Advanced Testing of Deep Learning Models: Towards Robust AI

Summer Semester - 2024

Vivek V. Vekariya
Prof. Dr. Alexander Pretschner

Lehrstuhl für Software and Systems Engineering
Technische Universität München
08.02.2024
The world of AI testing

- DNN Testing: How to identify corner cases?
- Dataset Quality Evaluation: How to evaluate dataset quality for safety assured performance?

Train & Val Data → DNN → Test Data Evaluation → Deployment in Real-World

- Hidden defects of DNN (Corner Cases)
- Test Dataset

- 1. Hand Picked Test set
- 2. Inadequate test dataset
- 3. Statistical measure about the quality of test dataset?
Exploring Latent Space Coverage

- **Dataset Quality Aspects:**
  - Robust test dataset: e.g. Accuracy- 0%
  - Diverse test dataset: Test more underlying faults

- **Latent Space Coverage:**
  - Coverage, Density & Sparsity Estimation
    - Verify training policies
    - Estimate potential data collection gap
Exploring Latent Space Coverage

- **Directly using Latent space vectors:**
  - GANs & VAEs

- **Corner Case Identification:**
  - Coverage-guided Fuzz Testing
  - Latent Space based Testing
  - Metamorphic Relation Testing

Is this a true maximization of latent space coverage?

Dense and Sparse test data points in Latent Space

Ideal test data points in latent space
Coverage-Guided Fuzzing

Input seeds → Seed Queue Handler → Mutator → Unit Under Test

Feedback Unit ➔ 1) Failed test
 ➔ 2) Coverage Information
Learning Outcomes

- **Implementation, testing & evaluation** of state-of-the-art Classification & 2D Object Detectors DNNs
- Corner Case data generation using fuzzing and latent space properties
- GANs & VAEs for latent space coverage maximization
- Adversarial Attacks for state-of-the-art Classifiers and 2D Object Detectors
Prerequisites

**Required**

- Python (of course 😊)
- Deep Learning Frameworks (PyTorch, Keras, TensorFlow)
- Linux / Windows

**Good to have**

- Insights of 2D Object Detector Networks (SSD, Yolo, RCNN)
- Understanding of latent space and vector space modelling
- Passion for Safe AI

...But every smart work requires **sincere dedication & commitment**! 😊
Agenda

- **Pre-course Meeting**: 05.02.2024 and 08.02.2024
- **Apply with additional documents**: till 15.02.2024
- **Acceptance Notification**: 23.02.2024
- **Kick–off Meeting - 1**: 18.04.2024 (Do.)
- **Project Discussions & Allocation**: 25.04.2024 (Do.)
- **Weekly Follow-ups**
- **Mid-term Presentations**: 30.05/06.06.2024 (Preliminary-Do.)
- **Final Presentations**: July.2024 (Preliminary-Do.)
Evaluation

- *We work in TEAMs & get evaluated based on TEAM*
- *Peer Reviews for code and merge requests (Let’s Learn Together)*
- **Evaluation Criteria & Deliverables:**
  - Code & results (5/10)
  - Team & Individual reports (3/10)
  - Final Presentation (2/10)
  - Bonus: innovative ideas & extensive evaluation of the approaches
1. Give your 1\textsuperscript{st} priority to this course in the matching system
2. Tell us more about you (motivation, CV, transcripts & Gitlab link) by filling out:  
   \url{TUM_I4_student_wiki}
Thank you for your attention 😊

Vivek V. Vekariya
Garching bei München