CITY
ST GEORGE'S
UNIVERSITY OF LONDON

MSc in Data Science

Project Report

2025

Advancing Building Damage Assessment from Satellite Imagery: Evaluating Two-Stage Deep Learning Pipelines for Automated Analysis

Sajjaad Jurawon

Supervised by:

Dr. Giacomo Tarroni

Date of submission: 22/09/2025

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the assessment instructions and any other relevant programme and module documentation. In submitting this work I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

Signed: Sajjaad Jurawon

ABSTRACT

The increasing frequency of natural disasters due to climate change necessitates automated systems for rapid building damage assessment from satellite imagery. This study develops and evaluates a two-stage deep learning pipeline that combines object detection for building localisation with classification models for damage severity predictions. Using the xBD dataset, we systematically compare state-of-the-art CNN and transformer architectures. YOLOv8s achieved superior building detection performance compared to Faster R-CNN and FCOS. CNN architectures (ResNet-50, EfficientNet-B3) demonstrated better adaptability to progressive fine-tuning than transformers (ViT-B/16, DeiT-B/16), with EfficientNet-B3 achieving the highest test accuracy (87.8%). Cross-Entropy loss outperformed focal and ordinal loss across all architectures. The integrated pipeline achieved an end-to-end F1-score of 0.503, processing 1.28 seconds per image while maintaining strong performance on extreme damage categories (no damage: 0.887, destroyed: 0.723). However, intermediate damage categories showed poor recall (minor: 0.214, major: 0.143), highlighting the challenge of sequential error propagation in two-stage architectures. This research provides empirical evidence comparing modern architectures for satellite-based disaster response and establishes reproducible baselines for future work.

Keywords: Satellite Imagery, Object Detection, Damage Classification, Transfer Learning, Two-Stage Pipeline

Table of Contents

1 Introduction and Objectives

The frequency and severity of natural disasters have risen dramatically in recent decades due to anthropogenic climate change. This shift has resulted in a marked increase in the occurrence and duration of extreme weather events worldwide (Proma et al., 2022; Morozov et al., 2023), manifesting in more frequent floods, droughts, hurricanes, storms, wildfires, and landslides (Morozov et al., 2023; Zhang et al., 2022). Historical data confirm this trend: for instance, climate disasters are reported to have increased from 1,171 in 1960–1979 to 6,641 in 2000–2019 (Chavez-Demoulin et al., 2021). The consequences of such disasters extend far beyond immediate human casualties, with built infrastructure suffering extensive damage, underscoring the need for rapid damage assessment for effective resource allocation and recovery planning. Traditional ground surveys and sensor-based methods are inadequate at this scale, creating a critical demand for automated, remote-sensing approaches based on satellite imagery and computer vision.

Computer Vision has revolutionized satellite imagery analysis, transforming static overhead images into rich sources of actionable information across diverse applications. Beyond building damage assessments, computer vision techniques contribute to broader satellite imagery analysis, including footprint extraction, urban planning insights and even infrastructure network mapping (Hoque et al., 2025; Maniyar et al., 2025; Wang et al., 2024). Key methodological approaches frequently employ deep learning architectures, notably Convolutional Neural Networks (CNNs) have enabled automated feature extraction from high-resolution satellite data, while advanced models like U-Net and transformer-based approaches have achieved precise pixel-level analysis for land use classification, object detection, and change detection (Dosovitskiy et al., 2021; Ronneberger et al., 2015**.** Recent advances in attention mechanisms, multimodal fusion, and transfer learning have expanded the capabilities of computer vision for satellite imagery analysis. Yet challenges remain in achieving robust generalization across regions and sensors, and in meeting the computational demands of real-time and large-scale processing. These limitations continue to drive innovation toward more efficient and scalable Earth observation solutions.

This study addresses these challenges by developing and evaluating a two-stage pipeline for automated building damage assessment from satellite imagery. The approach first applies object detection to localize individual buildings, followed by classification models to estimate damage severity. The study systematically evaluates state-of-the-art CNN and transformer architectures for both tasks through transfer learning, with attention to performance across disaster types and geographic contexts. Key aspects explored include the use of pre- and post-disaster imagery for change detection, the optimization of models for computational efficiency without sacrificing accuracy, and the design of evaluation metrics suited to the challenges of satellite-based assessment.

1.1 Research Questions

1. How effectively can state-of-the-art object detection architectures localize buildings in post-disaster satellite imagery?
2. How do modern CNN and transformer architectures compare for building damage classification from satellite imagery
3. What training strategies and loss functions optimise performance for imbalanced satellite imagery datasets for building damage assessment?
4. How effectively can optimised models be integrated into a two-stage damage assessment pipeline?

1.2 Purpose

The purpose of this research is to develop, evaluate and analyse critically a two-stage pipeline for automated building damage assessment from satellite imagery. By integrating object detection for building localisation with classification models for damage severity estimation, the study investigates and compare effectiveness of CNN and transformer architectures, as well as the impact of training strategies and loss functions on performance under class imbalance. Furthermore, the research examines how detection and classification perform within an integrated pipeline, with particular attention to error propagation effects.

1.3 Beneficiaries

The primary beneficiaries of this study are researchers in the field of AI, computer vision, remote sensing and disaster informatics. By providing systematic comparisons of CNN and transformer architectures, detailed evaluation of training strategies under class imbalances and an end-to-end assessment of sequential pipeline, the study establishes baselines and methodological insights that can inform future work.

Indirectly, the findings could benefit practitioners and organisations involved in developing operational disaster-response systems, the analysis of error propagation and computational efficiency highlights trade-offs relevant to real-world deployment.

1.4 Scope

The scope of this research is limited to the development and evaluation of machine learning approaches for building damage assessment from satellite imagery. Specifically, the study includes:

- **Object Detection:** Evaluating state-of-the-art models for localising individual buildings in post-disaster imagery
- **Damage Classification:** Comparing CNN and transformer architectures in assigning damage severity categories
- **Training Strategies:** Investigating the effects of loss functions on classification performance.
- **Pipeline integration:** Assessing the performance and limitations of a two-stage detection-classification framework.

The report is structured as follows:

1. **Introduction and Objectives:** We have outlined the motivation for automated building damage assessment from satellite imagery, defined the research purpose, and presented the guiding research questions, scope, and intended beneficiaries
2. **Context:** Reviews existing literature on disaster response, satellite imagery analysis, and computer vision methods, highlighting current approaches to building damage assessment and their limitations.
3. **Data:** Describes the dataset used in this study in detail, including its composition, annotation framework, disaster coverage, dataset splits, and inherent challenges.
4. **Methodology:** Outlines the proposed two-stage pipeline, beginning with building localisation using object detection models (Faster R-CNN, FCOS, YOLOv8) and followed by building-level damage classification (ResNet-50, EfficientNet-B3, ViT, DeiT). It further details data preparation strategies for building crop extraction, training procedures, and the loss function ablation study.
5. **Results:** Presents experimental findings, including baseline and fine-tuned performance of object detection models, classification results across different

architectures and loss functions and final test set evaluations. The section concludes with an end-to-end pipeline assessment integrating detection and classification.

6. **Discussion:** Interprets the experimental findings with respect to the research questions, including evaluation of object detection, CNN versus transformer classification, and the impact of class imbalance and loss functions. The section further analyses the integration of detection and classification within a two-stage pipeline, situates results in the context of existing literature, and reflects on methodological contributions, limitations, and directions for future research.
7. **Conclusion:** Summarises the overall findings of the study, highlighting the key contributions of the proposed two-stage pipeline for building damage assessment. The section reflects on the methodological and empirical insights gained and offers brief recommendations for future research directions.

1.5 Use of AI Tools

Generative AI tools have been used in limited capacity to refine the text by reducing repetitions and improve the flow the flow between paragraphs. The initial drafting and writing of the report were completed without the use of AI. For coding, AI tools were used to assist with debugging/syntax errors and to identify options to improve runtime performance. All research design, analysis, literature reviews, interpretation of results and final writing are of my own.

2 Context

This chapter reviews existing literature on building damage assessment using satellite imagery, summarising key approaches, highlighting methodological trends, and outlining the main challenges reported in prior work. We also explore broader computer vision research to inform the modelling choices of this study.

2.1 Conventional Ground-Based Approaches for Damage Assessment

Building damage assessment has traditionally relied on manual surveys and sensor-based monitoring systems. Manual surveys involve trained personnel to physically inspect and record damage levels, often in the first 48 hours after the disaster (Gupta et al., 2019). Sensor networks including accelerometers, strain gauges, and vibration-based monitoring devices, have also been employed to track structural integrity in real time (Avci et al., 2021).

While these methods provide valuable information, they face major limitations when applied at larger-scales. Manual surveys are slow and labour-intensive , delaying critical decision-making (Doshi et al., 2018; Xu et al., 2019). They also expose personnel to hazardous environments, where debris and unstable structures can make on-site inspections dangerous or unfeasible (Trekin et al., 2018). Furthermore, reliance on human judgment introduces subjectivity and inconsistency, leading to variable and sometimes unreliable assessments (Doshi et al., 2018). Sensor networks face challenges such as pre-installation and maintenance costs and they can be damaged during disasters, restricting their usefulness (Avci et al., 2021; Xu et al., 2019).

2.2 Satellite Imagery Analysis for Damage Assessment

To overcome the limitations of ground-based methods, satellite imagery analysis has become increasingly recognised as a key component of modern building damage assessment. Doshi et al. (2018) highlighted the key advantages of satellite imagery analysis. Satellites can image vast regions affected by a disaster including remote and inaccessible areas. Modern satellite constellations offer rapid revisit times, allowing for near-real-time acquisition of pre- and post-disaster imagery. This enables swift change detection and damage assessment, crucial for immediate response. They also emphasized the advent of computer vision and machine learning algorithms for automating satellite imagery analysis. This ensures consistent, objective assessments and allows for the processing of massive amounts of data efficiently, a task impossible for manual methods.

A major step in operationalising satellite imagery for damage assessment was the release of the xBD dataset (Gupta et al., 2019). This large-scale benchmark provides pre- and post-disaster imagery with over 850,000 annotated building footprints across multiple disaster types and regions, labelled with a four-level damage taxonomy, offering a standardised resource for training and evaluation.

2.3 Computer Vision for Satellite Imagery Analysis

2.3.1 Convolutional Neural Networks (CNNs)

CNNs have demonstrated consistent success in satellite imagery analysis, in tasks such as land cover classification, building extraction and change detection. Among these, ResNet has become one of the most widely adopted backbones due to its skip-connection design, which enables the training of very deep networks without degradation in performance (He et al., 2015). For instance, Mommert et al. (2021) demonstrated the effectiveness of a modified ResNet-50 architecture for satellite imagery analysis by classifying power plant types from medium-resolution Sentinel-2 data. By adapting the first convolutional layer to handle ten input channels and adjusting kernel sizes, their model preserved sensitivity to small-scale structures, a key requirement for satellite applications. The network achieved a mean accuracy of 90% across plant classes. This study highlights robustness of ResNet and its interpretability in real-world satellite tasks, supporting its use as a strong baseline CNN for building damage classification.

EfficientNet has emerged as a highly effective CNN architecture in satellite imagery analysis due to its unique compound scaling strategy, which balances network depth, width, and input resolution for optimal accuracy and efficiency (Tan and Le, 2020). By systematically scaling these dimensions, EfficientNet achieves state-of-the-art performance with significantly fewer parameters than traditional CNNs.

Le et al. (2022) systematically evaluated lightweight CNNs for remote sensing image classification and found EfficientNet-B0 outperformed MobileNet and NASNetMobile with 92.0**%** accuracy using only 4.6M parameters. Their key contribution was a novel multi-head attention mechanism that operates across three dimensional perspectives (spatial, channelheight, and channel-width), capturing both spatial and channel dependencies. This attention design, applied to intermediate feature maps, achieved an improvement to 93.8% accuracy. The study underscores the accuracy and efficiency advantages of the EfficientNet architecture family and flexibility in satellite imagery analysis. These findings motivate our choice of EfficientNet-B3 as a stronger capacity–efficiency balance for building-damage classification.

2.3.2 Object Detection

Object Detectors are the algorithms of choice for building localisation in our two-stage pipeline. ResNet-based detectors have demonstrated strong performance in remote sensing contexts. For instance, Groener et al. (2019) provided compelling evidence for ResNet backbones through their evaluation of state-of-the-art detection models on WorldView-3 and xView datasets, showing that Faster R-CNN with ResNet-50 offered the best trade-off between speed and accuracy for small object detection, achieving average precision scores of 0.685–0.691 for targets as small as 14 pixels. This confirms the robustness of ResNet architectures for extracting spatial features in satellite imagery and justifies the adoption of ResNet50-v2 as a baseline backbone in our building localisation stage.

Other object detection architectures such as YOLO are also widely used as single-stage frameworks. Ghazouali et al. (2024) showed the effectiveness of YOLOv8 in the context of aircraft detection from satellite images. The model demonstrated strong performance with average precision of 90.7% on the GDIT dataset while retaining efficient inference speed. The study highlighted the architectural strengths of YOLOv8, including its capacity to detect small objects, robust handling of multi-scale targets and robustness in cluttered backgrounds. These characteristics are directly relevant to building localisation, where structures vary widely in size, appear in dense urban layouts, and are embedded in visually complex environments. The demonstrated balance of accuracy and efficiency positions YOLOv8 as a particularly well-suited model for integration into disaster response pipelines, where rapid and scalable building detection is essential.

2.3.3 Vision Transformers (ViTs)

Dosovitskiy et al. (2021) introduced Vision Transformers (ViT), which apply self-attention mechanisms directly to image patches, enabling global receptive fields earlier in the network compared to CNNs. While ViTs lack certain inductive biases like translation equivariance, large-scale pretraining has shown they can outperform CNNs in many tasks. Their attention mechanisms provide interpretability by highlighting semantically relevant regions, and their strong transfer learning capabilities make them well-suited for domains with limited labelled data, such as disaster imagery. These properties position ViTs as a promising architecture for building damage assessment, where capturing global context and efficiently processing high-resolution imagery are critical.

Le et al. (2025) provide further empirical support for ViTs in remote sensing by systematically comparing CNN, ResNet, and Transformer-based models on the EuroSAT and PatternNet datasets. Their results show that pre-trained ViT models, particularly MobileViTV2 and EfficientViT-M2, substantially outperform CNN baselines while being more energy efficient. EfficientViT-M2 required only 38.19 MB storage and consumed less power than larger transformer variants. These findings demonstrate that ViTs not only surpass CNNs in classification performance but also meet the efficiency and robustness requirements of operational satellite systems, strengthening their case as a viable architecture for building damage assessment.

2.4 Existing Approaches to Building Damage Assessment

2.4.1 Two-Stage Pipelines

Gupta et al. (2019) made a landmark contribution to building damage assessment with the introduction of the xBD dataset, the largest publicly available benchmark for this task. The dataset provides over 850,000 annotated building footprints paired with damage labels across multiple disaster types, accompanied by the Joint Damage Scale. Their proposed baseline methodology adopts a two-stage pipeline: first, a U-Net model is used for building localisation, followed by a ResNet-based classification network that predicts damage severity from cropped building regions. This separation of tasks reflects the annotation process of xBD itself, where building footprints are delineated before damage labels are applied, and highlights the operational advantages of modular pipelines. By allowing each stage to be independently optimized, the framework facilitates accurate footprint extraction while enabling the classification stage to leverage both generic ImageNet-pretrained features and disaster-specific representations. xBD establishes both a standardized dataset and a methodological foundation that has shaped subsequent research. Its design closely aligns with a two-stage pipeline structure, where building localisation and damage classification are addressed as distinct but complementary tasks, positioning this approach as a natural and practical framework for building damage assessment.

