.6	73.4	70.9	70.7	78.7	57.7	58.6	58.7	58.3
	P-CAE	67.8	86.5	77.0	78.2	85.2	78.0	71.1	73.3	84.8	60.8	53.8	59.2	61.3
	P-ConvLSTM	59.1	44.2	67.7	75.6	89.7	60.2	65.0	64.8	83.4	59.0	58.2	52.8	53.9
	R-ViT-AE	61.5	58.6	68.2	62.8	71.9	66.2	69.5	68.5	73.6	58.1	56.0	56.3	55.7
	P-ViT-AE	60.3	55.4	68.0	68.3	79.2	63.9	67.4	67.1	77.4	58.7	56.0	53.9	54.0

Table 4. Evaluation of performance on the top-down view data. Table cells are color-coded according to the AUC.