Results of the paper:
”Benchmarking Safety Monitors for
Image Classifiers with Machine Learning”

May 10, 2021

1 Novelty Class

Table 1: Comparing data-based monitors for novelty class.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
GTSRB-BTSC	ALOOC	0.01	**0.50**	0.49	0.17	0.51	**0.57**
	OOB	**0.23**	0.61	0.11	**0.24**	0.90	0.52
	ODIN	0.03	0.99	**0.0**	0.16	**1.0**	0.06
CIFAR10-GTSRB	ALOOC	0.02	0.63	0.34	0.18	0.66	0.47
	OOB	0.11	0.72	0.16	0.20	0.84	0.41
	ODIN	**0.23**	**0.61**	**0.10**	**0.24**	**0.9**	**0.52**
GTSRB-CIFAR10	ALOOC	0.05	**0.56**	0.37	0.81	0.63	0.63
	OOB	**0.15**	0.79	0.09	**0.82**	0.91	**0.74**
	ODIN	0.07	1.0	**0.02**	0.17	**0.98**	0.06

2 Adversarial Attack

Table 2: Comparing data-based monitors for CIFAR-10 and GTSRB datasets with a FGSM attack.

CIFAR-10

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
FGSM	ALOOC	-0.23	0.89	0.29	0.28	0.71	0.25
	OOB	-0.13	0.92	0.16	0.31	0.84	0.24
	ODIN	0.06	0.14	0.81	0.37	0.19	0.62

GTSRB

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
FGSM	ALOOC	0.19	0.22	0.59	0.44	0.41	0.66
	OOB	-0.01	1.0	0.0	0.31	1.0	0.14
	ODIN	0.11	0.92	0.02	0.34	0.98	0.26

3 Anomaly

Table 3: Comparing data-based monitors for cifar10 dataset with different types of anomalies.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
pixel trap (3)	ALOOC	0.6	0.0	0.59	0.97	0.41	0.9
	OOB	0.02	1.0	0.0	0.14	1.0	0.04
	ODIN	-0.17	0.43	0.81	0.07	0.19	0.59
row add logic (3)	ALOOC	-0.01	0.42	0.59	0.14	0.41	0.62
	OOB	0.01	1.0	0.0	0.14	1.0	0.04
	ODIN	0.01	0.18	0.81	0.15	0.19	0.74
shifted pixel (3)	ALOOC	0.16	0.21	0.59	0.24	0.41	0.76
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.0	0.19	0.81	0.15	0.19	0.74
pixel trap (1)	ALOOC	0.51	0.02	0.59	0.77	0.41	0.88
	OOB	0.02	1.0	0.0	0.14	1.0	0.04
	ODIN	-0.52	0.84	0.81	0.04	0.19	0.22
row add logic (1)	ALOOC	0.0	0.41	0.59	0.14	0.41	0.63
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.11	0.09	0.81	0.26	0.19	0.79
shifted pixel (1)	ALOOC	0.04	0.36	0.59	0.16	0.41	0.67
	OOB	-0.01	1.0	0.0	0.14	1.0	0.04
	ODIN	-0.05	0.26	0.81	0.11	0.19	0.7

Table 4: Comparing data-based monitors for gtsrb dataset with different types of anomalies.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
pixel trap (3)	ALOOC	0.6	0.09	0.29	0.66	0.71	0.87
	OOB	-0.01	0.85	0.16	0.19	0.84	0.27
	ODIN	-0.01	0.98	0.02	0.2	0.98	0.09
row add logic (3)	ALOOC	-0.02	0.73	0.29	0.19	0.71	0.38
	OOB	0.03	0.81	0.16	0.2	0.84	0.32
	ODIN	0.02	0.97	0.02	0.2	0.98	0.11
shifted pixel (3)	ALOOC	0.02	0.69	0.29	0.2	0.71	0.42
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.02	0.97	0.02	0.2	0.98	0.11
pixel trap (1)	ALOOC	0.19	0.47	0.29	0.27	0.71	0.61
	OOB	0.01	0.83	0.16	0.2	0.84	0.28
	ODIN	0.01	0.97	0.02	0.2	0.98	0.1
row add logic (1)	ALOOC	0.0	0.7	0.29	0.2	0.71	0.41
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.11	0.9	0.02	0.21	0.98	0.21
shifted pixel (1)	ALOOC	0.0	0.7	0.29	0.2	0.71	0.41
	OOB	0.06	0.77	0.16	0.21	0.84	0.35
	ODIN	0.12	0.89	0.02	0.21	0.98	0.23

