Benchmarking Review, Standardization, and Exploratory Analysis of Public Remote Sensing Disaster Datasets

Project Description

This project begins with exploring recent benchmarking frameworks, commonly used annotation formats, and current practices in disaster-related remote sensing datasets. The intern will review relevant datasets and documentation to understand annotation standards (e.g., segmentation masks, bounding boxes, GeoJSON, and raster labels) and how these datasets are used in modern machine learning and disaster assessment workflows. Following this, the project focuses on compiling, organizing, and analyzing publicly available remote sensing datasets related to natural disasters such as floods, wildfires, earthquakes, and hurricanes. The intern will download selected datasets, explore their annotation formats and metadata structures, and reorganize them into a standardized and well-documented repository. The project also includes performing preliminary statistical analysis to summarize dataset characteristics such as image counts, class distributions, spatial coverage, and annotation types. The final outcome will be a structured dataset repository, metadata catalog, and technical documentation to support future research and development efforts.

Project Type: Engineering

Internship Batch: **Batch 1**: May 10 to July 9, suitable for Education City students, i.e., CMUQ, TAMUQ and HBKU students

Duties/Activities:
- Review recent disaster-related remote sensing datasets, benchmarking models, and annotation formats.
- Download and manage datasets. Explore RS dataset structures, annotation formats, and metadata
- Design and implement a standardized directory and metadata schema
- Reorganize datasets into a consistent structure and perform descriptive statistical analysis.
- Document processing workflows and prepare a final report

Required Skills:
- Proficiency in Python and experience with Linux command line
- Basic data analysis skills
- Familiarity of geospatial data
- Familiarity with file handling and scripting

- Familiarity with a No-SQL Document Store such as Elastic or MongoDB or Opensearch and rich Database visualization tools such as Kibana orApache Superset or Atlas Charts etc.

Preferred Intern Academic Level: Senior undergraduate student or Master's student

Learning Opportunities:
- Exposure to geospatial data formats and annotation standards
- Practical data engineering and dataset curation experience
- Understanding of modern benchmarking datasets and annotation standards in disaster remote sensing
- Hands-on experience with real-world remote sensing disaster datasets
- Development of reproducible data processing workflows
- Experience in metadata standardization and dataset documentation
- Foundation for future research in disaster monitoring and machine learning applications

Expected Team Size: 2 interns

Mentors:
Noora Al-Emadi | nalemadi@hbku.edu.qa
Anas Al-Nuiami | analnuaimi@hbku.edu.qa
Ferda Ofli | fofli@hbku.edu.qa
Mumhammed Imran | mimran@hbku.edu.qa