Advanced Testing of Deep Learning Models: Towards Robust AI

Winter Semester 2025-26 (Pre-Course Meeting)

Vivek V. Vekariya

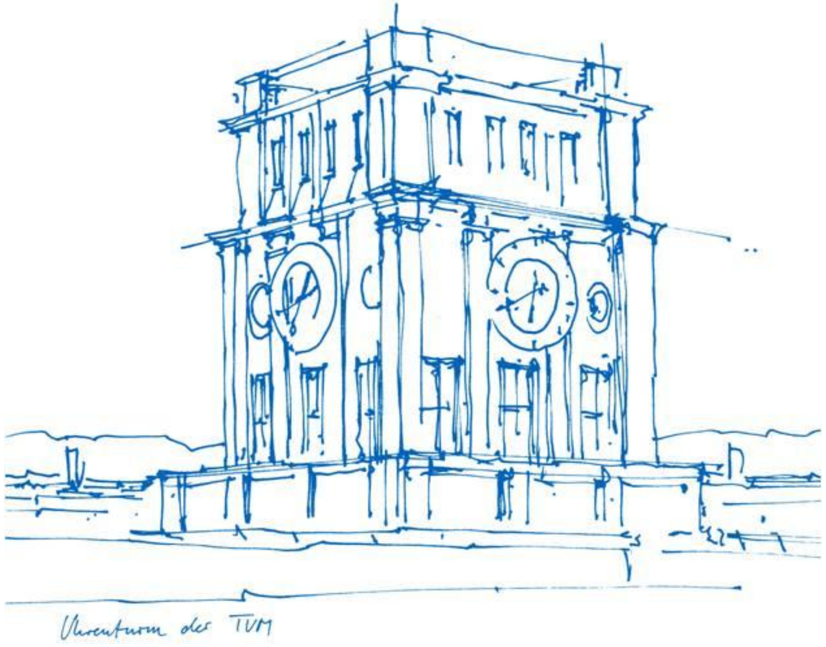
Simon Speth

Prof. Dr. Alexander Pretschner

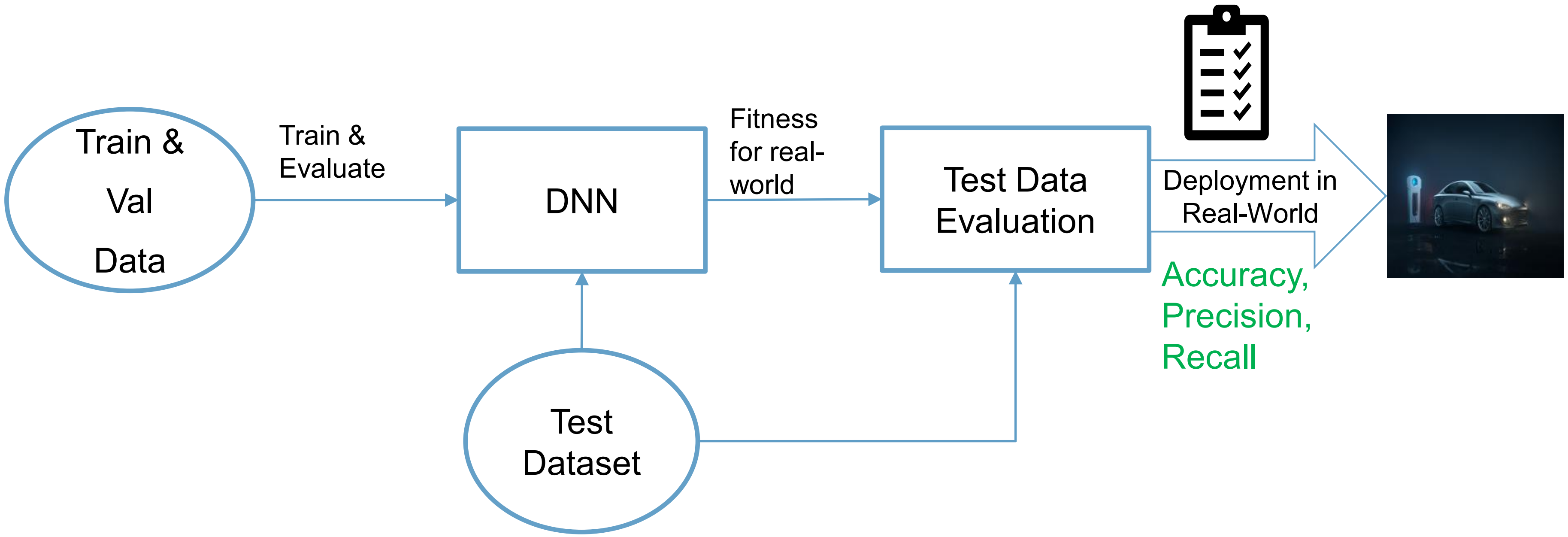
Chair of Software and Systems Engineering