Alisjahbana et al. (2024) proposed DeepDamageNet, also a two-stage deep learning framework that highlights the advantages of separating building localisation from damage classification. In the first stage, they compared semantic segmentation (ResNet-50 FPN) with instance segmentation (Mask R-CNN) for footprint extraction, concluding that semantic segmentation was more reliable in dense urban scenes (mIoU 0.85 vs. 0.70). The second stage employed a twin-tower ResNet-50 architecture that processed pre- and post-disaster image patches to classify damage levels. By incorporating contextual priors such as disaster-type labels, classification accuracy improved from 0.80 to 0.86, demonstrating the value of auxiliary features. Overall, DeepDamageNet achieved an F1 score of 0.66. The study underscores key strengths of a two-stage design: modularity, the ability to optimize localisation and classification separately, and enhanced robustness across multiple disaster types.

Shen et al. (2022) proposed BDANet, a two-stage convolutional framework designed to overcome the limitations of single-stage change detection in building damage assessment. The pipeline first segments buildings from pre-disaster imagery using a U-Net with a ResNet backbone, then classifies damage levels with a dual-branch network that processes pre- and post-disaster features, initialized with weights from Stage 1. To improve robustness, BDANet integrates a multi-scale feature fusion module to handle varied building sizes and a cross-directional attention mechanism to capture correlations between temporal features. Evaluated on the xBD dataset, the model achieved state-of-the-art performance with an overall F1 score of 0.806. These results show the effectiveness of a two-stage pipeline where separating localisation and classification allows each sub-task to be optimized independently while still benefiting from shared representations.

2.4.2 End-to-End and Alternative Frameworks

While two-stage pipelines have been widely adopted for building damage assessment, they present several drawbacks. Gupta and Shah (2020) highlighted these limitations, noting issues such as error propagation between stages, the lack of end-to-end trainability, and the need for stage-wise optimisation. To address these challenges, they designed RescueNet, a unified framework that performs both building segmentation and damage classification in a single forward pass. Built on a dilated ResNet-50 backbone, RescueNet integrates separate segmentation and change-detection heads, underpinned by a novel localisation-aware loss. This loss combines binary cross-entropy for building segmentation with selective categorical cross-entropy applied only to building pixels, ensuring damage classification is tied to correctly detected structures. By enabling end-to-end optimisation, RescueNet overcomes error propagation and learns shared feature representations across both tasks. On the xBD dataset, it delivered a dramatic improvement in harmonic mean F1 scores for damage classification.

Weber and Kané (2020) explored a multi-temporal damage assessment system that jointly predicts building localisation and damage levels within a single segmentation framework. They experimented with instance segmentation and semantic segmentation and found that the latter was more effective. Their model processes pre- and post-disaster imagery through shared ResNet-50 backbones, concatenates features, and applies a semantic segmentation head. This design avoids the inefficiencies of instance segmentation for small buildings and allows pixel-level classification across four damage categories. Like many other studies, they also used a cross-entropy loss function but also suggested the use of ordinal loss, recognising that misclassifications should be penalised according to their severity gap which we explore later in this study.

Kaur et al. (2023) introduced DaHiTrA, a hierarchical transformer-based framework that explicitly models temporal differences between pre- and post-disaster imagery. Unlike earlier CNN-based pipelines such as Siamese U-Net and RescueNet, which fuse features only at late stages, DaHiTrA employs transformer-based difference blocks to directly capture changes across temporal domains. This design forces the network to focus directly on structural changes rather than single-image representations, significantly improving the localisation of damage patterns. It achieved state-of-the-art performance on the xBD dataset and successfully transferred to the new Ida-BD dataset, highlighting its robustness. By combining global context through transformers with hierarchical difference learning, DaHiTrA sets a new benchmark for unified end-to-end models. For this project, DaHiTrA is particularly relevant as it illustrates the growing shift toward integrated transformer-based pipelines versus the modularity and efficiency of two-stage frameworks.

3 Dataset: The xBD Benchmark

3.1 Dataset Overview

The xBD dataset represents a seminal contribution to the field of automated building damage assessment using satellite imagery. The dataset was developed, with the collaboration of multiple disaster response agencies, to advance change detection and building damage assessment for humanitarian assistance and disaster recovery research. It addresses the critical need for rapid and accurate damage evaluation after a disaster and overcome the limitations of traditional and labour intensive methods (Gupta et al., 2019). By enabling the development of computer vision algorithms that can automate this process, xBD can potentially help accelerate response times and reduces risks to human assessors.

The dataset encompasses over 45,362 km^2 of polygon-labelled pre- and post-disaster imagery across 22,068 satellite image scenes. With 850,736 annotated building footprints spanning diverse disaster events, geographical regions, and environmental conditions, xBD provides researchers with unprecedented coverage for developing robust damage assessment models.

3.2 Dataset Composition and Structure

3.2.1 Imagery Characteristics

The xBD dataset consists of high-resolution satellite imagery sourced primarily from the Maxar/DigitalGlobe Open Data Program. The imagery features a ground sample distance (GSD) typically below 0.8 meters, providing sufficient spatial resolution for detailed building-level damage assessment. Each disaster event in the dataset includes paired pre-disaster and post-disaster RGB satellite images. Figure 1 shows representative examples of such image pairs.

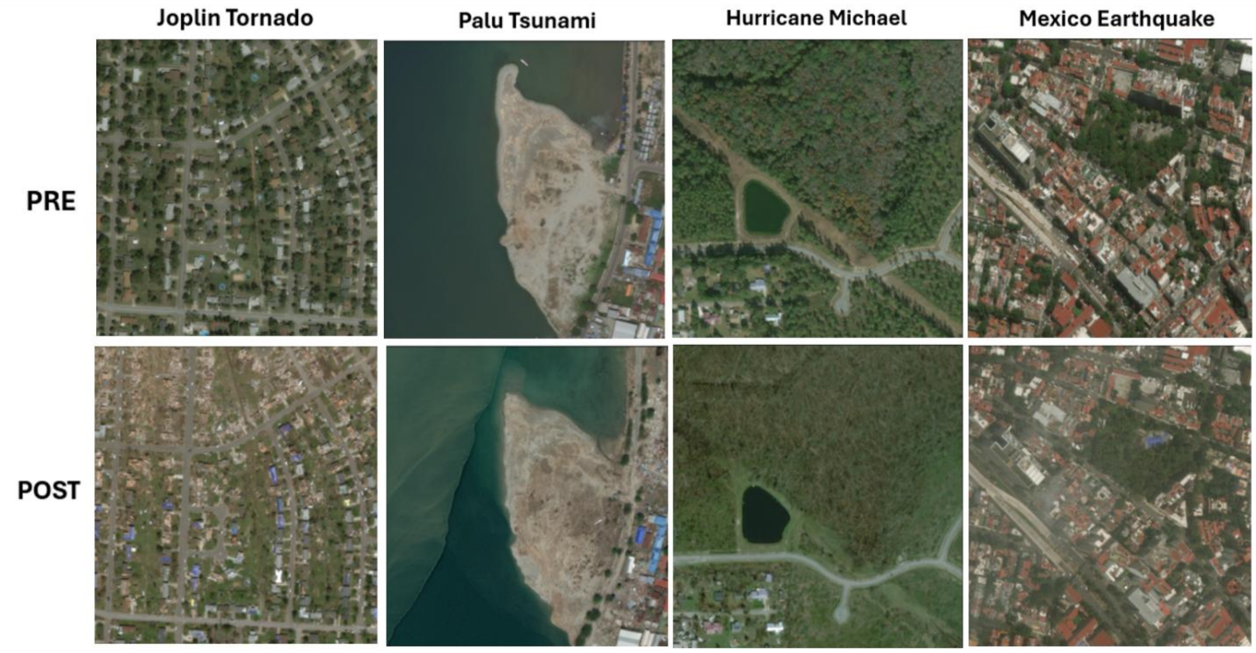


Figure 1: Pre-Disaster and Post-Disaster Images

3.2.2 Data Organisation and Metadata

Each image in the xBD dataset is accompanied by comprehensive metadata that is essential for contextual understanding and model development. The metadata includes precise geographic coordinates for spatial analysis, specific disaster type classifications facilitating disaster-specific model development, and detailed timestamps for both pre- and post-disaster imagery acquisition.

The dataset follows a structured organization with disaster-specific directories containing separate folders for images and corresponding JSON label files.

The annotation format provides building polygons in geospatial coordinate systems, enabling integration with geographic information systems (GIS) and supporting spatial analysis workflows. Each building annotation includes both geometric information (pixel and geographical coordinates) and semantic information (damage classification), providing a complete framework for supervised learning approaches.

3.2.3 Building Damage Annotation Framework

The xBD dataset provides building-level annotations comprising 850,736 instances across all disaster events. Each building is represented by a precise polygon delineating its footprint, accompanied by an ordinal damage classification based on the Joint Damage Scale. This standardized damage assessment framework categorizes building damage into four distinct levels as shown in Table 1 below.

Table 1: Joint Damage Scale descriptions (Gupta et al., 2019)

Damage Level	Description
0 (No Damage)	Undisturbed. No sign of water, structural or shingle damage or burn marks
1 (Minor Damage)	Building partially burnt, water surrounding structure, volcanic flow nearby, roof elements missing, or visible cracks.
2 (Major Damage)	Partial wall or roof collapse, encroaching volcanic flow, or surrounded by water/mud
3 (Destroyed)	Scorched, completely collapsed, partially/completely covered with water/mud, or no longer present

The class distribution within the dataset reveals a significant imbalance characteristic of real-world disaster scenarios: approximately 313,033 instances of "no damage," 36,860 instances of "minor damage," 29,904 instances of "major damage," and 31,560 instances of "destroyed" buildings.

3.2.4 Data Coverage

The xBD dataset encompasses 19 distinct natural disaster events spanning various geographical regions worldwide. The diversity of disaster types included in the dataset ensures coverage of different damage patterns and environmental conditions.

The dataset includes multiple hurricane events such as Hurricane Harvey (2017), Hurricane Michael (2018), and Hurricane Florence (2018), providing extensive examples of wind-related building damage patterns. These events capture roof damage, structural deformation, and flood-related impacts.

Earthquake damage is represented through events including the Mexico City Earthquake (2017) and other seismic events, reflecting the unique damage patterns associated with ground shaking, including building collapse, foundation failure, and structural separation.

Flood events and tsunami damage, including the devastating Palu Tsunami (2018) in Indonesia, provide examples of water-related building damage.

The dataset includes volcanic eruptions and wildfire events such as the Santa Rosa Wildfires (2017), offering coverage of fire-related damage patterns, ash deposition effects, and thermal damage to building structures.

This geographical and meteorological diversity ensures that models trained on xBD can potentially generalize across different disaster scenarios and environmental conditions, making the dataset particularly valuable for developing robust damage assessment systems.

3.2.5 Data Splits

The xBD dataset employs a carefully designed splitting strategy to support rigorous model development and evaluation. it is divided into three distinct subsets with specific purposes.

The training set comprises of 80% of the total data (18,336 images), this subset is used for model training and parameter optimization. The training set maintains the same disaster type diversity as the complete dataset, ensuring models are exposed to the full range of damage patterns during training.

The holdout set consists of 10% of the total data (1,866 images) and serves as a final, unbiased evaluation set for comprehensive performance assessment.

The remaining 10% of the dataset is the test set, and it serves to evaluate generalizability of the models through inference on unseen data.

3.3 Dataset Challenges and Limitations

Despite its comprehensive nature, the xBD dataset presents several inherent challenges that researchers must address when developing damage assessment models.

Class imbalance: The significant over-representation of the "no damage" class relative to damaged categories reflects real-world disaster scenarios but poses challenges for training balanced classifiers. This imbalance necessitates careful consideration of sampling strategies, loss function design, and evaluation metrics to ensure models can effectively detect damaged buildings rather than defaulting to the majority class.

Visual similarities: The distinction between adjacent damage levels, particularly between "minor damage" and "major damage", can be visually subtle and requires sophisticated feature extraction. This challenge is further compounded by varying viewing angles, lighting conditions, and image quality across different satellite acquisitions.

Environmental obscuration: Post-disaster imagery frequently contains environmental factors that complicate damage assessment, including cloud cover, smoke, debris accumulation, and flooding. These conditions can obscure building structures and reduce annotation quality and hence making damage classification challenging.

Incomplete Annotation Coverage: In some instances, buildings visible in post-disaster imagery may lack corresponding annotations if they were not present or were heavily obscured in pre-disaster imagery. In the test set provided, certain images contain incomplete or missing annotations, which can affect the reliability of performance evaluation.

Despite these challenges, the xBD dataset provides an ideal foundation for building damage classification tasks. Its comprehensive polygon annotations enable precise building-level analysis, while its diverse disaster coverage supports robust model evaluation across a variety of scenarios and damage types. Moreover, the standardised evaluation framework facilitates meaningful comparison with existing approaches, supporting rigorous methodology validation and performance assessment.

4 Methodology

This chapter outlines the methodology adopted for developing and evaluating a two-stage pipeline for automated building damage assessment from satellite imagery. It first presents the overall design of the proposed approach, detailing the building localisation and damage classification stages, before describing the training strategies and evaluation protocols. Together, these components provide a structured framework for assessing model performance and validating the effectiveness of the pipeline.

4.1 Pipeline Overview

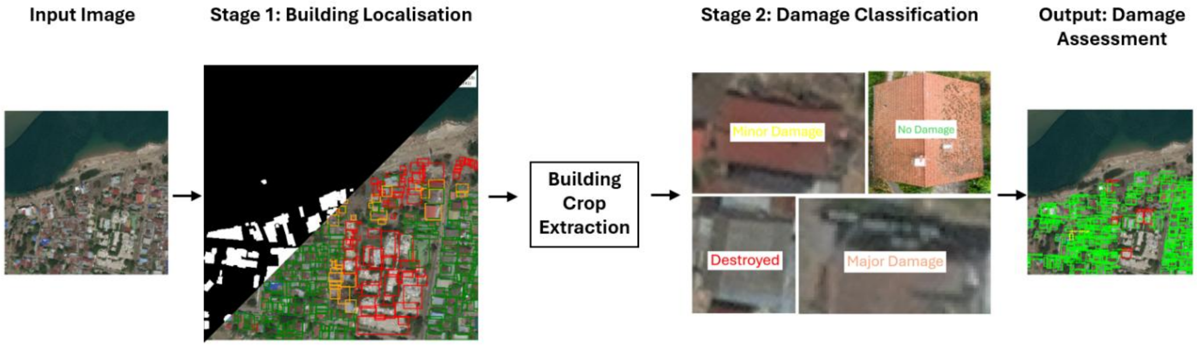

Input Image
Stage 1: Building Localisation
Building Crop Extraction
Stage 2: Damage Classification
Minor Damage
No Damage
Destroyed
Major Damage
Output: Damage Assessment

Figure 2: Proposed two-stage pipeline for post-disaster building damage assessment from satellite imagery

This study proposes a two-stage pipeline for post-disaster building damage assessment from satellite imagery. The pipeline consists of :

1. Inputting post-disaster satellite imagery
2. Stage 1 – Building Localisation: Detection and boundary extraction of buildings
3. Cropping and Context Padding: Each identified building is cropped from the original high-resolution image, with a fixed padding ratio to retain relevant surroundings.
4. Stage 2 – Damage Classification: Assigning each detected building to a predefined damage category.
5. Post-Processing and Visualisation: Classified results are mapped back to the original scene and displayed as bounding boxes colour-coded by damage level.