4 Distributional Shift

Table 5: Comparing data-based monitors for cifar10 dataset with different types of distributional shift.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
rotated	ALOOC	0.0	0.0	0.29	1.0	0.71	0.83
	OOB	0.02	1.0	0.0	0.14	1.0	0.04
	ODIN	-0.1	0.32	0.81	0.09	0.19	0.66
snow (5)	ALOOC	-0.01	0.42	0.59	0.14	0.41	0.62
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.14	0.08	0.81	0.29	0.19	0.8
fog (5)	ALOOC	0.47	0.03	0.59	0.68	0.41	0.88
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	-0.01	0.21	0.81	0.13	0.19	0.73
brightness (5)	ALOOC	0.0	0.4	0.59	0.14	0.41	0.63
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.22	0.04	0.81	0.45	0.19	0.82
contrast (5)	ALOOC	0.27	0.12	0.59	0.36	0.41	0.82
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.03	0.16	0.81	0.17	0.19	0.75
saturate (5)	ALOOC	0.16	0.21	0.59	0.24	0.41	0.76
	OOB	0.0	0.0	0.13	1.0	0.87	0.93
	ODIN	0.0	0.0	0.84	1.0	0.16	0.27
snow (2)	ALOOC	0.56	0.01	0.59	0.88	0.41	0.89
	OOB	0.02	1.0	0.0	0.14	1.0	0.04
	ODIN	0.06	0.14	0.81	0.19	0.19	0.77
fog (2)	ALOOC	0.29	0.11	0.59	0.38	0.41	0.82
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.32	0.02	0.81	0.68	0.19	0.84
brightness (2)	ALOOC	0.07	0.32	0.59	0.18	0.41	0.69
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.0	0.2	0.81	0.14	0.19	0.73
contrast (2)	ALOOC	0.22	0.16	0.59	0.3	0.41	0.79
	OOB	-0.01	1.0	0.0	0.14	1.0	0.04
	ODIN	0.37	0.01	0.81	0.86	0.19	0.85
saturate (2)	ALOOC	-0.01	0.42	0.59	0.14	0.41	0.62
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	-0.06	0.27	0.81	0.11	0.19	0.69

Table 6: Comparing data-based monitors for gtsrb dataset with different types of distributional shift.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
rotated	ALOOC	0.0	0.0	0.55	1.0	0.45	0.62
	OOB	0.09	0.75	0.16	0.21	0.84	0.38
	ODIN	-0.12	1.0	0.02	0.19	0.98	0.06
snow (5)	ALOOC	0.81	0.0	0.29	1.0	0.71	0.94
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.16	0.85	0.02	0.22	0.98	0.28
fog (5)	ALOOC	-0.28	0.93	0.29	0.16	0.71	0.15
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.0	0.98	0.02	0.2	0.98	0.1
brightness (5)	ALOOC	0.2	0.45	0.29	0.28	0.71	0.62
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.13	0.88	0.02	0.21	0.98	0.24
contrast (5)	ALOOC	-0.32	0.95	0.29	0.15	0.71	0.13
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.03	0.97	0.02	0.2	0.98	0.12
saturate (5)	ALOOC	0.18	0.48	0.29	0.26	0.71	0.6
	OOB	0.0	0.0	0.18	1.0	0.82	0.9
	ODIN	0.0	0.0	0.04	1.0	0.96	0.98
snow (2)	ALOOC	0.81	0.0	0.29	1.0	0.71	0.94
	OOB	0.06	0.77	0.16	0.21	0.84	0.35
	ODIN	0.14	0.86	0.02	0.22	0.98	0.26
fog (2)	ALOOC	-0.27	0.92	0.29	0.16	0.71	0.16
	OOB	0.0	0.84	0.16	0.2	0.84	0.27
	ODIN	0.22	0.75	0.02	0.24	0.98	0.39
brightness (2)	ALOOC	0.06	0.64	0.29	0.21	0.71	0.47
	OOB	0.01	0.83	0.16	0.2	0.84	0.28
	ODIN	0.0	0.98	0.02	0.2	0.98	0.1
contrast (2)	ALOOC	-0.29	0.93	0.29	0.16	0.71	0.14
	OOB	-0.05	0.88	0.16	0.19	0.84	0.23
	ODIN	0.54	0.3	0.02	0.44	0.98	0.78
saturate (2)	ALOOC	0.02	0.68	0.29	0.2	0.71	0.43
	OOB	0.07	0.77	0.16	0.21	0.84	0.36
	ODIN	0.03	0.96	0.02	0.2	0.98	0.12