Technical University of Munich

16.07.2024



Deep Learning Application Development



Train & Val Data

Train & Evaluate

DNN

Fitness for real-world

Test Data Evaluation

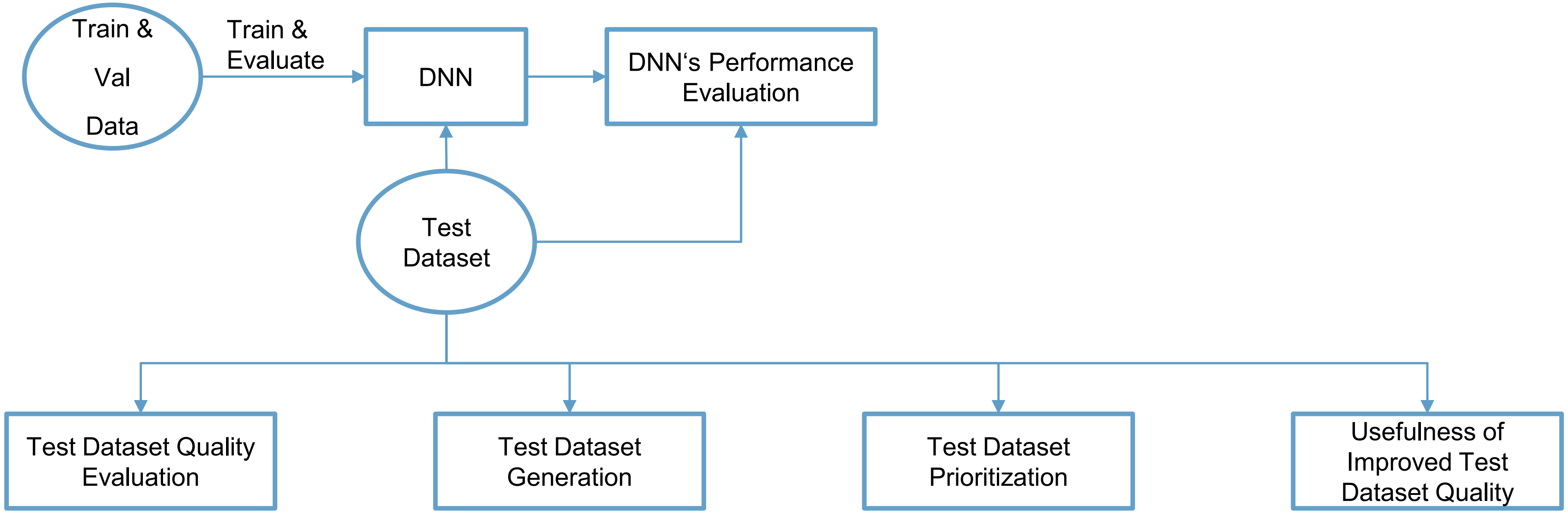
Deployment in Real-World

Accuracy, Precision, Recall

Test Dataset

- Performance of Deep Neural Networks (DNNs) is reflected in evaluation metrics on test dataset
- Overestimation of DNN's capabilities in real-world applications