4.1.1 Rationale for a Two-Stage Design

Rather than training a single model to jointly detect buildings and assess their damage level, we employ a decomposed approach. Prior work suggests that this can be advantageous in numerous ways. For example, BDANet (Shen et al., 2022) highlights the benefits of task specialisation whereby U-Net is used to first localise buildings from pre-disaster imagery, suggesting that this avoids missing structures that may be heavily damaged in *post-disaster* images. Similarly, DeepDamageNet (Alisjahbana et al., 2024) shows that using cropped building images from Stage 1 significantly improves classification accuracy, as the damage classifier focuses only on the relevant regions.

A two-stage approach can also be robust to data challenges such as class imbalance. In DeepDamageNet (Alisjahbana et al., 2024) class imbalance is addressed more effectively by training the classifier separately with balanced or augmented samples.

Another advantage is the ability to employ enhanced feature representation in the second stage. For example, BDANet (Shen et al., 2022) incorporates a multi-scale U-Net with cross-directional attention to better capture correlations between pre- and post-disaster imagery which would be harder to achieve in a single-pass model.

However, some studies, such as RescueNet (Gupta and Shah, 2020), argue in favour of an end-to-end workflow. They highlight that two-stage methods are often not jointly trainable, potentially leading to sub-optimal overall performance, and that errors from the localisation stage can propagate to the classification stage.

4.2 Stage 1: Building Localisation

The first stage of the pipeline focuses on detecting and localising buildings within post-disaster satellite imagery. This stage is crucial as it provides the foundation for subsequent damage assessment and requires robust identification of building structures that may exhibit varying degrees of structural integrity. Object detection techniques are employed to generate bounding boxes around individual buildings, which are then cropped and passed to the classification stage.

4.2.1 Data Preparation and Annotation Processing

The xBD dataset provides building footprint polygons in WKT format within JSON files. These polygons are converted into axis-aligned bounding boxes using the Shapely geometry library, with a 10-pixel padding applied to ensure full building coverage.

4.2.2 Model Architectures

4.2.2.1 Faster R-CNN ResNet-50 FPN V2

The primary architecture selected for building localisation is Faster R-CNN with ResNet-50 FPN V2 backbone, representing a state-of-the-art two-stage object detection framework. This architecture combines the feature extraction capabilities of ResNet-50 with Feature Pyramid Network (FPN) enhancements and the precision of region-based detection (Qi et al., 2023).

ResNet-50 backbone provides robust feature extraction through residual connections enabling effective flow of gradient. The FPN component enhances detection performance across multiple scales, which is particularly important for identifying buildings of varied sizes in satellite imagery. The two-stage design of Faster R-CNN aligns well with the precision requirements of building localisation. The Region Proposal Network (RPN) generates high-quality object proposals, while the subsequent classification and refinement stage provides accurate bounding box localisation.

4.2.2.2 FCOS (Fully Convolutional One-Stage)

FCOS is evaluated as a modern anchor-free object detection approach. By eliminating the need for predefined anchor boxes, FCOS simplifies the detection process, which can be advantageous for irregular shaped buildings or varied orientations. The architecture leverages multi-level FPN for robust multi-scale representation and introduces a centredness branch to improve localisation accuracy, especially for objects of different sizes and aspect ratios (Tian et al., 2019).

4.2.2.3 YOLOv8s

YOLOv8 is explored as a modern single-stage detector that adopts a fundamentally different architectural philosophy compared to two-stage methods. Unlike Faster R-CNN, which separates region proposal and classification, YOLOv8 performs detection in a single forward pass through a unified architecture. It incorporates recent advances in object detection, including an improved backbone design and optimised anchor-free detection heads. The

YOLOv8s variant is selected to balance accuracy and efficiency between the lightweight nano and the larger, more computationally expensive large versions.

This model is implemented using the Python library: Ultralytics. Additionally, images are converted to the Portable Network Graphics (.png) format to be compatible with the model.

4.2.3 Training Procedure

To ensure a fair comparison, the three object detection models (Faster RCNN, FCOS and YOLOv8s) have been trained under identical baseline settings. All baseline use resized input images of 512x512, a batch size of 4, the Adam optimiser and early stopping with a patience of 8 epochs. This setup enables consistent evaluation of their relative performance on the building detection task.

A two-stage fine-tuning strategy is planned for the CNN-based architectures (Faster R-CNN and FCOS), involving selective layer unfreezing followed by full networks fine-tuning at discriminative and reduced learning rates. However, as will be shown in the *Results chapter*, these models have underperformed compared to YOLOv8 in baseline experiments and therefore extended fine-tuning has not been pursued.

YOLOv8 requires a different training approach due to its distinct architectural design and pre-training methodology. Rather than the progressive unfreezing strategy, YOLOv8 training relies on extended epoch schedules to achieve convergence. The YOLOv8s variant has been trained initially for 30 epochs with standard hyperparameters including 512×512 input resolution, batch size of 4, 0.002 learning rate, and early stopping patience of 8. The model is further optimised by increasing training to 50 epochs, reducing early stopping patience to 5, and raising the input resolution to 1024×1024 to capture finer details that may be lost when images are down sampled to 512×512. The batch size remains fixed due to GPU memory constraints. Under this configuration, the Ultralytics auto-configuration selects the AdamW optimiser with a learning rate of 0.002, which also provides adaptive learning-rate scheduling.

Data Augmentation: For Faster R-CNN and FCOS, only basic preprocessing is applied, consisting of resizing to 512×512 and normalisation. No additional augmentation is introduced in order to maintain a fair and controlled baseline. In contrast, YOLOv8 is trained using the Ultralytics implementation, which applies built-in augmentation strategies by default (e.g., flipping, colour adjustments, and geometric transformations). These augmentations are retained to follow the recommended YOLOv8 training configuration and to improve model generalisation.

All experiments are conducted on an NVIDIA Tesla T4 GPU with 15 GB of memory, using PyTorch 2.6.0 and CUDA 12.4.

4.2.4 Evaluation Framework

The training process employs robust evaluation metrics tailored to object detection performance. Mean Average Precision (mAP) serves as the primary metric, with mAP@0.5 used for model selection. This choice follows established practice in remote sensing object detection, where mAP effectively captures detector performance across varying object scales (Shermeyer and Etten, 2019).

Additional metrics including precision, recall and mAP@0.5:0.95 are also analysed for the best-performing model to guide deployment in the final pipeline. Precision is defined as the ratio of correct detections to total detections, recall as the ratio of detected buildings to total ground-truth buildings, and mAP@0.5:0.95 as the mean average precision across IoU thresholds from 0.5 to 0.95.

4.3 Stage 2: Damage Classification

The second stage of the pipeline focuses on classifying the damage level of individual buildings identified during the localisation stage. Using cropped building images extracted from post-disaster satellite imagery, the goal is to predict one of four pre-defined damage categories: *no damage*, *minor damage*, *major damage* or *destroyed*. This is a crucial step in post-disaster response as it supports prioritization of resources and recovery planning. Visual indicators of damage can vary widely; subtle damage cues may be overlooked by the human eye but can be captured by deep learning models.

In this stage, state-of-the-art classification models such as ResNet, EfficientNet, ViT and DeiT are evaluated. An ablation study is conducted to compare the effectiveness of different loss functions on each model, and a multi-stage fine-tuning strategy is employed to optimise model performance. The following sections describe data preparation, model architectures, training procedures and experimental setup in detail.

4.3.1 Building Crop Extraction for Classification

To train the damage classification models, building-level image crops are extracted from post-disaster satellite imagery in the xBD dataset using the provided building footprint annotations. Each crop corresponds to an individual building and is labelled with one of four predefined damage categories. Several extraction strategies were explored to identify an approach that balances visual clarity, computational efficiency, and memory usage. The adopted method is presented in detail, followed by a brief discussion of alternative approaches and the reasons they were not selected.

4.3.1.1 Adaptive Polygon-Based Extraction

The adopted building extraction method implements an adaptive polygon-based extraction system that dynamically adjusts crop sizes based on individual building characteristics while also employing a memory-efficient metadata-only processing strategy to address the memory management challenges. This approach is designed to preserve spatial resolution, maintaining building focus with enough contextual information and ensure scalability for large-scale satellite imagery datasets.

The adaptive sizing algorithm calculates optimal crop dimensions for each building individually, rather than applying a uniform sizing. The system computes building-specific bounding boxes from polygon coordinates and applies proportional padding (30% of building size) that scales with building size, ensuring adequate surrounding information while maintaining building focus. This proportional approach means that a 50-pixel building would receive approximately 15 pixels of padding, providing essential context without diluting the building signal, while a 200-pixel building would receive 60 pixels of padding, maintaining appropriate spatial relationships without introducing excessive background noise.

To address memory management challenges that emerged when processing a dataset with over 119,000 building instances, a metadata-only strategy is used. This approach differs from other extraction methods explored by storing only lightweight metadata for each building instance rather than the actual image crops. The metadata includes essential information such as image paths, bounding box coordinates, damage labels, building characteristics, and spatial relationships, requiring only a few megabytes of memory regardless of dataset size. Actual crops are generated on-demand during training via the data loading pipeline, enabling efficient handling of large datasets without exceeding system memory. This deferred processing approach enables the system to handle datasets that would otherwise exceed available system memory.

Visual analysis of the extracted dataset (Figure 3) shows that this method produces higher quality crops than the other approaches. Extracted crops exhibit excellent building focus while maintaining sufficient contextual information for accurate damage assessment. The adaptive sizing ensures that buildings consistently occupied appropriate portions of their crops. No-damage samples have displayed clear structural definition with minimal background interference. With minor-damage and major-damage categories, visual cues are still challenging to identify as they can easily be classed as no-damage or destroyed respectively.

This method effectively balances resolution, focus, and efficiency, addressing the critical requirements of the damage classification task. Its combination of adaptive sizing, proportional padding, and metadata-only processing makes it a robust and scalable solution for building damage assessment.

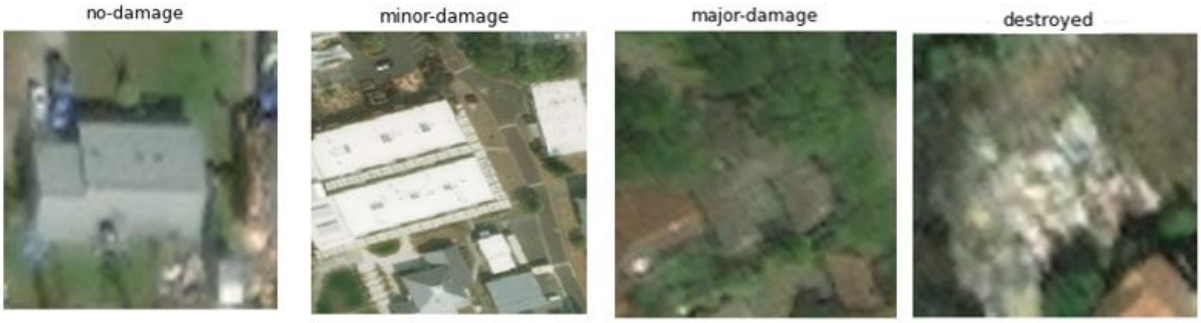


Figure 3: Visual examples from the adaptive polygon-based extraction method

4.3.1.2 Alternative Approaches Considered

Ground-Truth Polygon

The first approach used the ground truth polygon annotations in the xBD dataset to generate building-level crops. Polygons stored in Well-Known Text (WKT) format were parsed with the Shapely library, converted into axis-aligned bounding boxes with a 10-pixel padding, and filtered to remove very small buildings (<32 pixels). This method leveraged precise building boundaries and offered a conceptually straightforward way to extract pixel-accurate building instances.

However, the extracted crops were typically very small (50–100 pixels), requiring significant up sampling to meet the input requirements of modern CNNs (224×224 or 256×256). This produced severe interpolation artifacts, loss of fine detail, and poor visual quality, especially for distinguishing subtle damage categories such as minor vs. major damage. Given these limitations, the method was deemed unsuitable for robust classification and was abandoned.

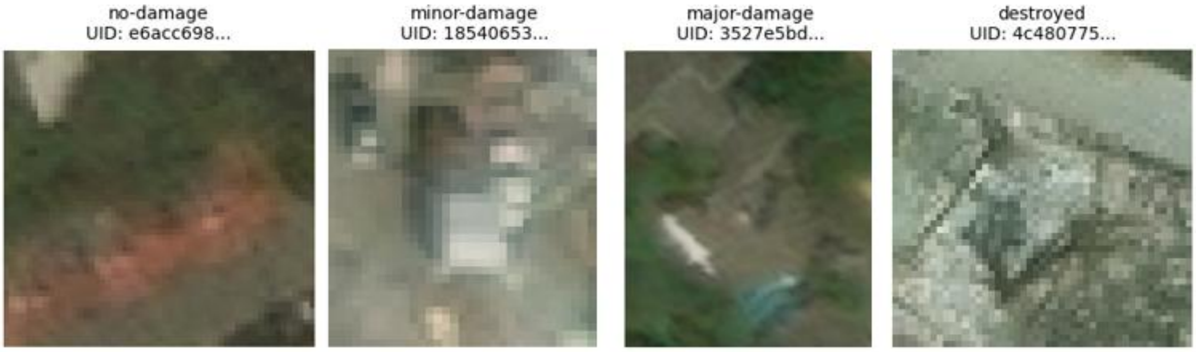


Figure 4: Example building crops generated with the polygon-based extraction method, showing low resolution and limited detail across damage categories

Fixed-Sized Centroid

To overcome the low-resolution limitations of the polygon-based method, a fixed-size centroid-based approach was tested. For each building polygon, a 224x224 crop centred on the polygon centroid was extracted, aligning with the input size requirements of most ImageNet-pretrained architectures. This ensured consistent dimensions across samples, eliminating the severe up sampling artifacts from previous approach and produced sharper and more detailed crops with visible damage features.

However, the method introduced excessive background noise and reduced focus on target buildings. Many crops contained multiple or partial structures as well as irrelevant context such as roads and vegetation, while small buildings appeared tiny and larger ones often exceeded crop boundaries. In addition, the fixed-size strategy produced a very large dataset, leading to inefficient memory usage. These limitations reduced the method's suitability for precise and scalable damage classification, and it was not adopted for the final pipeline.

