

RECAP: Rapid Event-level Classification of Affected Properties

Project Proposal

Problem Statement and Motivation

After a major disaster, responders need a quick, trustworthy view of which buildings are safe, damaged, or destroyed. Field surveys are slow and risky, while raw satellite images are not directly actionable. The gap is a tool that converts paired pre- and post-disaster satellite images into building-level damage assessments with usable confidence.

Our goal is to build a compact, end-to-end prototype. The system will train on the public xView2 dataset, predict one of four damage levels (no, minor, major, destroyed), calibrate probabilities for predictable thresholds, and present results on an interactive Streamlit map. For class demos, we will precompute predictions for one or two events so the interface runs smoothly on a standard laptop.

Related Work (brief)

- **xBD / xView2**: Paired pre/post imagery with building polygons and ordinal damage labels; standard benchmark for automated building damage assessment (Gupta et al., CVPRW 2019).
 - **Siamese change detection**: Shared-encoder Siamese networks that learn from image pairs and are effective for change classification; a practical baseline for our task (Daudt et al., ICIP 2018).
 - **CrisisMMD**: Disaster-related tweets with annotations; motivates an optional late-fusion re-ranker on top of image predictions (Alam et al., ICWSM 2018).
-

Initial Hypotheses

- **H1:** A simple image-only Siamese ResNet-18 will achieve macro-F1 ≥ 0.60 on held-out events; most errors will be confusions between “minor” and “major” damage.
 - **H2:** Temperature scaling will reduce Expected Calibration Error (ECE) by $\geq 30\%$ compared to the uncalibrated baseline, producing steadier threshold behavior.
 - **H3 (Stretch):** A lightweight re-ranking step using crisis tweet embeddings for one event will improve Top-K recall for “major” and “destroyed” classes.
-

Goals and Scope

- **Data preparation:** Generate 224–256 px pre/post building chips. Split by event to prevent leakage and track class balance.
- **Baseline model:** Train a Siamese ResNet-18 with class-weighted cross-entropy and basic augmentations; report per-class precision, recall, and F1.
- **Calibration and explainability:** Apply temperature scaling; report calibration curves and ECE. Produce Grad-CAM overlays for qualitative inspection.
- **Demo application:** Build a Streamlit map with color-coded buildings, event selector, and confidence slider. On click, show pre/post chips, predicted label, confidence, and optional Grad-CAM. Include CSV and GeoJSON export. Use precomputed predictions for smooth demos.
- **Stretch goals (time permitting):** Add an inspection queue ranked by severity \times confidence. Explore tweet-based late fusion for one event.
- **Out of scope:** No live satellite tasking or operational claims; this is a reproducible classroom prototype, not a production service.