5 Noise

Table 7: Comparing data-based monitors for cifar10 dataset with different types of noise.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
gaussian noise (2)	ALOOC	-0.03	0.45	0.59	0.13	0.41	0.6
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.0	1.0	0.0	0.14	1.0	0.04
gaussian noise (5)	ALOOC	-0.01	0.42	0.59	0.14	0.41	0.62
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.01	0.18	0.81	0.15	0.19	0.74
impulse noise (2)	ALOOC	-0.02	0.44	0.59	0.13	0.41	0.61
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.02	0.17	0.81	0.16	0.19	0.75
impulse noise (5)	ALOOC	-0.02	0.44	0.59	0.13	0.41	0.61
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.02	0.17	0.81	0.16	0.19	0.75
shot noise (2)	ALOOC	-0.02	0.44	0.59	0.13	0.41	0.61
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.03	0.16	0.81	0.17	0.19	0.76
shot noise (5)	ALOOC	-0.04	0.46	0.59	0.13	0.41	0.59
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.0	0.19	0.81	0.14	0.19	0.74
spatter (2)	ALOOC	-0.02	0.44	0.59	0.13	0.41	0.61
	OOB	0.01	1.0	0.0	0.14	1.0	0.04
	ODIN	0.01	0.18	0.81	0.15	0.19	0.74
spatter (5)	ALOOC	-0.01	0.42	0.59	0.14	0.41	0.62
	OOB	0.01	1.0	0.0	0.14	1.0	0.04
	ODIN	0.01	0.19	0.81	0.15	0.19	0.74
speckle noise (2)	ALOOC	-0.02	0.44	0.59	0.13	0.41	0.61
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.01	0.18	0.81	0.15	0.19	0.74
speckle noise (5)	ALOOC	0.01	0.4	0.59	0.15	0.41	0.64
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.0	0.19	0.81	0.15	0.19	0.74
defocus blur (2)	ALOOC	0.14	0.23	0.59	0.23	0.41	0.75
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.01	0.18	0.81	0.15	0.19	0.74
defocus blur (5)	ALOOC	0.48	0.03	0.59	0.7	0.41	0.88
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.04	0.16	0.81	0.17	0.19	0.76
elastic transform (2)	ALOOC	0.43	0.04	0.59	0.6	0.41	0.87
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.04	0.15	0.81	0.18	0.19	0.76
elastic transform (5)	ALOOC	0.61	0.0	0.59	1.0	0.41	0.9
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.09	0.11	0.81	0.23	0.19	0.78
glass blur (2)	ALOOC	0.06	0.32	0.59	0.17	0.41	0.69
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.05	0.14	0.81	0.19	0.19	0.77
glass blur (5)	ALOOC	0.3	0.1	0.59	0.4	0.41	0.83
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.1	0.11	0.81	0.23	0.19	0.79
zoom blur (2)	ALOOC	0.49	0.03	0.59	0.72	0.41	0.88
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.07	0.13	0.81	0.2	0.19	0.77
zoom blur (5)	ALOOC	0.47	0.03	0.59	0.69	0.41	0.88
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.1	0.1	0.81	0.24	0.19	0.79
gaussian blur (2)	ALOOC	0.13	0.24	0.59	0.22	0.41	0.74
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.06	0.13	0.81	0.19	0.19	0.77
gaussian blur (5)	ALOOC	0.29	0.11	0.59	0.39	0.41	0.82
	OOB	0.0	1.0	0.0	0.14	1.0	0.04
	ODIN	0.13	0.08	0.81	0.28	0.19	0.8

Table 8: Comparing data-based monitors for gtsrb dataset with different types of noise.