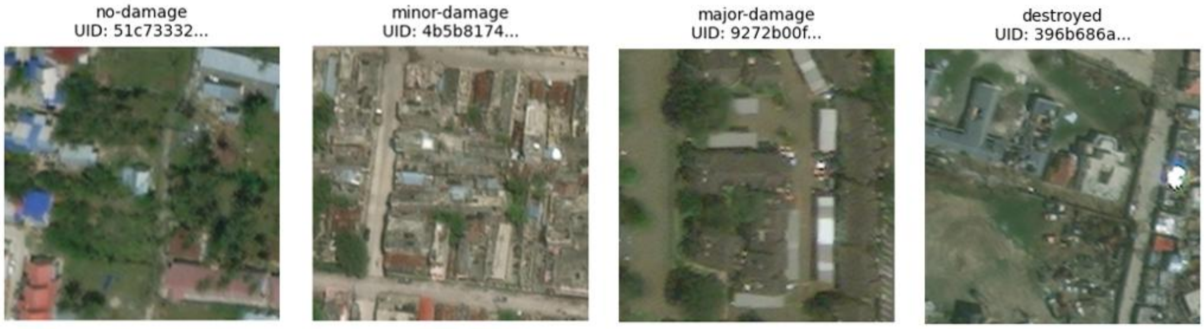


Figure 5: Example crops generated with fixed-size centroid, showing improved resolution but excessive background noise

4.3.2 Model Architectures

Several deep learning architectures are evaluated for the building damage classification task, with the goal of comparing standard and lightweight variants from convolutional and transformer-based architectures. All models are initialized with ImageNet-pretrained weights and fine-tuned on the prepared building crop dataset. The following subsections briefly summarise the key characteristics and motivations for selecting each architecture.

4.3.2.1 ResNet-50

ResNet, a widely used CNN architecture with residual connections, is selected as the baseline for this study. It was first introduced in 2015 with the key innovation of skip connections that allow gradients to flow directly through the network, addressing the vanishing gradient problem (He et al., 2015). Structurally, ResNet-50 consists of an initial convolution and max-pooling layer, followed by four sequential stages of residual blocks. Each residual block includes a series of convolutions with batch normalisation and ReLU activation, combined with an identity shortcut connection that preserves information from earlier layers. The architecture concludes with global average pooling and a fully connected layer for classification.

The ResNet architecture has proven its effectiveness in remote sensing applications, including satellite imagery classification tasks (Shabbir et al., 2021), with recent studies specifically showing its effectiveness for post-disaster building damage assessments (Bhardwaj et al., 2024). The residual connections make the architecture suitable for complex feature extraction from satellite imagery. Therefore, ResNet-50 serves as a robust baseline for performance comparison against more recent architectures.

4.3.2.2 EfficientNet-B3

EfficientNet-B3 has been selected to assess the impact of recent CNN innovations on building damage classification. The key innovation of EfficientNet lies in compound scaling, which simultaneously optimizes network depth, width, and resolution rather than scaling individual dimensions independently (Tan and Le, 2020). This approach is implemented through a series of MBConv (Mobile Inverted Bottleneck Convolution) layers that combine depth-wise separable convolutions and shortcut connections to maximize computational efficiency. In practice, this design allows EfficientNet to achieve strong accuracy with significantly fewer parameters and lower computational cost compared to conventional CNNs.

Recent study by Saricayir and Ozcan (2024) has demonstrated exceptional performance from EfficientNet on satellite imagery classification using the EuroSAT dataset while maintaining computational efficiency. We opted for EfficientNet-B3 over the baseline B0 variant to provide additional capacity for a more complex task of building damage assessment, following the recommendation from Saricayir and Ozcan (2024); to explore variants that trade some efficiency for higher capacity while maintaining the model computational advantage over other CNNs.

4.3.2.3 ViT-B/16

Vision Transformer (ViT) is selected to evaluate the potential of transformer-based architectures for building damage classification. Unlike CNNs, which use convolutional filters to extract local features, ViT divides an input image into fixed-size patches (e.g., 16x16), flattens them, and linearly embeds each patch into a token sequence. A learnable [CLS] token is added to this sequence, and positional encodings are added to retain spatial information. The sequence is then processed through a stack of transformer encoder blocks using multi-head self-attention and feed-forward networks (Dosovitskiy et al., 2021). We have opted for ViT-B/16 (Base variant with 16x16 patch size) due to its optimal balance between model capacity and computational efficiency. This model would be the most appropriate for a transformer baseline configuration.

Recent research has demonstrated that transformers can be superior in remote sensing applications. The self-attention mechanism provides the ability to capture global spatial dependencies and contextual relationships which can be crucial for understanding damage patterns across different scales within satellite imagery. Although transformer-based approaches are emerging in remote sensing research, most studies have focused on custom architectures such as DaHiTra (Kaur et al., 2023) rather than leveraging the power of pretrained models. We aim to investigate if ViTs can rival the performance of CNNs in building damage classification.

4.3.2.4 DeiT-B/16

The Data-efficient Image Transformer (DeiT) is evaluated to test whether transformer-based models optimised for data efficiency can perform well for building damage classification. DeiT builds upon the Vision Transformer (ViT) architecture but introduces a knowledge distillation framework, where a CNN teacher model guides a student transformer during training (Touvron et al., 2021). This design reduces dependence on very large training datasets, making it well suited for satellite imagery analysis where labelled data is often scarce.

DeiT has demonstrated strong performance in remote sensing applications (Bashmal et al., 2021), showing that transformers can generalise effectively in data-limited scenarios. Its efficiency and competitive accuracy make it an ideal candidate to test whether data-efficient transformers can achieve performance on par with CNNs in building damage classification tasks.

4.3.3 Loss Function Ablation Study

4.3.3.1 Cross-Entropy Loss

Cross-Entropy loss serves as the baseline loss function for the multiclass classification tasks. It is employed as the +primary comparison in this building damage assessment study from satellite imagery. Cross-Entropy provides a principled probabilistic framework to train neural networks since we have buildings assigned with one of four discrete damage levels. The loss increases as predicted probability diverges away from the actual label. It is particularly suitable for gradient based-optimisation methods.

The loss is defined as:

$$\text{Cross Entropy} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(p_{i,c}\right)$$

where N represents the number of samples, C is the number of classes, y_i is ground truth label and p_i is the predicted probability for class c. In practice, we implement this using PyTorch's built-in 'CrossEntropyLoss' function.

4.3.3.2 Focal Loss

Focal Loss was introduced by Lin et al. (2017) to address the class imbalance problem encountered in dense object detection tasks. By applying a modulating term to cross entropy loss, it enables learning to be focused on hard misclassified examples, automatically down-weighting the contribution of easy examples during training. This is particularly useful in the context of building damage assessment where certain damage categories are underrepresented.

The loss is defined as:

$$\text{Focal Loss} = -\alpha_t(1 - p_t)^{\gamma}\log\left(p_t\right)$$

where p_t is the predicted probability for the true class, γ(gamma) is the rate at which easy examples are down-weighted, α is a class balancing weight factor to address class imbalance.

Focal loss is expected to perform well in our study as it addresses two critical challenges. Through the alpha parameter, we can mitigate class imbalance. The modulating factor, gamma, can help lay more emphasis on samples that are more difficult to classify correctly. Focal Loss also dynamically adjusts sample importance based on prediction confidence, allowing the model to adaptively focus computational resources on challenging cases throughout training (Lin et al., 2017).

4.3.3.3 Ordinal Loss

Building damage assessment involves ordinal relationships where damage categories follow a severity progression from no damage to minor damage to major damage and finally to destroyed. Traditional, multi-class classification treats these categories as independent classes. This can lead to suboptimal performance. Misclassification errors should have varying significance where confusing “minor damage” with “major damage” should incur less penalty than “confusing “no damage” to “destroyed”. The implementation of Ordinal Loss was also recommended by Weber and Kane in their 2020 article to penalize errors based on damage scales.

We implement Ordinal Loss based on CORAL: Cumulative Ordinal Regression (Cao et al., 2020). Instead of predicting class membership directly, the approach models the cumulative probability of exceeding each threshold level. CORAL (COnsistent RAnk Logits) models the cumulative probability that the true class exceeds a set of thresholds. Specifically, for K ordered classes, the model outputs K−1 logits, each representing a binary classification task: whether the true class is greater than a given threshold (Cao et al., 2020). The original formulation and detailed loss equation can be found in Cao et al. (2020).

4.3.3.4 Class Weighted Loss Functions

The xBD dataset exhibits clear class imbalance, with “no damage” dominating the samples of the dataset. To mitigate this, class weights were computed using the balanced weighting strategy:

$$\text{Weight} = \frac{\text{Total Number of Training Samples}}{\text{Number of Classes} * \text{Number of Samples in a Class}}$$

Table 2: Distribution of building damage categories and weights (Training Set)

Damage Class	Sample Count	Percentage	Weight
No Damage	88,961	74.6%	0.335
Minor Damage	11,040	9.3%	2.702
Major Damage	11,802	9.9%	2.527
Destroyed	7,507	6.3%	3.973

The weights inversely correlate with class frequency, ensuring that rare classes (major-damage, minor-damage) receive proportionally higher attention during training. For Cross Entropy, weights are applied directly through the weight parameter from PyTorch's cross entropy function. For Focal Loss, weights are implemented through the alpha parameter and for Ordinal Loss, sample-specific weights were applied by multiplying the base loss with the corresponding class weight.

All loss functions have been evaluated under identical training conditions to ensure fair comparison. The experimental setup consisted of 10 training epochs using the Adam optimizer with a learning rate of 0.001 and 'ReduceLROnPlateau' scheduler (factor=0.1, patience=2 epochs). Training employed a batch size of 32 with models evaluated on both training and validation sets using accuracy, F1-score, and detailed classification reports.

4.3.4 Training Procedure

The training procedure for building damage classification has been implemented in a two-phases: the initial loss function ablation followed by three-stage progressive fine-tuning using the best-performing loss function from the ablation study. This approach ensures both a fair evaluation of different loss functions and optimal adaptation of pretrained models to the task.

During the ablation phase, a conservative fine-tuning approach is employed to maintain stability across all loss function variants. For all models, only the classifier head and the final feature extraction block were unfrozen, while earlier layers remained fixed. The best-performing loss function for each architecture is selected based on validation accuracy and class-balanced performance metrics, prioritizing configurations that shows strong performance across all damage categories rather than just overall accuracy.

Following the ablation study, the optimal model-loss combinations have undergone intensive three-stage fine-tuning to maximize performance through progressive layer unfreezing and adjustments in hyperparameters. This approach allows gradual adaptation of pretrained features while maintaining training stability. A consistent three-stage unfreezing pattern is adopted for each model:

- Stage 1: Classification Head + Final Feature Extraction layers
- Stage 2: Extended to include 2-3 additional deeper layers/blocks
- Stage 3: Unfreezing covering approximately 50% of network parameters

For CNN architectures (ResNet-50 and EfficientNet-B3), stage-wise unfreezing of residual blocks or feature stages from deepest layers backward with 15 epochs have been implemented.

For Vision Transformers (ViT-B/16 and DeiT-B/16), progressive transformer block unfreezing with patch embedding adaptation in final stage with varying number of epochs (10-15) have been implemented.

Alongside the above configurations, discriminative learning rates have been used; earlier layers have received lower learning rate (0.0001-0.0002) to preserve pretrained features, while later layers and classifiers used higher learning rates (0.0003-0.0005) for task adaptation. Similarly stronger weight decay (0.0001) has been employed for later layers and reduced decay (0.00001 – 0.00005) for deeper layers. Early stopping with patience 3 is adopted to prevent overfitting with learning rate scheduling.

This systematic approach is to ensure controlled adaptation of pre-trained representations to building damage assessment while maintaining model stability and generalization capabilities.

Data Augmentation: To prepare the building image crops for classification, standard preprocessing and light augmentation strategies are applied. On the training set, each image was resized to 224×224 pixels using bicubic interpolation, followed by random horizontal flipping, random rotations, and colour jitter. These augmentations increase robustness to orientation, lighting, and appearance variability in satellite imagery. Images were then converted to tensors and normalised using the standard ImageNet mean and standard deviation values. To ensure consistency during evaluation, only resizing (224×224) and normalisation are applied, without augmentation for the validation set.

4.3.5 Evaluation Framework

Model performance is assessed using standard classification metrics including accuracy, precision, recall, and F1-score. Given the multi-class nature of building damage assessment, macro-averaged metrics are employed to assign equal weight to each damage class, ensuring balanced evaluation across all severity levels regardless of class frequency in the dataset.

F1-Score as the Primary Metric

The F1-score, defined as the harmonic mean of precision and recall, serves as our primary evaluation metric:

$$\mathrm{F1-Score} = \frac{2 * (\mathrm{Precision} * \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})}$$

where:

$$\mathrm{Precision} = \frac{\text{True Positive (TP)}}{(\text{True Positive (TP)} + \text{False Positive(FP)})}$$

$$\mathrm{Recall} = \frac{\text{True Positive (TP)}}{\big(\text{True Positive (TP)} + \text{False Negative (FN)}\big)}$$

F1-score is prioritised as it provides a balanced measure of both precision and recall, penalising false positives and false negatives equally. This is crucial for accurate resource allocation and efficient recovery planning after disasters.

Emphasis on Recall for Sever Damage Classes

While F1-score provides overall measure of balance, particular emphasis is placed on recall performance for severe damage classes (major damage and destroyed). Recall measures the ability of the model to correctly identify all instances of actual damage. High recall ensures that all truly damaged structures are correctly identified, reducing the likelihood of false negatives. This is critical in disaster response, where missed detections of severely damaged buildings could result in misallocation of resources, delayed rescue efforts, or failure to issue evacuation orders for unsafe structures.

5 Results

This section presents the experimental results for the two components of the proposed pipeline. The first part evaluates building localisation using object detectors while the second part focuses on damage classification of cropped building instances. Each stage is trained and assessed independently, and the final section demonstrates how the two stages are integrated into the full pipeline and the adjustments that have been made.

5.1 Stage 1: Building Localisation

This section reports the performance of the evaluated object detectors for building localisation. Three different models have initially been compared under a common training setup to ensure fairness, and the results are presented for both the baseline experiments and for extended fine-tuning of the best-performing model.

5.1.1 Baseline Performance

The baseline comparison includes Faster R-CNN (ResNet-50 FPN V2), FCOS and YOLOv8s. Their respective mAP scores at an IoU threshold of 0.5 are shown in Table 3 below. Among the three models, YOLOv8s has achieved the highest mAP.

Table 3 : Baseline performance of object detectors

	mAP (IoU=0.5)
Faster R-CNN (ResNet FPN v2)	0.3308
FCOS	0.2878
YOLOv8s	**0.5300**

Faster R-CNN has 0.3308 mAP (IoU=0.5) at epoch 9. A steady increase in validation mAP is observed for the first 10 epochs before plateauing and triggering early stopping at epoch 17. The average number of predictions per image has decreased from 78 to 65 while the share of high-confidence detections (>0.5) increases from 47.3% to 61.3% in the last epoch, indicating improved calibration and less spurious boxes.

FCOS improves quickly from epoch 6 to 10, then plateaus, with peak performance observed at epoch 16 with mAP of 0.2878. The FCOS detector consistently produced approximately 97 to 100 proposals per image with only 33% above 0.5 confidence, suggesting weaker precision at higher IoUs.

YOLOv8s substantially outperforms both Faster R-CNN and FCOS baselines. On the validation split, it has achieved mAP@0.5 of 0.530 and mAP@0.5:0.95 of 0.279. The training shows stable optimization, with all loss components showing consistent monotonic reduction: box loss decreased from 2.26 to 1.807, classification loss from 1.726 to 1.108, and distribution focal loss from 1.26 to 1.099 over the 30-epoch training period. Validation performance improved steadily throughout training, with mAP@0.5 rising from 0.371 in the first epoch to 0.530 at convergence, indicating effective learning without overfitting. The final precision-recall balance of 0.695/0.478 demonstrates reliable detection performance on this dense dataset containing approximately 58 building instances per image on average.

