Project Proposal: Disaster Damage Project (DDP)

Problem Statement and Motivation

After a major disaster, responders urgently need to know which buildings are safe, damaged, or destroyed. Field surveys are slow, dangerous, and resource-intensive, often taking days when decisions are needed in hours. Satellite imagery is available rapidly, but raw pre- and post-event photos are not directly actionable. Human analysts can interpret imagery, but scaling this work is difficult during emergencies.

The gap is a tool that can automatically convert pre- and post-disaster satellite image pairs into actionable, building-level assessments. A successful system should assign each building a damage category (no, minor, major, destroyed), provide calibrated confidence scores, and present the results in an intuitive map interface. This would allow emergency managers to quickly filter for high-risk buildings, generate inspection queues, and download results for use in planning.

Our project will build a small end-to-end prototype. The core deliverable is an image-only model trained on public xView2 data, paired with a minimal Streamlit application where users can interact with predictions on a map. We will focus on calibration and interpretability so that the system produces not only predictions but also confidence measures and visual explanations.

Related Work

- **xBD / xView2 Dataset**: Gupta et al. (CVPRW 2019) introduced the xBD dataset, which contains paired pre- and post-disaster imagery with building polygons and four ordinal damage labels. It has become the benchmark for automated disaster damage assessment and enables event-level splits for testing generalization.

- **Siamese Networks for Change Detection**: Daudt et al. (ICIP 2018) demonstrated that Siamese architectures with shared encoders improve classification on paired imagery tasks. This approach is widely used in xView2 baselines and provides a practical foundation for our prototype.

- **CrisisMMD**: Alam et al. (ICWSM 2018) released a dataset of disaster-related tweets labeled for information type. While our MVP is image-only, this motivates a stretch goal:

experimenting with simple text fusion to re-rank ambiguous predictions using social media signals.

Initial Hypotheses

- **H1.** A simple image-only Siamese ResNet-18 will achieve macro-F1 ≥ 0.60 on a held-out event split, with most confusions occurring between minor and major damage classes.
- **H2.** Applying temperature scaling to model logits will reduce Expected Calibration Error (ECE) by at least 30% compared to the uncalibrated baseline, resulting in more reliable thresholding in the demo app.
- **H3 (Stretch).** Adding a lightweight re-ranking step with tweet embeddings for one event will improve Top-K recall for the most severe damage categories.

Goals and Scope

- **Data Preparation**: Generate 224–256 px chips centered on building footprints from xView2. Split by event to avoid leakage and compute class balance.
- **Baseline Model**: Train a Siamese ResNet-18 on pre/post chips using class-weighted cross-entropy and basic augmentations. Track per-class precision, recall, and F1.
- **Calibration & Explainability**: Apply temperature scaling on validation logits and report calibration curves. Produce Grad-CAM overlays to highlight image regions that drive predictions.
- **Demo Application**: Build a minimal Streamlit app with a Leafmap/Folium map. Users will see color-coded building polygons, select events, adjust a confidence slider, and click to view pre/post chips, predicted labels, and optional Grad-CAM overlays. Export buttons will allow CSV/GeoJSON downloads.
- **Stretch Goals (time permitting)**: Add an inspection queue sorted by severity × confidence, and experiment with late fusion using tweet embeddings.
- **Out of Scope**: No live satellite tasking or operational claims. The deliverable is a reproducible classroom prototype, not a production service.