A Fast and Scalable Deep Learning Solution for Large-Scale Disaster Damage Mapping from Multi-Sensor Satellite Imagery

Nicolas Dublé nicolas.duble@thalesgroup.com, Amélie Rougé amelie.rouge@thalesgroup.com, Thales Services Numériques, Labège, France

AI4Earthquake Challenge Φ-lab Challenges

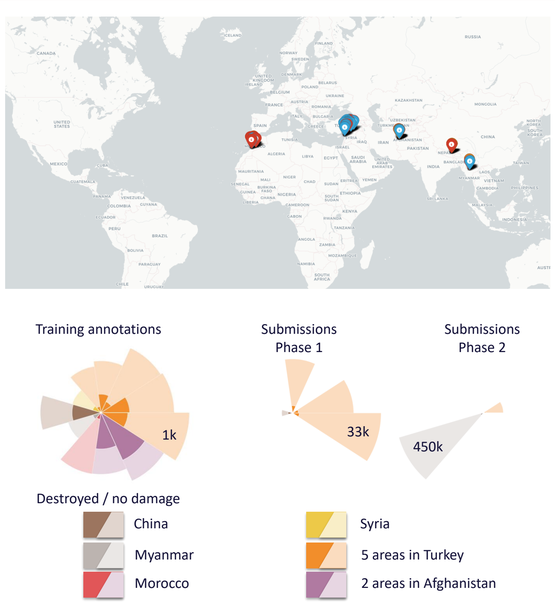
Context

Earth Observation is essential for rapid disaster response by providing information on affected regions. After major natural disasters, fast and reliable building damage assessment is critical for emergency coordination and recovery planning. The International Charter “Space and Major Disasters” enables rapid access to multi-source satellite imagery within hours of an event. These data support large-scale situational awareness over disaster-affected areas.

However, operational **EO data are highly heterogeneous**, with **differences in sensors**, **resolution**, **geometry**, and **radiometry**, making **fast and robust automated analysis challenging**. In this context, ESA Phi-Lab and Disaster Charter proposed a challenge in 2025 to build machine learning models capable of rapidly identifying post-earthquake building damage using EO data.

The competition was conducted in a two-phase setup. During the first phase, participants could train and test their models on partially annotated scenes. In the second phase, participants were required to process a new set of completely unannotated scenes within a limited time frame.

Real data

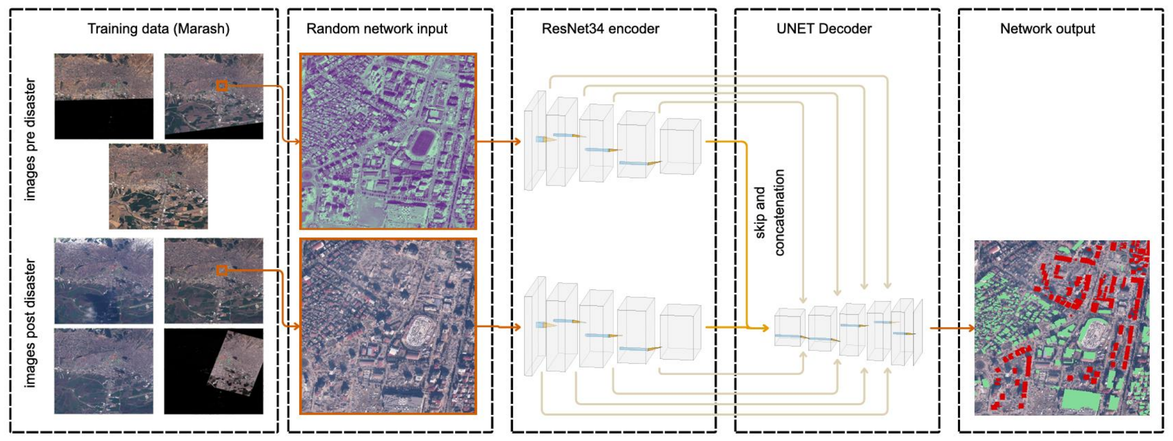


The dataset is based on **International Charter activations covering major earthquakes from 2020 to 2025**. It includes multi-mission very high-resolution optical imagery acquired under operational conditions. Sensors include Pléiades-1A/1B, WorldView-2/3 and GeoEye-1, BlackSky, Gaofen-2, and Kompsat-3.

Images are provided as Cloud Optimized **GeoTIFFs** (COG), with both 16-bit radiometrically calibrated reflectance images or 8-bit RGB products. Due to operational constraints, images are not systematically co-registered, **leading to spatial misalignment between acquisitions**.

Building annotations are provided as polygons in GeoPackage format, mainly derived from **OpenStreetMap** and annotated manually in selected regions. Labels contain damaged and intact buildings.