These results justify selecting YOLOv8s as the detector for further fine-tuning and for generating building crops in the damage classification stage.

5.1.2 YOLOv8s Fine-Tuning

The fine-tuned YOLOv8s model achieved substantial improvements on the validation set of 1,866 images containing 108,784 building instances, with precision of 0.729, recall of 0.531, mAP@0.5 of 0.585, and mAP@0.5:0.95 of 0.336. The extended 50-epoch training schedule has continued to yield performance gains without triggering early stopping while loss components have showed consistent optimization: box loss decreased from 1.934 to 1.587, classification loss from 1.488 to 1.014, and distribution focal loss from 1.313 to 1.150.

Despite processing full-resolution imagery, the model maintained practical inference speeds of 10.8 ms with 5.2 ms post-processing. The balanced precision-recall performance indicates effective building detection across diverse disaster scenarios while maintaining enhanced sensitivity to smaller structures that benefit from full-resolution processing, positioning the model optimally for the subsequent damage classification pipeline.

5.1.2.1 Baseline Model vs Fine-Tuned Model

The extended fine-tuning strategy has produced consistent improvements across all key performance metrics. Table 4 presents a comprehensive comparison between the baseline and fine-tuned configurations.

Table 4: Baseline and Fine-Tuned YOLOv8s Performance Analysis

Configuration	Input Size	Epochs	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Baseline YOLOv8s	512x512	30	0.695	0.478	0.530	0.279
Fine-Tuned YOLOv8s	1024x1024	50	0.729	0.531	0.585	0.336

The fine-tuned model has achieved an 11.1% relative improvement in recall, indicating significantly better capability in identifying building instances. This improvement is essential for disaster response scenarios where missing damaged buildings can have severe consequences. Precision also improved by 4.9%, reducing spurious detections. The mAP@0.5:0.95 metric showed the most significant improvement (+20.4%), indicating enhanced performance across different IoU thresholds. This suggests the fine-tuned model produces more accurate bounding box localisations, particularly important for precise building extraction in the classification stage.

Loss convergence further supports these improvements: the fine-tuned model achieved final box, classification, and distribution focal losses of 1.587, 1.014, and 1.150, respectively, outperforming the baseline losses of 1.807, 1.108, and 1.099. Processing higher-resolution (1024×1024) imagery enabled the detection of finer structural details, particularly benefiting small buildings (50–200 pixels), which make up a substantial portion of the dataset.

The fine-tuned model maintained practical inference speeds of 10.8ms per image despite processing larger images. The computational overhead is reasonable considering the substantial performance gains. GPU memory usage increases moderately, peaking at 14.1GB compared to 7.83 GB for the baseline.

Figure 7 presents a qualitative analysis, comparing building detection performance across ground truth annotations (green), baseline YOLOv8s (blue), and fine-tuned YOLOv8s (red). The fine-tuned model demonstrates substantially improved detection coverage. In the top row, the baseline model captures only prominent structures while missing numerous smaller buildings. The fine-tuned model achieves comprehensive coverage, detecting most structures present in ground truth annotations. In the second row, the baseline has numerous false positives while the predictions of the fine-tuned model are more in line with the ground truths. In the third row, the fine-tuned model is more precise in detecting buildings while the baseline model has grouped nearby buildings in some instances.

The visual comparison (Figure 6) reveals improved multi-scale detection capability. The fine-tuned model successfully identifies small residential structures, medium-scale commercial buildings, and larger facilities across diverse urban morphologies. This improvement validates the decision to increase input resolution.

However, several detections in the fine-tuned model lack corresponding ground truth annotations, which may indicate either false positives or incomplete ground truth annotation which is a common limitation in large-scale satellite imagery datasets. A small number of heavily damaged or shadowed buildings also remain undetected, indicating ongoing challenges in complex visual conditions.

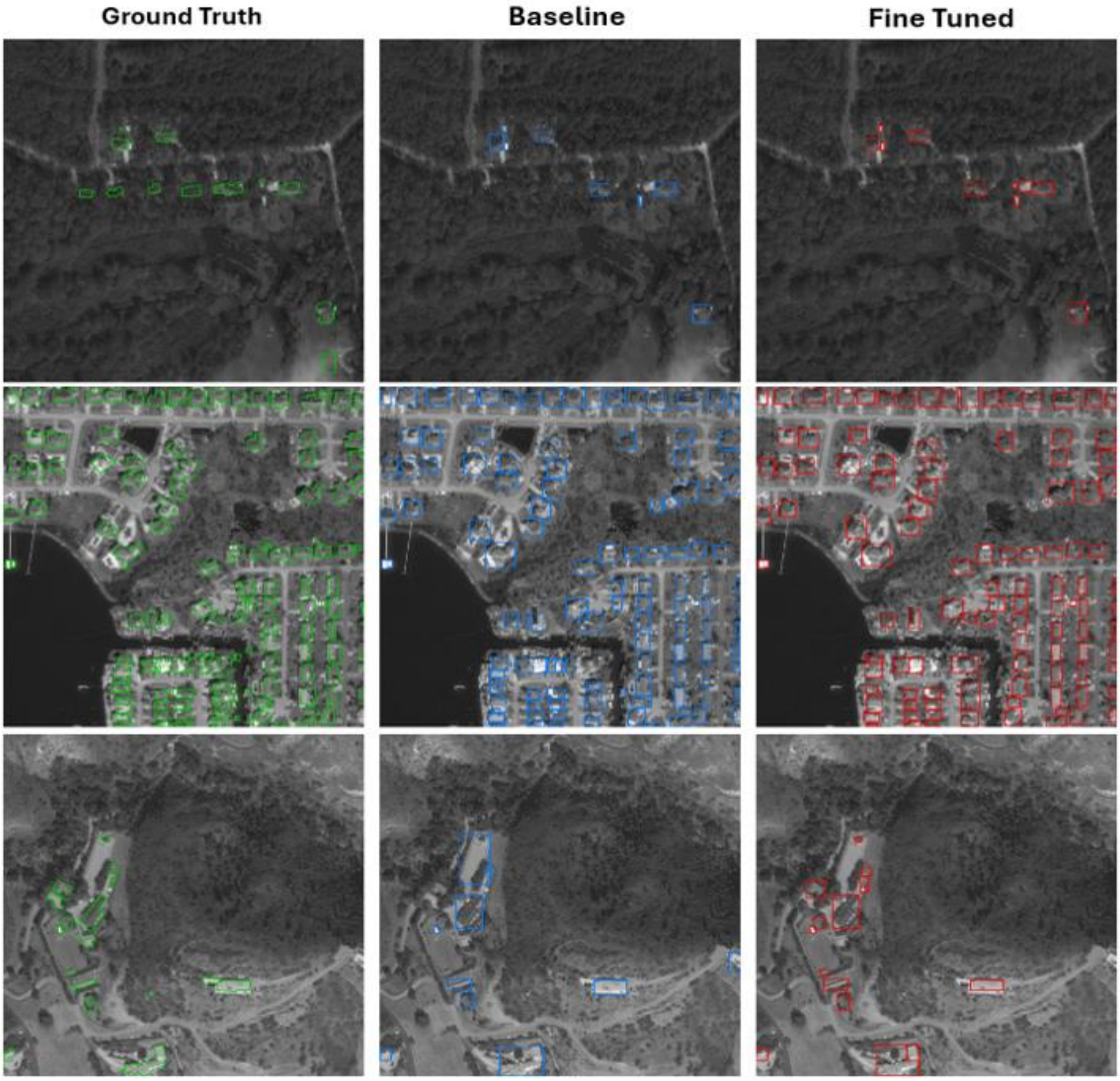


Figure 6: Qualitative Analysis of Baseline and Fine-Tuned YOLOv8s

5.1.3 Test Set Evaluation

The fine-tuned YOLOv8s model is evaluated on the test set comprising of 1,866 images, containing 109,724 ground truth buildings to assess generalization performance on unseen disaster scenarios.

Table 5: Test Set Performance Metrics

Metric	**Value**
Precision	0.7399
Recall	0.4969
F1-Score	0.5945
mAP@0.5	0.4294
mAP@0.5:0.95	0.2812

On the test set, the model has achieved a precision of 73.99% and recall of 49.69%, resulting in an F1-score of 59.45%. The mAP@0.5 of 0.4294 indicates moderate detection performance across varying IoU thresholds, while the lower mAP@0.5:0.95 of 0.2812 suggests reduced accuracy at stricter localisation requirements.

The detection analysis revealed 73,693 total predictions from 109,724 ground truth buildings, with 54,522 true positives, 19,171 false positives, and 55,202 false-negatives. These results highlight the ability of the model to achieve strong precision while maintaining room for improvement in recall, particularly for difficult-to-detect structures.

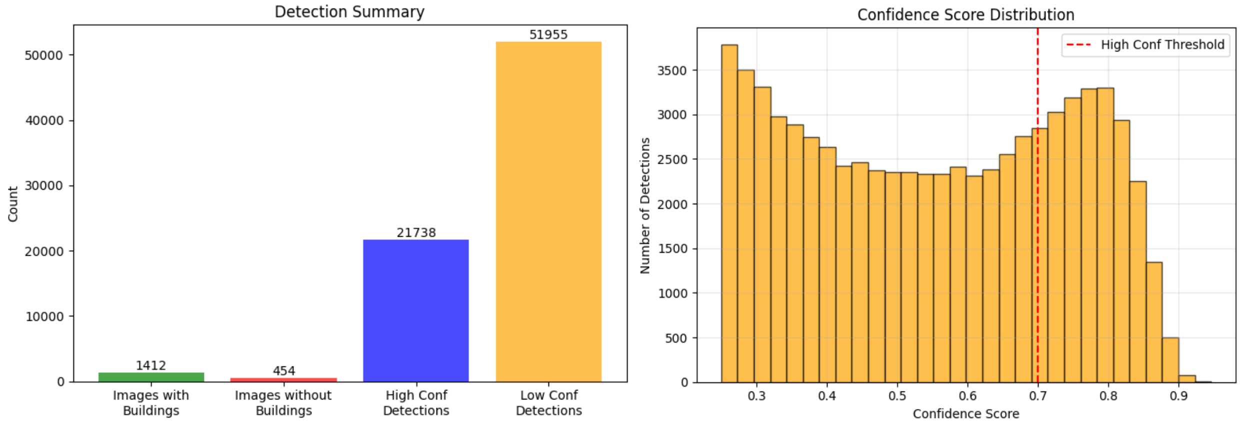


Figure 7: YOLOv8s Test Results

Figure 7 illustrates the detection behaviour of YOLOv8s on the test dataset. The *Detection Summary* shows 21,738 high-confidence detections (≥0.7),with 51,955 being in low-confidence detections. This means that about 70% of predictions fall below the high-confidence threshold. The model faces difficulties in consistently identifying buildings in post-disaster imagery.

The *Confidence Score Distribution* exhibits a bimodal pattern with peaks around 0.3 and 0.8, separated by the 0.7 threshold (red dashed line). This bimodal distribution indicates two distinct detection regimes: challenging scenarios where the model exhibits low confidence (<0.3) and clear scenarios where high confidence is achieved (>0.8). The substantial volume of low-confidence detections suggests many buildings in post-disaster imagery present ambiguous visual characteristics that may challenge automated detection.

For pipeline integration, this distribution highlights a critical threshold selection trade-off. A high threshold (≥0.7) improves precision but excludes a majority of detections, while a lower threshold (0.25–0.4) captures more buildings but risks increased false positives. The final threshold will therefore be selected empirically to balance precision and recall for downstream classification.

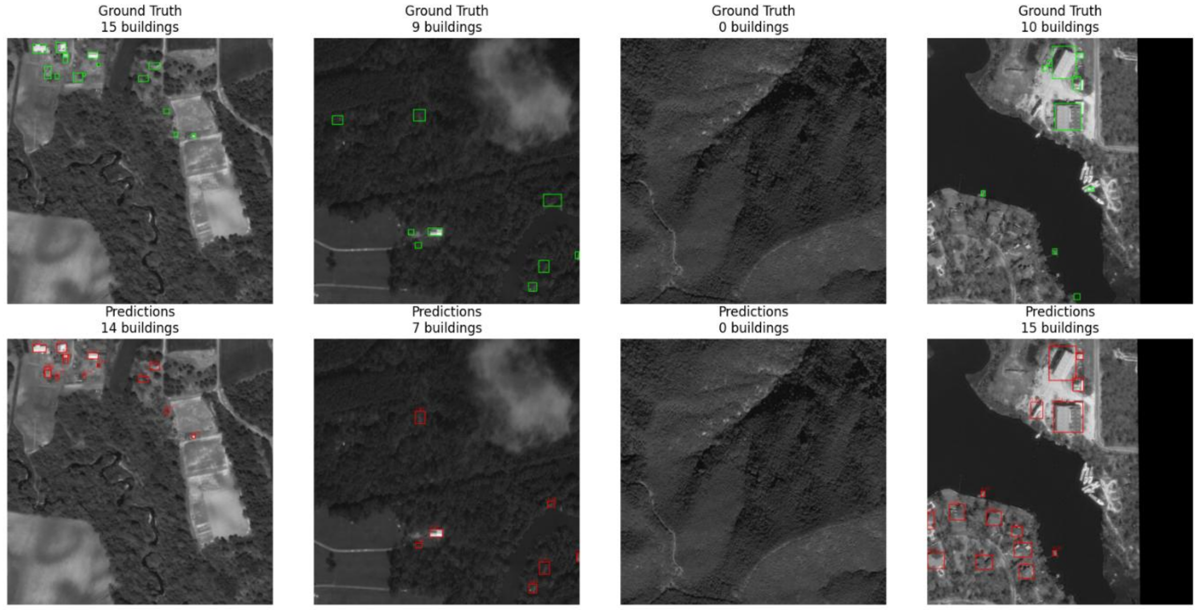


Figure 8: Qualitative Analysis of YOLOv8s on Test Set

Visual analysis of representative test cases reveals scenarios where the model succeeds and encounters difficulties. The model shows good performance in dense urban areas with clear building boundaries and moderate damage levels as seen in Figure 8 (first column) where 14 out of 15 buildings have been identified. It also detects successfully well-defined structural elements with regular geometric shapes and sufficient contrast against the background.

By contrast, the model struggles in more challenging scenarios. In Figure 8 (second column) isolated buildings in agricultural or mountainous terrain are often missed due to limited contextual cues and poor background contrast. The model also struggles with complex urban layouts with dense informal settlements, presenting challenges due to overlapping structures and unclear boundaries. In Figure 8 (fourth column), over detection is caused by misclassification of infrastructure elements as buildings, with 15 out of 10 buildings being detected.

5.2 Stage 2: Damage Classification

This section presents the experimental results for building damage classification using cropped building instances extracted from post-disaster satellite imagery. The evaluation follows the two-phase training approach described in Section 4.3: first, an ablation study of loss functions across all architectures followed by three-stage progressive fine-tuning of the best performing model-loss combinations.