My Research Areas



Train & Val Data

Train & Evaluate

DNN

DNN‘s Performance Evaluation

Test Dataset

Test Dataset Quality Evaluation

Test Dataset Generation

Test Dataset Prioritization

Usefulness of Improved Test Dataset Quality

Test Dataset Quality Evaluation

1. New metric aware of latent space – LSCD
2. Computationally efficient alternative to the Mutation Score

Test Dataset Generation

1. Use LSCD in coverage-guided fuzzing
2. LSCD is the most effective and efficient coverage criteria during fuzzing

Test Dataset Prioritization

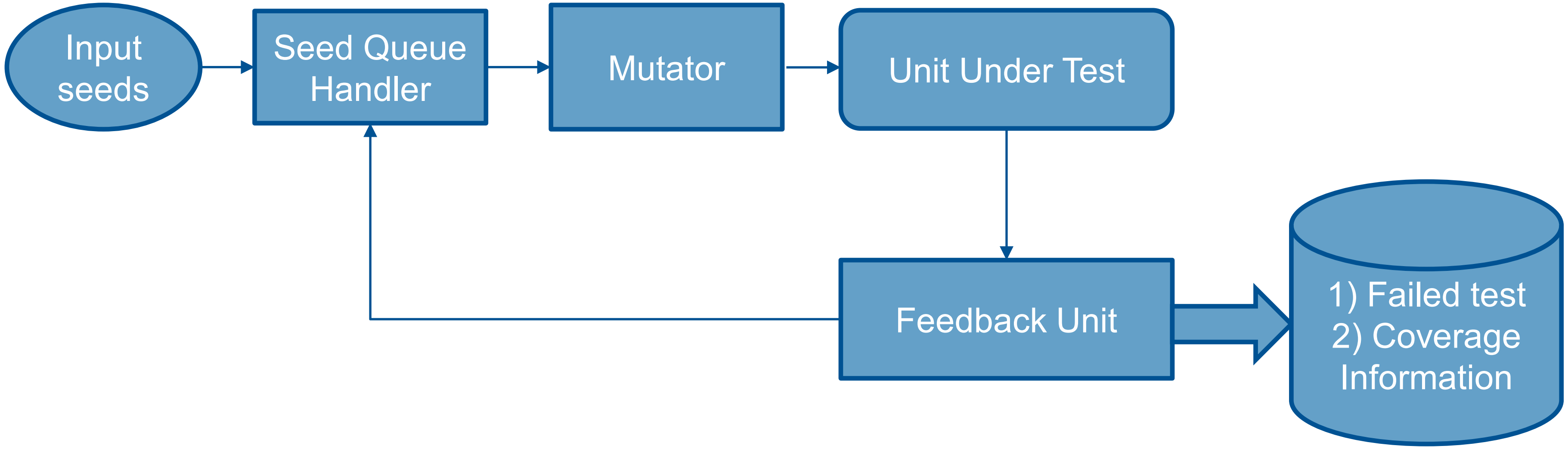
1. Propose advanced methods to partition the latent space
2. Use it to prioritize test suites
3. Compare it with random prioritization, DeepGini

Usefulness of Improved Test Dataset Quality

1. Use fuzzing data for re-training
2. How to use LSCD for robustness evaluation
3. Navigate Neural Maze- find limitation of DNN‘s performance

Ongoing

Coverage-Guided Fuzzing



Metamorphic Testing

- Metamorphic Testing (MT) is one method to solve the oracle problem for Deep Learning Models
    - There are usually no oracles for DL models
    - Metamorphic testing can be seen as a pseudo-oracle/model
    - Reverse engineering of a part of the specification

- Metamorphic Relations (MR) need to be defined in order to compute test cases
    - Source test inputs are used to compute follow-up inputs
    - Both inputs (source and follow-up) are fed into the System Under Test (SUT)
    - Both outputs and both inputs are compared to check whether the MR holds true

Metamorphic Testing

- **Example:** Testing the implementation of the $\sin(x)$ function
- **Assumption:** We implement a test case $\sin(2)$ but don't know what the correct output
- **Metamorphic Testing:** Creation of a *follow-up test case* $\sin(2 + 2\pi)$ which is expected to have the same output as the *source test case* $\sin(2)$
- **Test Case Evaluation:** We check if the relation $\sin(2) = \sin(2 + 2\pi)$ holds. If yes, the test case *passed*

$$\begin{tikzcd}
 & \text{Inputs} & \text{Outputs} \\
\text{sin() Example:} & 2 \arrow[r] \arrow[d] & sin(2) \arrow[d, leftrightarrow, "="] \\
 & 2 + 2\pi \arrow[r] & sin(2 + 2\pi)
\end{tikzcd}$$

What‘s new this year?

- 🧠 **Test Case Generation for LLMs**
  - Developing diverse prompts to rigorously evaluate Large Language Models.
- ✅ **LLM Test Dataset Quality**
  - Ensuring the integrity and effectiveness of data used for LLMs.
- 🚙 💭 **Automated Driving Use Cases**
  - Applying existing methodologies to expand capabilities (e.g., advanced steering angle prediction).

Evaluation

- *We work in TEAMs*
- ***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

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) and Large Language Models
- Understanding of latent space and vector space modelling
- Passion for Safe AI

….But every smart work requires sincere dedication & commitment! 😉

Agenda

- **Pre-course Meeting:** 16.07.2025
- **Apply with additional documents: till 22.07.2025**
- **Acceptance Notification:** 31.07.2025
- **Kick–off Meeting - 1:** XX.10.2025 (Mo. / Tue.)
- **Project Discussions & Allocation:** XX.10.2025 (Mo. / Tue.)
- ***Weekly Follow-ups***
- **Final Presentations:** Feb.2026 (Preliminary-Do.)

Interested?

1. Give your 1st priority to this course in the matching system
2. Tell us more about you (motivation, CV, transcripts & Gitlab link) by filling out:

   TUM_I4_student_wiki

Thank you for your attention 😇

Vivek V. Vekariya

16.07.2025

Garching bei München