Methods



We use a **lightweight Siamese CNN** for building damage detection from pre- and post-event imagery. The model processes **two RGB inputs** and learns a similarity representation. A **shared convolutional backbone** extracts features from both images. Feature differences are fused and passed through a sigmoid layer for pixel-wise scoring to predict a **continuous damage score in [0,1]**.
Building footprints are used to aggregate pixel predictions at object level, producing a single damage score per building.
Training is performed on heterogeneous multi-sensor data with standard augmentation to improve robustness. No test-time adaptation is used to ensure fast inference in operational settings.

Results

Models were ranked using instance **F1-score evaluation among a set of damaged and intact buildings**.

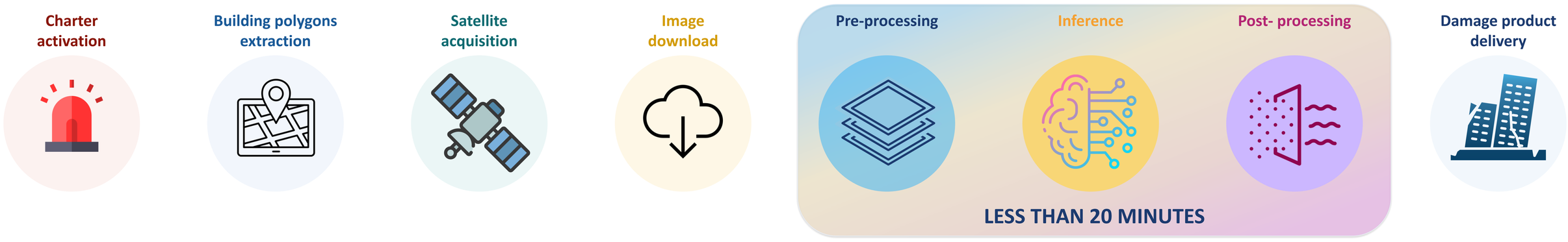
The Antakya area was partially annotated during training and used for the Phase 1 evaluation, leading to performance levels comparable to those obtained in Phase 1.

The **Mandalay** (Myanmar) area is a **completely unseen scene**, with distinct visual characteristics and a different damaged/undamaged building distribution, resulting in lower overall scores.

Rank	Team	Mean F1	Antakya	Mandalay
1	**Badaboum (ours)**	**0.5710**	0.7135	**0.4285**
2	**DisasterM3**	0.5382	0.7463	0.3300
3	**TelePIX**	0.4232	**0.7750**	0.0715
4	**Datalayer**	0.4046	0.7179	0.0913
5	**Blackpink**	0.3466	0.6931	0

Table : Results on phase 2 results evaluation

FROM CHALLENGE TO OPERATIONS



Discussion

Several key aspects arise from the competition setup and results. First, clear performance discrepancies were observed between well-represented regions such as the Middle East and underrepresented areas such as Southeast Asia, highlighting limitations in training data diversity. Retraining on a more geographically balanced dataset, as well as increasing sample size, would **improve 1. robustness and 2. the reliability of evaluations.**

Second, leveraging the OSM annotations available at the time of the challenge provided a valuable baseline, but before real-world utilization, it would be interesting to **evaluate deeper the available annotations at the time of the concerned disasters**. Also, post-hoc annotation corrections could further improve evaluation consistency.

Finally, **the challenge focused exclusively on earthquake disaster management**, a particularly **complex use case**. Extending the analysis to other International Charter events such as floods, volcanic eruptions, and storms would improve the generality of the conclusions.

Conclusions and perspectives

So far, Siamese approaches have demonstrated strong capabilities, reaching the top of the leaderboard while effectively leveraging building polygons during post-processing. The proposed fully convolutional Siamese neural network also showed excellent scalability for rapid large-scale damage assessment, enabling the **processing** of the entire **600 km² urban area** of Mandalay, comprising approximately **450.000 buildings**, **in less than 20 minutes.**

Operational requirements could motivate **extending the Siamese approach to multi-temporal image sequences**, allowing the exploitation of successive satellite acquisitions and supporting **continuous updates** of building damage maps throughout the disaster response cycle.

Furthermore, the significantly better performance achieved on partially annotated scenes suggests that an **active learning strategy**, developed jointly with photo-interpretation experts, could **streamline the labeling process while improving model performance**.



Key takeaway

Multi-sensor ready
Harmonized results with multi-sensor images.

Lightweight and fast deployment
Siamese CNNs provide efficient pre- and post-event change detection. The models can be trained within a few days, while inference time is on the order of minutes.

Robust ranking-based damage assessment
A ranking-based selection of destroyed buildings improves robustness compared to threshold-based detection methods.

Nicolas DUBLÉ, Thales Services Numériques
nicolas.duble@thalesgroup.com
Toulouse, France