Presentation on *DeepTest: Automated Testing of Deep-Neural-Network-driven Autonomous Car*

Presented by : Ji Gao

[1]Department of Computer Science, University of Virginia
https://qdata.github.io/deep2Read/

August 26, 2018

DeepTest: Automated Testing of Deep-Neural-Network-driven Autonomous Cars

- DNN is popular. DNN based autonomous car is popular.
- DNN still get erroneous behaviors: Accidents happen.
- Existing testing techniques: manual collection of test data. They miss fatal corner cases.
- Need for automatically detecting erroneous behaviors of DNN-driven vehicles

Motivation: Previous work

- Research problem: Build automatic and systematic ways to test DNN-based autonomous cars.
- Issue: DNN is different from traditional software.
  - According to previous paper, the logic flow in DNN is not encoded in the training code. Therefore, traditional branch or code coverage.
  - Current Satisfiability Modulo Theory (SMT) solvers can't handle such formulas involving floating-point arithmetic and highly nonlinear constraints
  - Several research projects try to test DNN, but doesn't scale well to real-world-sized DNNs. [Verification papers, Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks.]

Neuron coverage

- Systematic way to partition the test space into equivalence classes: neuron coverage.
- Neuron coverage: $\frac{\#\text{activated neurons}}{\#\text{total neurons}}$
- Activated :=? (Output $>$ Threshold)
- Work on the simple DNN in the DeepXplore paper.
- CNN: Take an average on the output of all filters.
- RNN: Unrolling all the layers.

Increasing coverage

- Generating arbitrary inputs that maximize neuron coverage may not be very useful if the inputs are not likely to appear in the real-world.
- Nine different realistic image transformations (changing brightness, changing contrast, translation, scaling, horizontal shearing, rotation, blurring, fog effect, and rain effect)
- Justify by experiment result

Combining Transformations to Increase Coverage

- Greedy search to generate sample that maximize neuron coverage.

Algorithm 1: Greedy search for combining image tranformations to increase neuron coverage

Input : Transformations T, Seed images I
Output : Synthetically generated test images
Variable : S: stack for storing newly generated images
Tqueue: transformation queue

```
Push all seed imgs ∈ I to Stack S
genTests = ϕ
while S is not empty do
    img = S.pop()
    Tqueue = ϕ
    numFailedTries = 0
    while numFailedTries ≤ maxFailedTries do
        if Tqueue is not empty then
            T1 = Tqueue.dequeue()
        else
            Randomly pick transformation T1 from T
        end
        Randomly pick parameter P1 for T1
        Randomly pick transformation T2 from T
        Randomly pick parameter P2 for T2
        newImage = ApplyTransforms(image, T1, P1, T2, P2)
        if covInc(newimage) then
            Tqueue.enqueue(T1)
            Tqueue.enqueue(T2)
            UpdateCoverage()
            genTest = genTests ∪ newimage S.push(newImage)
        else
            numFailedTries = numFailedTries + 1
        end
    end
end
return genTests
```

Creating a Test Oracle with Metamorphic Relations

- Metamorphic relations: Necessary properties of the system or function to be implemented.
- if a DNN model infers a steering angle θ_o for an input seed image I_o and a steering angle θ_t for a new synthetic image I_t, which is generated by applying the transformation t on I_o , one may define a simple metamorphic relation where θ_o and θ_t are identical
- In their case, they didn't have the only solution.
- What they do is:

$$(\hat{\theta}_i - \theta_{ti})^2 \leq \lambda \text{MSE}_{orig}$$

Experiment

- 1. Do different input-output pairs result in different neuron coverage?
- Categorize inputs in different categories.
- Calculate Spearman rank correlation between neurons coverage and steering angle
- Result: Correlation exists.

Experiment

- 2. Do different transformations activate different neurons?
- Measure the dissimilarities between N1 and N2 by measuring their Jaccard distance:

$$1 - \frac{|N_1 \cap N_2|}{|N_1 \cup N_2|}$$

- Result: They activate different neurons.

Experiment

- 3. Can neuron coverage be further increased by combining different image transformations?
- Two ways: Cumulative/ Their greedy guided search
- Result: Combining image transformations does increase the neuron coverage.

Experiment

- 4. Can we automatically detect erroneous behaviors using metamorphic relations?

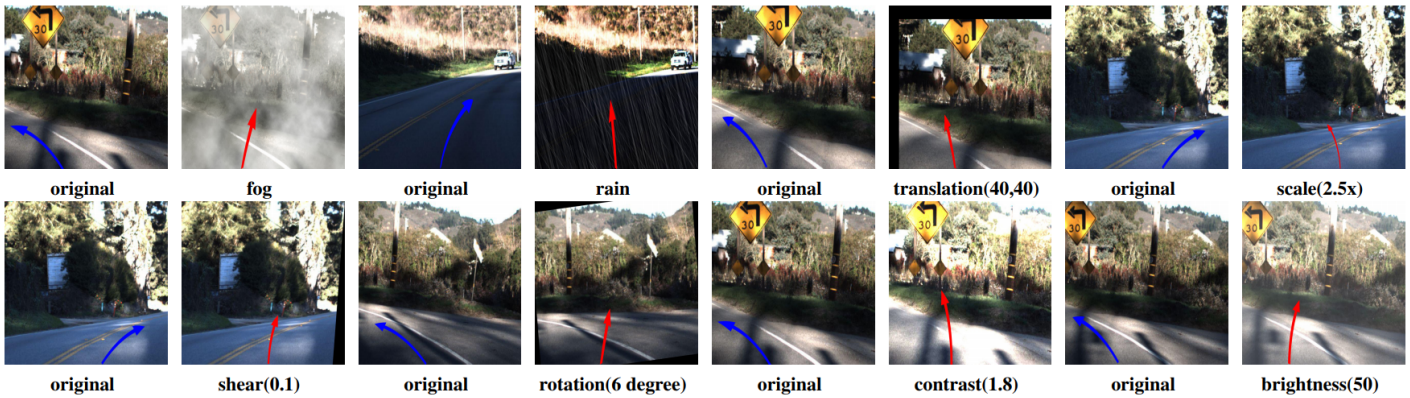

Figure 7: Sample images showing erroneous behaviors detected by DeepTest using synthetic images. For original images the arrows are marked in blue, while for the synthetic images they are marked in red. More such samples can be viewed at https://deeplearningtest.github.io/deepTest/ **(anonymous link).**

Retraining

- Accuracy of a DNN can be improved by up to 46% by retraining the DNN with synthetic data generated by DeepTest.