Datasets used in the paper:
"Benchmarking Safety Monitors for
Image Classifiers with Machine Learning"

May 12, 2021

1 Datasets

Figure 1 gives a general illustration about the generated types of out-of-distribution (OOD) data.

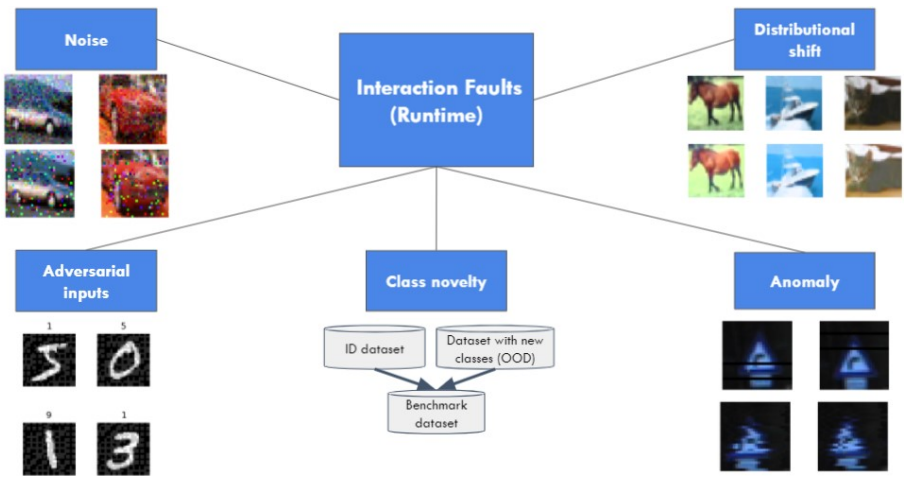

Figure 1: Example of OOD data generated by the framework.

Adversarial attack on images generated with the fast gradient sign method (FGSM). Novelty class datasets generated as ID dataset - OOD dataset. Anomalies based on common anomalies on sensors described in the report "Evaluating Functional Safety in Automotive Image Sensors, Semiconductor Components Industries" [a]. For distributional shifts, we applied simulations of weather conditions (snow, fog) and possible interference of the environment on the cameras such as brightness, contrast and saturation. We also generated rotated images, varying from -30 degrees to +30 degrees. For noise variations, we applied several types of corruptions in the pixels as can be seen in the Table 1 All the variations for distributional shift and noise (except for the rotated data) were proposed on the paper "Benchmarking neural network robustness to common corruptions and perturbations" [b].

Table 1: Dataset details.

Variation	OOD type	ID Instances	OOD Instances
FGSM	Adversarial Attack		
BTSC-GTSRB	Novelty Class		
GTSRB-CIFAR10	Novelty Class		
CIFAR10-GTSRB	Novelty Class		
Pixel Trap	Anomaly		
Row Add Logic	Anomaly		
Shifted Pixel	Anomaly		
Snow	Distributional shift		
Fog	Distributional shift		
Brightness	Distributional shift		
Contrast	Distributional shift		
Saturate	Distributional shift		
Rotated	Distributional shift		
Spatter	Noise		
Gaussian	Noise		
Shot	Noise		
Speckle	Noise		
Defocus Blur	Noise		
Elastic Transform	Noise		
Impulse	Noise		
Glass Blur	Noise		
Zoom Blur	Noise		
Gaussian Blur	Noise		

[a] `https://www.onsemi.cn/pub/Collateral/TND6233-D.PDF`

[b] Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261.