5.2.1 Loss Function Ablation Study

The ablation study evaluates Cross Entropy, Focal Loss and Ordinal Loss across four architectures: ResNet-50, EfficientNet-B3, ViT-B/16, and DeiT-B/16. All models have been trained under identical conditions with conservative fine-tuning (classifier head + final feature extraction layer unfrozen) for 10 epochs to ensure stable comparison.

F1-score serves as the primary evaluation metric, with particular emphasis on recall for severe damage classes (major damage and destroyed), where false negatives have the greatest impact on disaster response.

Table 6: Performance Results from Loss Function Ablation study

	Cross Entropy		Focal Loss		Ordinal Loss	
	Accuracy	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
ResNet-50	0.83	0.73	0.71	0.65	0.80	0.72
EfficientNet-B3	0.80	0.70	0.65	0.61	0.78	0.71
ViT-B/16	0.81	0.72	0.66	0.61	0.79	0.71
DeiT-B/16	0.79	0.67	0.59	0.54	0.74	0.67

Cross Entropy consistently achieved the highest performance across all architectures. From Table 6, ResNet-50 with Cross-Entropy has achieved the best overall performance with 83% accuracy and 0.73 F1-score, while maintaining 79% recall for severe damage classes (macro-average of major damage: 73% and destroyed: 86%). The weighted Cross-Entropy implementation effectively mitigated class imbalance while maintaining balanced performance across categories.

Focal Loss has significantly underperformed expectations in overall accuracy and F1-Score, despite being specifically designed to address class imbalances. However, the value of recall for severe damage classes are higher than Cross Entropy and Ordinal Loss across all models, Table 7 below illustrates the performance of recall for severe damage classes that reveals important trade-offs in classification strategy.

Table 7: Recall of Severe Damage Classes (Major Damage and Destroyed)

	Cross Entropy		**Focal Loss**		**Ordinal Loss**	
	Major Damage	**Destroyed**	**Major Damage**	**Destroyed**	**Major Damage**	**Destroyed**
ResNet-50	0.73	0.86	0.74	0.87	0.73	0.84
EfficientNet-B3	0.75	0.85	0.75	0.86	0.71	0.84
ViT-B/16	0.71	0.83	0.71	0.84	0.71	0.82
DeiT-B/16	0.70	0.85	0.73	0.87	0.68	0.81

The Focal Loss mechanism to down weight easy example (“No Damage” instances in our case) has resulted in increases attention to minority classes, particularly severe damage categories. This creates an inherent trade-off: while overall accuracy may decrease as the model becomes more sensitive to potential damage indicators, it achieves significantly better detection of damaged buildings, demonstrating how focal loss prioritizes minority class sensitivity over balanced classification performance.

Ordinal Loss has demonstrated moderate performance across all architectures, with results consistently falling between Cross-Entropy and Focal Loss. ResNet-50 with Ordinal Loss achieved 80% accuracy and 0.72 F1-score, with severe damage recall of 78% (major damage: 73%, destroyed: 84%). While theoretically well-suited for the ordinal nature of damage categories, the CORAL-based implementation shows only marginal improvements in ordinal consistency compared to standard Cross-Entropy. Ordinal Loss shows similar performance to Cross Entropy; not surpassing Focal Loss.

In summary, Cross-Entropy emerged as the most balanced option, maintaining strong overall classification quality while achieving competitive recall on severe damage categories. Although Focal Loss slightly improved severe-class recall (1–2%), this came at the cost of substantial reductions in overall performance. Therefore, all four architectures proceeded to extended three-stage fine-tuning using Cross-Entropy.

5.2.2 Three-Stage Fine-Tuning

Following the ablation study, the best-performing model–loss combinations undergo three-stage progressive fine-tuning (Section 4.3.4). This strategy is applied to ResNet-50, EfficientNet-B3, ViT-B/16, and DeiT-B/16, all trained with Cross-Entropy Loss.

Table 8: Three-Stage Fine-Tuning Results

	Stage 1		Stage 2		Stage 3	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
ResNet-50	0.82	0.74	0.82	0.74	**0.85**	**0.75**
EfficientNet-B3	0.78	0.70	0.84	0.74	**0.85**	**0.75**
ViT-B/16	**0.87**	**0.77**	0.79	0.72	0.81	0.71
DeiT-B/16	0.79	0.67	0.80	0.71	**0.83**	**0.75**

ResNet-50 has benefitted from the three-stage fine-tuning approach, achieving its optimal performance in the final stage. Stage 1 (Layer 4 + Classifier) reaches 82.51% validation accuracy after 6 epochs with early stopping, establishing a solid baseline with balanced performance across damage categories. Stage 2 (Layer 3-4 + Classifier) shows marginal improvement to 82.64% validation accuracy indicating that moderate layer unfreezing provided limited additional benefit. However, Stage 2 (Layers 2-4 + Classifier) has yielded substantial gains, achieving the highest validation accuracy of 85.33% after 13 epochs. The model has maintained strong recall for severe damage classes (73% major damage, 85% destroyed), confirming its robustness for disaster scenarios. Loss curves showed steady convergence (from 0.1268 to 0.6675 validation loss), validating the progressive unfreezing strategy for ResNet-50.

EfficientNet-B3 has shown a similar progressive improvement pattern as ResNet-50, with consistent performance gains across all three stages, ultimately achieving the highest validation accuracy among all the explored architectures. Stage 1 (last 2 MBConv blocks + Classifier) establishes a foundation with 79.37% validation accuracy after triggering early stopping at 6 epochs. Stage 2 (features[6-7] + classifier) shows significant improvement, reaching 83.69% validation accuracy after 15 full epochs without triggering early stopping, indicating the architecture's capacity to benefit from extended training. Stage 3 (features[5-7] + classifier) achieves the peak performance of 85.40% validation accuracy at epoch 12. The final model demonstrates excellent balanced classification with macro-averaged precision, recall, and F1-scores of 0.74, 0.77, and 0.75 respectively, while achieving strong severe damage class recall (major damage: 71%, destroyed: 83%). The training pattern reveals the ability of EfficientNet-

B3 to leverage progressive unfreezing effectively, with consistent training loss reduction and stable validation performance.

ViT-B/16 has demonstrated a contrasting performance pattern compared to CNN architectures, achieving its optimal performance in the initial stage before experiencing degradation in subsequent stages. Stage 1 (last block + head) has achieved the highest validation accuracy of 86.97% after 8 epochs, establishing exceptionally strong baseline performance with macro-averaged precision, recall, and F1-scores of 0.78, 0.75, and 0.77 respectively, while maintaining strong severe damage class recall (major damage: 71%, destroyed: 80%). Stage 2 (last 3 blocks + head) has shown significant performance degradation to 82.93% validation accuracy after 3 epochs with early stopping. This indicates that unfreezing additional transformer blocks disrupted optimal feature representations. Stage 3 (patch embedding + last 6 blocks + classifier) partially recovers performance to 84.23% validation accuracy at epoch 7, though still falling 2.7 percentage points below Stage 1 performance. The performance trajectory suggests that transformer architectures may benefit from more conservative fine-tuning approaches, as ViT-B/16 achieved optimal performance with minimal parameter adjustment, indicating the pretrained attention mechanisms were already well-suited for spatial relationship analysis required in damage assessment.

DeiT-B/16 has shown moderate performance improvements through the three-stage fine-tuning approach, though with less consistent gains compared to CNN architectures. Stage 1 (last block + head) has achieved 79.57% validation accuracy after 10 epochs with early stopping. Stage 2 (last 3 blocks + head) has shown improvement to 80.43% validation accuracy after 6 epochs with early stopping. Stage 3 (last 6 blocks + enhanced head) has achieved the highest performance of 83.67% validation accuracy at epoch 4. The final model has demonstrated good classification performance with macro-averaged precision, recall, and F1-scores of 0.73, 0.79, and 0.75 respectively, while maintaining excellent severe damage class recall (major damage: 73%, destroyed: 83%). However, the training exhibits signs of instability with frequent early stopping and validation performance fluctuations.

5.2.3 Test Set Evaluation

Following three-stage fine-tuning, the best-performing variant of each architecture was evaluated on the test set to assess generalizability and identify the optimal model for pipeline integration. The test set contains 39,192 building instances across 1,866 images from previously unseen disaster scenarios, with a damage category distribution consistent with the validation set.

Table 9: Test Set Performance Results

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Major Damage Recall**	**Destroyed Recall**
ResNet-50	0.873	0.727	**0.792**	0.755	0.748	**0.863**
EfficientNet-B3	**0.878**	**0.739**	0.780	**0.756**	**0.757**	0.844
ViT-B/16	0.858	0.701	0.784	0.734	0.787	0.863
DeiT-B-16	0.852	0.687	0.792	0.730	0.755	0.879

ResNet-50 demonstrates robust generalisation performance, achieving 87.3% test accuracy compared to 85.3% validation accuracy, indicating effective learning without overfitting. The model maintains strong classification balance with macro-averaged precision, recall, and F1-scores of 0.727, 0.792, and 0.755 respectively. It performs well on severe damage classes as observed in Table 9 above, reliably identifying critical damage scenarios essential for disaster response prioritization. The consistent performance across validation and test sets confirms ResNet-50 as suitable for robust damage classification in diverse post-disaster scenarios.

EfficientNet-B3 achieves the highest test accuracy among all the architectures at 87.8%. The model maintains superior classification balance with macro-averaged precision, recall, and F1-scores of 0.739, 0.780, and 0.756 respectively, marginally outperforming ResNet-50 across all metrics. For severe damage classes, the model achieves strong performance with 75.7% recall for major damage and 84.4% recall for destroyed buildings, though slightly lower than ResNet-50's destroyed class recall (86.3%). The consistent improvement from validation to test performance, combined with the highest overall accuracy, establishes EfficientNet-B3 as the leading architecture for reliable damage classification across diverse post-disaster scenarios.

ViT-B/16 achieves 85.8% test accuracy compared to 86.97% validation accuracy, indicating a modest 1.2 percentage point decline but still robust performance without significant overfitting. The model maintains reasonable classification balance with macro-averaged precision, recall, and F1-scores of 0.701, 0.784, and 0.734 respectively, though trailing behind both CNN architectures. For severe damage classes, ViT-B/16 achieves competitive performance with 78.7% recall for major damage and 86.3% recall for destroyed buildings, matching the performance of ResNet-50 on the destroyed category. Despite achieving the highest validation accuracy (86.97%) in Stage 1 training, the degradation in performance of the transformer architecture in subsequent fine-tuning stages and lower test performance compared to CNN counterparts suggests that vision transformers may require more specialized optimization strategies for building damage classification tasks.

DeiT-B/16 demonstrates reasonable generalization performance, achieving 85.2% test accuracy compared to 83.67% validation accuracy. The model maintains moderate classification balance with macro-averaged precision, recall, and F1-scores of 0.687, 0.792, and 0.730 respectively, ranking lowest among all architectures in precision and F1-score. However, DeiT-B/16 excels in severe damage class detection, achieving the highest recall rates with 75.5% for major damage and 87.9% for destroyed buildings, surpassing all other models in identifying the most critical damage category. Despite the lowest overall performance metrics and training instability evidenced by frequent early stopping, the superior sensitivity to severe damage classes makes DeiT-B/16 valuable for disaster response scenarios where maximizing detection of destroyed buildings is prioritized over overall classification accuracy.

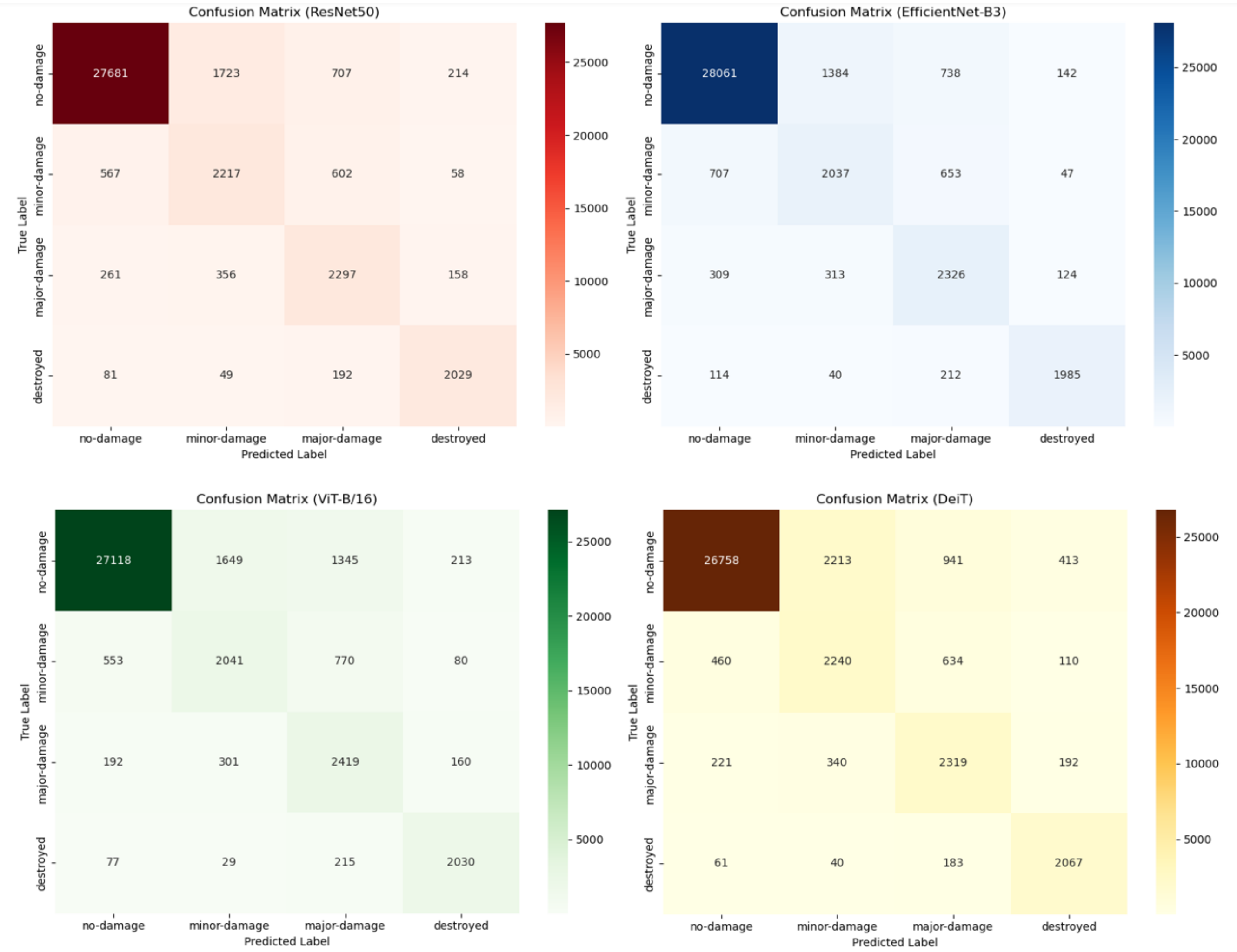


Figure 9: Test Set Confusion Matrices for Damage Classification Architectures

Figure 9 presents the confusion matrices for all four architectures evaluated on the test set, showing distinct classification patterns and error characteristics across damage categories. All models achieve strong performance on the dominant "no-damage" class, with correct classifications ranging from 26,758 (DeiT-B/16) to 28,061 (EfficientNet-B3). However, notable differences emerge in minority class handling, particularly for "minor-damage" where misclassification rates vary significantly across architectures. EfficientNet-B3 shows the most balanced performance, with slightly lower false positives, while DeiT-B/16 exhibits higher confusion between adjacent damage categories, evidenced by increased misclassifications between "no-damage" and "minor-damage" (2,213 instances) and between "minor-damage" and "major-damage" (340 instances).

