KI ABSICHERUNG
Safe AI for Automated Driving

Coverage Guided Fuzz Testing Framework and Dataset Quality Metrics

Vivek Vekariya, Mojdeh Golagha

Dataset Quality

Overview

The quality of the test dataset is very crucial for the trustworthiness of reported KPIs of DNN in safety-critical domains. Therefore, we propose the dataset quality metrics to evaluate the quality of the test dataset and a framework to sample the additional test data points using the coverage guided fuzz testing mechanism. The overall framework is as shown in figure 1.

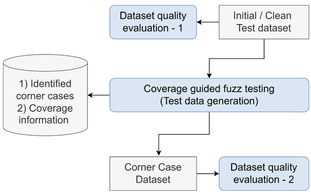

Figure 1: Overview of dataset quality metrics and fuzz testing framework

Dataset Quality Metrics

We propose three metrics [1] namely:

1. Equivalent Partitioning (EP - measures biasness in the dataset),
2. Boundary Conditioning (BC - measures % of samples in weak classification region),
3. Centroid Positioning (CP – measures % of samples into train class cluster). CP is calculated using latent space vector. Therefore, latent space coverage can be inferred from CP metric.

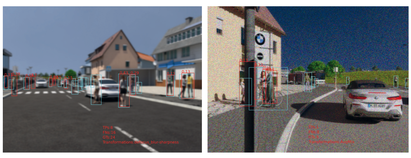

Figure 3: Sample corner case images

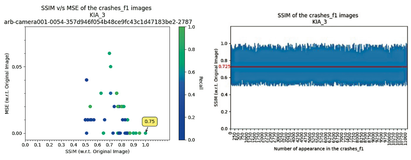

Figure 4: Quality of corner case images

Coverage guided fuzz testing framework

Fuzz testing framework [2] is used to generate corner case data by using various coverage criteria. Here, the focus is on the internal structure of DNN. The effectiveness of the fuzz testing lies in the corner case data which leads to the latent space maximization.

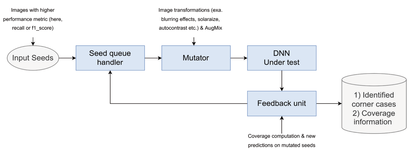

Figure 2: Overview of various components of fuzz testing framework

Evaluation

We evaluated TP1 Opel SSD & TP3 VW Robust SSD on TP2 tranche 5 test data. An augmented test image is classified as corner case when it has f1_score < 0.5, recall < 0.5 or hybrid objective function = (α * recall) + (1-α) * precision < 0.5.

Results

The identified corner cases have high structural similarity w.r.t. the original images. The decreasing CP value also shows that the identified corner cases leads to the latent space maximization.

Model	Test Dataset	Centroid Positioning*		
		Original Dataset	Corner Cases (Image trans.)	Corner Cases (AugMix)
Opel SSD	Tranche 5	0.669	0.320	0.518
VW Robust SSD	Tranche 5	0.732	0.596	0.615

* Ideally, CP metric close to 0 is expected for a test dataset well distributed in the latent space of DNN.

References:

[1] Coverage Testing of Deep Learning Models using Dataset Characterization (https://arxiv.org/abs/1911.07309)
[2] DeepHunter: a coverage-guided fuzz testing framework for deep neural networks (https://doi.org/10.1145/3293882.3330579)

fortiss

For more information contact:
vekariya@fortiss.org
golagha@fortiss.org

Images Sources:
Figures 1/2/4: Fortiss GmbH; Figure 3: BIT-TS/Fortiss GmBH

KI Absicherung is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag