Deep Learning Projects

Introduction

The goal of these projects is to design a deep learning methodology to solve a task on a provided dataset. This involves:

- Implementing methods to load and preprocess the data
- Implementing and training several relevant models and/or one model with several parameter sets
- Computing relevant metrics and analyzing the results
- Visualizing results

Choose one of the projects proposed in the next sections.

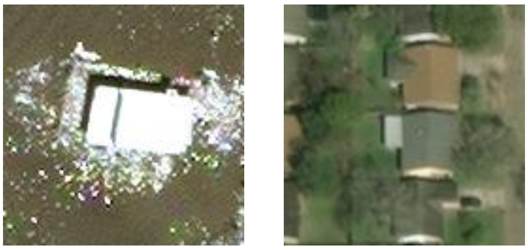
Specific expected outcomes are defined for each project, and the following criteria are used for **evaluation**:

Criterion	**Explanation**	**Points**
Method	- Relevant choice of methods - Several methods are compared - Decisions are justified	5
Reproducibility, code	- Provided code can be run easily - A readme file with instructions is provided	2
Evaluation of results	- Several relevant metrics are computed over train / val splits - Computational complexity (e.g. training time, inference time) is assessed	3
Report	- Report is clearly written - Graphics are readable and complete (e.g. axis titles) - Results are analysed - Discussion in terms of scientific (domain) output - Limitations are clearly discussed	2
TOTAL		**12**

Table 1: Evaluation criteria

4 Hurricane Damage Detection with Deep Learning

Task. After natural disasters such as hurricanes, rapid and reliable assessment of damaged areas is essential for coordinating response and recovery efforts. In this project, you will develop a deep learning model to automatically classify satellite image patches as damaged or undamaged based on post-hurricane observations. The task focuses on evaluating how well deep learning models can detect hurricane-induced damage patterns from optical imagery.
While achieving high accuracy is important, emergency response systems require models that are not only effective but also *trustworthy*. A model that confidently predicts the wrong outcome can be more harmful than one that acknowledges its limitations. Therefore, beyond building a classifier, you will evaluate the model's **calibration**, that is, how well predicted probabilities reflect actual correctness. A well-calibrated model should, for instance, be correct about 70% of the time when it assigns 70% confidence to its predictions. This project thus aims to explore not only what the model sees, but also how accurately its confidence aligns with reality.
For a short introduction to model calibration, see this tutorial.



(a) Damaged (b) Undamaged

Figure 5: Comparison of damaged vs undamaged buildings from satellite imagery.

Data. The dataset is based on the version provided by Cao et al. [1] and can be downloaded here.

- **Images:** RGB satellite patches of areas in Texas after Hurricane Harvey, captured by the GeoEye-1 satellite.
- **Classes:** *damage* and *no damage*.
- **Splits:** 19,000 training images (unbalanced), 2,000 validation images, and 2,000 test images (both balanced).
- **Data quality.** The validation and test sets include visually ambiguous or lower-quality samples compared to the training set, representing more realistic and noisy post-disaster conditions.

Challenges.

- High intra-class variability, "damage" can manifest in many ways (e.g., roof loss, flooding, vegetation destruction).
- Some image patches are visually ambiguous, and differences in lighting or sensor conditions introduce additional noise.
- Domain shift between training and test events.

Expected Outcome. Students are expected to develop a deep learning model for detecting hurricane-related damage from satellite imagery and evaluate its performance using standard classification metrics such as accuracy, and F1-score. Beyond these metrics, students should assess how well the model's predicted probabilities correspond to actual outcomes, its *calibration*, and explore ways to improve it where appropriate. This may involve analyzing reliability diagrams or related visualizations to examine whether the model tends to be over- or under-confident. The goal is to develop a calibrated model whose confidence values can be meaningfully interpreted, supporting more reliable decision-making in emergency response settings.

References. [1] Cao, Quoc Dung, and Youngjun Choe. "Detecting damaged buildings on post-hurricane satellite imagery based on customized convolutional neural networks." IEEE Dataport 1.206 (2018): 1-19.