The confusion matrices reveal critical insights into severe damage class performance, where accurate identification is essential for disaster respons. ResNet-50 and EfficientNet-B3 demonstrate superior precision in distinguishing between damage categories, with ResNet-50 achieving 2,029 correct "destroyed" classifications out of 2,351 total instances. ViT-B/16 and DeiT-B/16 show higher inter-class confusion, particularly DeiT-B/16's tendency to misclassify "destroyed" buildings as "no-damage" (61 instances) and "major-damage" as "minor-damage" (340 instances). Despite these classification errors, DeiT-B/16 achieves the highest true positive rate for destroyed buildings, confirming its superior recall performance for the most critical damage category.

Overall, EfficientNet-B3 achieves the most balanced results, combining the highest accuracy (87.8%), strong macro precision (73.9%), and F1-score (0.756).

5.3 Building Damage Assessment Pipeline

This section evaluates the performance of the complete two-stage pipeline on the xBD test set. Unlike the standalone experiments, this integrated evaluation reflects end-to-end capability, measuring the effectiveness of the pipeline in detecting buildings, classifying damage levels, and generating actionable outputs for disaster response.

5.3.1 Pipeline Configuration

The integrated two-stage pipeline combines optimized detection and classification models in a unified framework for real-world deployment. Configuration parameters are informed by prior component-level evaluations.

- **Stage 1 – Building Localisation:** YOLOv8s operates with a low confidence threshold (0.1) to maximize recall, since false negatives cannot be recovered downstream while false positives can be filtered in Stage 2. This choice is supported by test set analysis, which showed many valid detections in the 0.2–0.4 confidence range.
- **Stage 2 – Damage Classification:** EfficientNet-B3, adapted for the four-class taxonomy, processes building crops generated through adaptive polygon-based extraction with 30% proportional padding. For very small or oversized buildings, padding is adjusted to preserve building focus.
- **Integration:** The pipeline runs on CUDA (with CPU fallback) and employs in-memory operations with automatic coordinate scaling and error handling for diverse input formats. Outputs are provided as color-coded bounding boxes with aggregated damage statistics for disaster response applications.

5.3.2 Pipeline Performance on Test Data

The integrated pipeline is evaluated on the complete xBD test set of 933 post-disaster images, providing end-to-end damage assessment performance metrics.

Table 10: Overall Pipeline Performance Metrics

Metric	**Value**	**Description**
Buildings Processed	22,213	Total buildings detected and classified
Detection Rate	40.5%	Percentage of ground truth buildings detected at localisation stage
Classification Accuracy	79.4%	Accuracy of damage classification on detected buildings
End-to-End F1 Score	0.503	Overall pipeline F1 performance
Processing Time	1.28s/image	Average processing time per image

The integrated pipeline demonstrates both notable strengths and significant limitations. It is evaluated on 752 successfully processed images from the xBD test set (80.6% processing success rate from 933 total images). Overall, the pipeline has achieved a localisation F1-score of 0.457 and a classification F1-score of 0.523. The weighted end-to-end F1-score of the pipeline is 0.503, where 30% is from building localisation and 70% from building damage classification; this weighting follows the methodology described in BDANet (Shen et al., 2022) and is implemented for better comparison with other studies. The two-stage approach shows efficient computational performance of 1.28 second processing time per image.

Localisation/Detection Performance

The detection stage achieves 40.5% recall and 52.4% precision, identifying 22,213 true positive buildings from 54,862 ground truth instances. This substantial coverage gap represents a fundamental limitation, as the 32,649 false negatives cannot be recovered through downstream processing. The detection precision of 52.4% indicates that nearly half of all detections (20,208 false positives) are non-building objects, creating additional computational burden for the classification stage.

Classification Performance on Detected Buildings

Table 11: Damage Category Performance

Damage Level	**Precision**	**Recall**	**F1 Score**	**Count**
No Damage	0.835	0.946	0.887	17493
Minor Damage	0.279	0.174	0.214	1949
Major Damage	0.428	0.086	0.143	1923
Destroyed	0.682	0.770	0.723	848

The EfficientNet-B3 classifier achieved 79.4% accuracy on the 22,213 detected building crops. Performance varied significantly across damage categories as observed in Table 11 above. The pipeline achieves strong performance on extreme damage categories, with "no damage" achieving 94.6% recall and F1-score of 0.887, while "destroyed" buildings achieve 77.0% recall and F1-score of 0.723. This binary-like performance suggests the pipeline can reliably distinguish between undamaged and severely damaged structures. However, for “minor damage” and “major damage”, critical failures are observed. The pipeline misses over 82% of “minor damage” cases and is, moreover, not usable for “major damage” category as a critically low recall is observed.

The integrated system successfully detected and classified 40.5% of all buildings in the test set. However, the 59.5% undetected buildings received no assessment, reflecting the central limitation of the two-stage design: classification performance is constrained by upstream detection coverage.

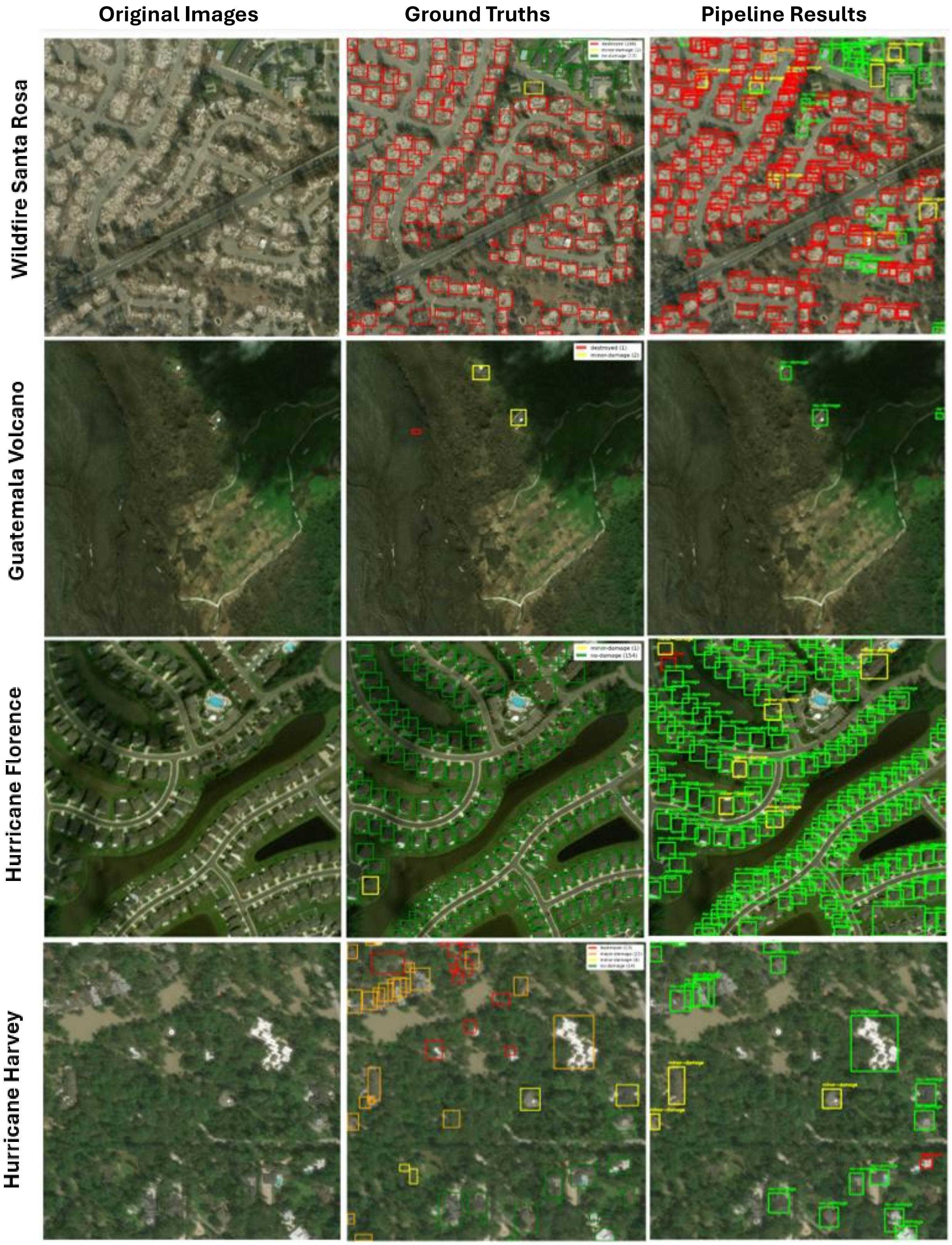


Figure 10: Two-Stage Pipeline Performance Visualisation with Ground Truth Comparison

Visual inspections of the pipeline outputs (Figure 10) across diverse scenarios highlights systematic errors. Detection performance correlates strongly with building boundary clarity. In urban areas (Santa Rosa Wildfire and Hurricane Florence) where distinct architectural features can be identified, the model achieves better detection rates than in areas with complex natural backgrounds (Guatemala Volcano and Hurricane Harvey).

False positive detections are evident across scenarios, with the pipeline misidentifying non-building objects such as vehicles, road intersections, and debris accumulations as buildings. Larger buildings have higher detection rates than smaller residential structures. Building orientation and partial occlusion by vegetation or shadows create additional detection challenges, particularly visible in the Hurricane Harvey forested areas.

Bounding box alignment remains a persistent issue: even for correctly detected buildings, overlaps with ground truth are often partial, with boxes shifted or imprecisely fitted.

An analysis of the classification shows pronounced bias towards “no damage” predictions (green). Categories such as “minor damage” and “major damage” rarely appear in outputs of the pipeline while “destroyed” buildings are detected with higher reliability. This behaviour effectively reduces the system to a binary classifier distinguishing between undamaged and severely damaged structures, while failing to capture the intermediate categories that are critical for nuanced disaster assessment.

6 Discussion

This study has developed and evaluated a two-stage building damage pipeline combining object detection and classification for automated post-disaster satellite imagery analysis. The research addresses four key questions about modern computer vision architectures for building damage assessment: performance of object detection on satellite images, the comparative performance of CNN versus transformer architectures, the effects of class imbalance and loss function selection, and the effectiveness of two-stage pipeline approaches. The research contributes empirical evidence about the effectiveness and limitations of combining object detection with classification for satellite-based building damage assessment.

6.1 Object Detection Performance

For building localisation, object detection has been used where the results demonstrate the substantial advantages of modern anchor-free architectures, with YOLOv8s achieving 60% better performance than traditional two-stage approaches like Faster R-CNN while also outperforming FCOS.

While many studies in building damage assessment literature employ semantic segmentation approaches using UNet-based architectures to generate dense pixel-wise building masks, these methods face distinct challenges including the inability to separate individual buildings within connected structures and significantly higher computational requirements for full-resolution processing (Li et al., 2021). Object detection was hence opted to distinguish building instances. However, we have observed that the model struggles in areas where clear architectural features are lacking, particularly in forested, agricultural or mountainous terrain where buildings lack sufficient contrast from natural backgrounds. The model also struggled with small buildings, which was an observation also made by Ghazouali et al. in their 2024 article. Conversely, the model performs relatively well in dense urban areas where distinct building boundaries and geometric patterns provide clear visual cues, though it occasionally misclassifies debris accumulations, road intersections, and other infrastructure elements as buildings, leading to a notable false positive burden.

6.2 Classification: CNN and Transformer Comparison

The architectural comparison reveals distinct performance patterns that address fundamental questions about transformer effectiveness in satellite image classification. CNN architectures have demonstrated superior adaptability to progressive fine-tuning strategies, with ResNet-50 and EfficientNet-B3 achieving consistent increase in performance across the different stages.

In contrast, ViT-B/16 achieved its peak performance accuracy in the initial fine-tuning stage but experienced significant degradation with extended training across all metrics. This counterintuitive result suggests that transformer architectures may require fundamentally different optimization strategies for satellite imagery tasks, with minimal parameter adjustment being more effective than aggressive fine-tuning. When evaluated on test set, DeiT-B/16, while less stable overall, achieved stronger recall for intermediate categories, outperforming CNNs in minor-damage and major-damage detection. Taken together, these results suggest that CNNs remain advantageous for stable, balanced feature extraction in satellite imagery, but that transformers may hold targeted value where sensitivity to severe or subtle damage cues is prioritised.

6.3 Classification: Class Imbalance and Loss Function Analysis

The loss function ablation study reveals critical trade-offs between overall classification performance and recall for severe damage categories that have significant implications for building damage assessment applications. Cross-Entropy achieved the best balance of accuracy and recall across all architectures, with ResNet-50 maintaining strong performance even on severe damage categories. However, Focal Loss demonstrated marginally superior recall for critical damage categories (1-2% improvement in major damage and destroyed classes) at the cost of substantial decreases in overall accuracy and F1-scores across all architectures. This performance trade-off presents a fundamental challenge: while missing severely damaged buildings could have life-threatening consequences, excessive false positives are not ideal either.

The implementation of Ordinal Loss, following the approach proposed by Weber and Kané (2020), produced performance between Cross-Entropy and Focal Loss. The weighted Cross-Entropy implementation had successfully addressed basic class imbalance issues, but the persistent poor performance on intermediate damage categories suggests that additional steps should be considered such as oversampling of classes where the model struggles ("minor damage" and "major damage"). Moreover, it can be inferred that the inherent difficulty lies in distinguishing subtle damage indicators at satellite resolution, requiring architectural or data collection innovations rather than training procedure modifications.

6.4 Two-Stage Pipeline Integration and Performance

Integrating YOLOv8s detection with EfficientNet-B3 classification highlights the inherent vulnerability of sequential architectures, where upstream errors irreversibly constrain downstream performance. Prior research has shown that multi-stage methods often suffer from error propagation, resulting in suboptimal outcomes (Gupta and Shah, 2020). In this study, the detection stage imposed two key constraints on classification: (1) missed buildings, which cannot be assessed for damage, and (2) false positives, which waste computational resources and introduce noise. The choice of a low confidence threshold (0.1) maximized recall but also generated a large volume of ambiguous detections, with most predictions falling below 0.7 confidence. This reflects a broader issue of confidence miscalibration in modern detectors, where probability scores are not reliable indicators of correctness and vary systematically with object size and context (Küppers et al., 2020). These uncertainties carried into the classification stage, collapsing intermediate categories, with recall for minor and major damage dropping to critically low levels. Consequently, the pipeline effectively operated as a binary classifier, distinguishing primarily between undamaged and destroyed buildings. This outcome illustrates a structural limitation of sequential designs: classification cannot recover information that is either absent or corrupted during detection, especially when subtle damage cues depend on precise localisation.