Variation	Method	MCC	FPR	FNR	Pr	Re	F1
gaussian noise (2)	ALOOC	0.06	0.63	0.29	0.21	0.71	0.47
	OOB	0.01	0.83	0.16	0.2	0.84	0.28
	ODIN	0.0	0.84	0.16	0.2	0.84	0.28
gaussian noise (5)	ALOOC	0.17	0.5	0.29	0.26	0.71	0.59
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.0	0.98	0.02	0.2	0.98	0.1
impulse noise (2)	ALOOC	0.09	0.59	0.29	0.23	0.71	0.51
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.0	0.98	0.02	0.2	0.98	0.1
impulse noise (5)	ALOOC	0.5	0.15	0.29	0.53	0.71	0.83
	OOB	0.02	0.82	0.16	0.2	0.84	0.3
	ODIN	-0.02	0.98	0.02	0.2	0.98	0.09
shot noise (2)	ALOOC	0.0	0.71	0.29	0.2	0.71	0.41
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	-0.06	0.99	0.02	0.19	0.98	0.07
shot noise (5)	ALOOC	0.05	0.64	0.29	0.21	0.71	0.47
	OOB	0.01	0.83	0.16	0.2	0.84	0.28
	ODIN	-0.01	0.98	0.02	0.2	0.98	0.09
spatter (2)	ALOOC	0.65	0.06	0.29	0.73	0.71	0.89
	OOB	0.05	0.78	0.16	0.21	0.84	0.34
	ODIN	-0.01	0.98	0.02	0.2	0.98	0.09
spatter (5)	ALOOC	0.66	0.06	0.29	0.74	0.71	0.89
	OOB	0.05	0.79	0.16	0.21	0.84	0.34
	ODIN	-0.13	1.0	0.02	0.19	0.98	0.06
speckle noise (2)	ALOOC	0.01	0.69	0.29	0.2	0.71	0.42
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	-0.12	1.0	0.02	0.19	0.98	0.06
speckle noise (5)	ALOOC	0.05	0.65	0.29	0.21	0.71	0.46
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	-0.01	0.98	0.02	0.2	0.98	0.09
defocus blur (2)	ALOOC	-0.08	0.79	0.29	0.18	0.71	0.31
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	-0.01	0.98	0.02	0.2	0.98	0.09
defocus blur (5)	ALOOC	-0.11	0.81	0.29	0.17	0.71	0.29
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.01	0.98	0.02	0.2	0.98	0.1
elastic transform (2)	ALOOC	-0.08	0.79	0.29	0.18	0.71	0.32
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.04	0.96	0.02	0.2	0.98	0.13
elastic transform (5)	ALOOC	-0.09	0.8	0.29	0.18	0.71	0.31
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.04	0.96	0.02	0.2	0.98	0.13
glass blur (2)	ALOOC	-0.07	0.78	0.29	0.18	0.71	0.33
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.02	0.97	0.02	0.2	0.98	0.11
glass blur (5)	ALOOC	-0.11	0.82	0.29	0.17	0.71	0.28
	OOB	0.01	0.83	0.16	0.2	0.84	0.29
	ODIN	0.06	0.95	0.02	0.2	0.98	0.14
zoom blur (2)	ALOOC	-0.09	0.8	0.29	0.18	0.71	0.31
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.01	0.97	0.02	0.2	0.98	0.1
zoom blur (5)	ALOOC	-0.11	0.82	0.29	0.17	0.71	0.28
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.02	0.97	0.02	0.2	0.98	0.11
gaussian blur (2)	ALOOC	-0.11	0.82	0.29	0.17	0.71	0.28
	OOB	0.0	0.84	0.16	0.2	0.84	0.28
	ODIN	0.02	0.97	0.02	0.2	0.98	0.11
gaussian blur (5)	ALOOC	-0.17	0.86	0.29	0.17	0.71	0.24
	OOB	0.0	0.84	0.16	0.2	0.84	0.27
	ODIN	0.06	0.95	0.02	0.2	0.98	0.14