6.5 Comparison with Existing Literature

The methodological approach adopted in this study differs from predominant strategies employed in building damage literature. The most common approach for building damage assessment using satellite imagery is to pose the problem as a combination of segmentation and classification tasks (Gupta et al., 2019; Kaur et al., 2023; Shen et al., 2022), with many studies using semantic segmentation based on U-Net for building localisation to capture irregular shaped entities as opposed to object detection frameworks. Despite achieving state-of-the-art performances (Table 12), a drawback of semantic segmentation is its tendency to treat multiple instances of the same class as a single contiguous entity (Alisjahbana et al., 2024) which we attempted to solve through object detection. A further challenge was the wide variation in building sizes, ranging from small residential units to large commercial complexes, which made it difficult to learn scale-invariant features. This limitation is consistent with observations reported in BDANet (Shen et al., 2022).

Table 12: Performance Comparison (F1-Score) with Existing Literature on the xBD Dataset

	F1 (Overall)	**No Damage**	**Minor Damage**	**Major Damage**	**Destroyed**
xBD Baseline	0.265	0.663	0.144	0.009	0.466
Our model	0.503	0.887	0.214	0.143	0.723
BDANet	0.806	0.925	0.616	0.788	0.876
DaHiTrA	0.819	0.978	0.711	0.765	0.772

Table 12 compares the performance of our pipeline with existing literature on the xBD dataset. Our pipeline achieved an overall F1-score of 0.503, substantially lower than BDANet (0.806) and DaHiTrA (0.819). The strengths of our approach lie in its ability to discriminate extreme categories, performing competitively on no damage (0.887) and destroyed (0.723). However, it struggled with intermediate categories (minor = 0.214, major = 0.143), where BDANet achieved 0.616 and 0.788, respectively. These results highlight the trade-off between modular and end-to-end approaches: while object detection provides modularity and computational efficiency, state-of-the-art end-to-end models like DaHiTrA demonstrate clear advantages in avoiding sequential error propagation and maintaining higher overall accuracy across all categories.

6.6 Methodological Contributions

This study provides several methodological contributions to the field of building damage assessment from satellite imagery. Firstly, the comprehensive architectural comparison across both CNNs and transformers, combined with loss function ablation, represents one of the few empirical evaluations of transformer effectiveness for this task; prior studies have relied exclusively of CNN models or custom transformer architectures.

Secondly, the study introduces an adaptive polygon-based extraction method with metadata-only processing to address memory constraints while preserving building focus. This innovation enabled efficient handling of large-scale datasets that would otherwise exceed system capacity. Alongside this, the rigorous evaluation framework incorporated component-level testing, progressive fine-tuning strategies, and end-to-end pipeline assessment. Finally, the standardized experimental protocols and ablation studies establish reproducible baselines for future research, while the detailed analysis of error propagation in sequential architectures provides actionable insights for future research.

6.7 Limitations and Future Research Directions

We have identified several critical limitations that create opportunities to advance building damage assessment pipelines.

For building detection, the 40.5% coverage rate underscores the difficulty of reliably localising all structures, particularly small buildings and those in environments lacking clear architectural features. Future work could investigate extended training schedules, multi-scale feature enhancement, or instance segmentation approaches to capture finer boundaries, though these would come with higher computational cost.

For damage classification, performance on intermediate categories remains a fundamental challenge. Accuracy proved misleading under class imbalance, reinforcing the importance of prioritising F1-score as the primary optimisation metric and early stopping criterion. Further work could explore advanced sampling strategies (e.g., SMOTE) to balance class representation and strengthen performance on minority classes.

At the pipeline level, sequential architectures inherently suffer from error propagation, with upstream detection failures constraining downstream stages. End-to-end frameworks that directly predict damage levels for detected buildings may mitigate this limitation at the expense of modularity and component-level flexibility. Architectures like YOLT may prove beneficial as either an initial detection stage or an end-to-end pipeline.

Finally, confidence calibration represents an additional avenue for improvement. Detection performance was hindered by ambiguous confidence scores and unreliable calibration, complicating threshold selection. Evaluating calibration methods and incorporating higher-resolution or multi-source imagery may improve both confidence reliability and the visibility of small, low-contrast structures that dominate missed detections.

7 Conclusion

This research developed and evaluated a two-stage automated building damage assessment pipeline, integrating YOLOv8s detection with EfficientNet-B3 classification on the xBD dataset. The study addressed critical gaps in comparing CNN and transformer architectures for disaster response, while also proposing a scalable polygon-based extraction method for large-scale satellite data.

The work makes three key contributions: (1) demonstrating the continued advantage of CNNs in satellite imagery tasks while highlighting optimisation challenges for transformers, (2) introducing an adaptive, metadata-driven extraction approach that enables efficient and detailed dataset handling, and (3) establishing a reproducible evaluation framework that clarifies the trade-offs between modular two-stage designs and end-to-end architectures.

Overall, the findings underscore both the potential and limitations of sequential pipelines. While they offer computational efficiency and modularity, their vulnerability to error propagation highlights the need for future research focused on improving building detection coverage, enhancing classification of intermediate damage categories, and developing more robust learning strategies for imbalanced and visually ambiguous satellite data.

References

Alisjahbana, I., Li, J., Ben, Strong, Zhang, Y., 2024. DeepDamageNet: A two-step deep-learning model for multi-disaster building damage segmentation and classification using satellite imagery. https://doi.org/10.48550/arXiv.2405.04800

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., Inman, D.J., 2021. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mech. Syst. Signal Process. 147, 107077. https://doi.org/10.1016/j.ymssp.2020.107077

Bashmal, L., Bazi, Y., Rahhal, M.A., 2021. Deep Vision Transformers for Remote Sensing Scene Classification, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Presented at the IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Brussels, Belgium, pp. 2815–2818. https://doi.org/10.1109/IGARSS47720.2021.9553684

Bhardwaj, D., Nagabhooshanam, N., Singh, A., Selvalakshmi, B., Angadi, S., Shargunam, S., Guha, T., Singh, G., Rajaram, A., 2024. Enhanced satellite imagery analysis for post-disaster building damage assessment using integrated ResNet-U-Net model. Multimed. Tools Appl. 84, 2689–2714. https://doi.org/10.1007/s11042-024-20300-0

Cao, W., Mirjalili, V., Raschka, S., 2020. Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognit. Lett. 140, 325–331. https://doi.org/10.1016/j.patrec.2020.11.008

Chavez-Demoulin, V., Jondeau, E., Mhalla, L., 2021. Climate-Related Disasters and the Death Toll. https://doi.org/10.48550/arXiv.2109.02111

Doshi, J., Basu, S., Pang, G., 2018. From Satellite Imagery to Disaster Insights. https://doi.org/10.48550/arXiv.1812.07033

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929

Ghazouali, S.E., Gucciardi, A., Venturini, F., Venturi, N., Rueegsegger, M., Michelucci, U., 2024. FlightScope: An Experimental Comparative Review of Aircraft Detection Algorithms in Satellite Imagery. Remote Sens. 16, 4715. https://doi.org/10.3390/rs16244715

Groener, A., Chern, G., Pritt, M., 2019. A Comparison of Deep Learning Object Detection Models for Satellite Imagery, in: 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). pp. 1–10. https://doi.org/10.1109/AIPR47015.2019.9174593

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2019. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. https://doi.org/10.48550/arXiv.1911.09296

Gupta, R., Shah, M., 2020. RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery. https://doi.org/10.48550/arXiv.2004.07312

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385

Hoque, O.B., Adiga, Abhijin, Adiga, Aniruddha, Chaudhary, S., Marathe, M.V., Ravi, S.S., Rajagopalan, K., Wilson, A., Swarup, S., 2025. IGraSS: Learning to Identify Infrastructure Networks from Satellite Imagery by Iterative Graph-constrained Semantic Segmentation. https://doi.org/10.48550/arXiv.2506.08137

Kaur, N., Lee, C.-C., Mostafavi, A., Mahdavi-Amiri, A., 2023. Large-scale Building Damage Assessment using a Novel Hierarchical Transformer Architecture on Satellite Images. https://doi.org/10.48550/arXiv.2208.02205

Küppers, F., Kronenberger, J., Shantia, A., Haselhoff, A., 2020. Multivariate Confidence Calibration for Object Detection, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1322–1330. https://doi.org/10.1109/CVPRW50498.2020.00171

Le, C., Pham, L., NVN, N., Nguyen, T., Trang, L.H., 2022. A Robust and Low Complexity Deep Learning Model for Remote Sensing Image Classification. https://doi.org/10.48550/arXiv.2211.02820

Le, T.-D., Ha, V.N., Nguyen, T.T., Eappen, G., Thiruvasagam, P., Chou, H., Tran, D.-D., Nguyen-Kha, H., Garces-Socarras, L.M., Gonzalez-Rios, J.L., Merlano-Duncan, J.C., Chatzinotas, S., 2025. Onboard Satellite Image Classification for Earth Observation: A Comparative Study of ViT Models. https://doi.org/10.48550/arXiv.2409.03901

Li, R., Zheng, S., Zhang, C., Duan, C., Wang, L., Atkinson, P.M., 2021. ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of Fine-Resolution remotely sensed imagery. ISPRS J. Photogramm. Remote Sens. 181, 84–98. https://doi.org/10.1016/j.isprsjprs.2021.09.005

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal Loss for Dense Object Detection, in: 2017 IEEE International Conference on Computer Vision (ICCV). Presented at the 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324

Maniyar, C.B., Kumar, M., Mai, G., 2025. Feature-Augmented Deep Networks for Multiscale Building Segmentation in High-Resolution UAV and Satellite Imagery. https://doi.org/10.48550/arXiv.2505.05321

Mommert, M., Scheibenreif, L., Hanna, J., Borth, D., 2021. Power Plant Classification from Remote Imaging with Deep Learning. https://doi.org/10.48550/arXiv.2107.10894

Morozov, V., Galliamov, A., Lukashevich, A., Kurdukova, A., Maximov, Y., 2023. CMIP X-MOS: Improving Climate Models with Extreme Model Output Statistics. https://doi.org/10.48550/arXiv.2311.03370

Proma, A.M., Islam, M.S., Ciko, S., Baten, R.A., Hoque, E., 2022. NADBenchmarks -- a compilation of Benchmark Datasets for Machine Learning Tasks related to Natural Disasters. https://doi.org/10.48550/arXiv.2212.10735

Qi, T., Xie, H., Li, P., Ge, J., Zhang, Y., 2023. Balanced Classification: A Unified Framework for Long-Tailed Object Detection. https://doi.org/10.48550/arXiv.2308.02213

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/arXiv.1505.04597

Saricayir, B., Ozcan, C., 2024. EfficientNet Deep Learning Model for Satellite Image Classification Using the EuroSAT Dataset.

Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., Dar, S.H., 2021. Satellite and Scene Image Classification Based on Transfer Learning and Fine Tuning of ResNet50. Math. Probl. Eng. 2021, 1–18. https://doi.org/10.1155/2021/5843816

Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., Xiao, L., Du, Q., 2022. BDANet: Multiscale Convolutional Neural Network With Cross-Directional Attention for Building Damage Assessment From Satellite Images. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3080580

Shermeyer, J., Etten, A.V., 2019. The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. https://doi.org/10.48550/arXiv.1812.04098

Tan, M., Le, Q.V., 2020. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1905.11946

Tian, Z., Shen, C., Chen, H., He, T., 2019. FCOS: Fully Convolutional One-Stage Object Detection. https://doi.org/10.48550/arXiv.1904.01355

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., 2021. Training data-efficient image transformers & distillation through attention. https://doi.org/10.48550/arXiv.2012.12877

Trekin, A., Novikov, G., Potapov, G., Ignatiev, V., Burnaev, E., 2018. Satellite imagery analysis for operational damage assessment in Emergency situations. https://doi.org/10.48550/arXiv.1803.00397

Wang, M., Su, L., Yan, C., Xu, S., Yuan, P., Jiang, X., Zhang, B., 2024. RSBuilding: Towards General Remote Sensing Image Building Extraction and Change Detection with Foundation Model. https://doi.org/10.48550/arXiv.2403.07564

Weber, E., Kané, H., 2020. Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion. https://doi.org/10.48550/arXiv.2004.05525

Xu, J.Z., Lu, W., Li, Z., Khaitan, P., Zaytseva, V., 2019. Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1910.06444

Zhang, J., Zhang, K.K., Zhang, M., Jiang, J.H., Rosen, P.E., Fahy, K.A., 2022. Avoiding the "Great Filter": An assessment of climate change solutions and combinations for effective implementation. Front. Clim. 4, 1042018. https://doi.org/10.3389/fclim.2022.1042018

Appendix A

All the python notebooks for each stage (Localisation and Classification) are provided in the Google Drive Folder below.

Google Drive Folder:
https://drive.google.com/drive/folders/1itws_DFUNdQqGB4SnW3CUmiPCv1vM-da?usp=drive_link

The folder includes the following notebooks:

1. Annotations for Object Detection Model Training
2. Object Detection Models
3. Crops Extraction Methods explored
4. Damage Classification Models
5. Pipeline Evaluation on Test Data
6. Pipeline for Single Image Processing

Alongside the python notebooks, the weights of the best models are provided, along with an example of an image to run the pipeline.

The instructions to run the code are provided in a text file. The xBD dataset is needed to run the code (model training and test set evaluation) successfully. The dataset can be downloaded at: https://xview2.org/

Another useful dataset for damage assessment for earthquake mainly: https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-4169

These resources are intended to ensure reproducibility and support future extensions of this work.

Appendix B

Project Proposal

Title: Advancing Building Damage Assessment from Satellite Imagery: Evaluating Modern Computer Vision Approaches for Disaster Response

Supervisor: Dr Giacomo Tarroni

INTRODUCTION

The aim of this project to evaluate the effectiveness of modern computer vision architectures beyond CNNs in building damage assessment from satellite imagery, comparing their performance against traditional approached to advance disaster response capabilities. Current disaster protocols typically require dangerous in-person inspections within 24-48 hours of a disaster event. With the advent of computer vision, we have seen the implementation of CNNs, which have been mostly successful and essential for the increasing number of disasters globally, natural and man-made.

Building damage assessment has been approached as a pixel-level semantic segmentation task using CNN-based backbones. While these methods have proven effective, they present several limitations in operational deployment. Segmentation-based approaches require dense pixel-level predictions across entire images, leading to high computational costs during both training and inference. Additionally, CNNs process images through local receptive fields, potentially missing important spatial relationships and global damage patterns that are crucial for accurate assessment. Recent advances in computer vision have introduced new architectures that offer unique advantages through novel attention mechanisms and ability to capture long-range spatial dependencies.

Research question, objectives and outcomes

Effective disaster response requires rapid and accurate building damage assessment to guide emergency resources and recovery efforts. In post-disaster scenarios where computational resources may be limited and speed is essential, any improvement in efficiency and accuracy can be crucial for successful emergency operations. The emergence of new architectures offers potential advantages, but their effectiveness compared to established methods remains unclear. This leads to the question of: How do modern computer vision architectures beyond CNNs compare to traditional CNN architectures for building damage classification from satellite imagery?

The objectives are to evaluate the performance of different modern architectures against CNN baselines for building damage classification and analyse the trade-offs.

The main outcome would be a comprehensive comparison of modern and traditional architectures for building damage assessment and insights into how new architecture perform in the domain. Emergency Response Teams would benefit from a rapid and accurate building damage assessment enables better resource allocation during disaster response. They will be equipped with tools more effective in resource-constrained environments. Disaster management agencies would obtain an operational system for future disaster response. Finally, this work would contribute toward the research community whereby modern architectures are applied in remote sensing and disaster response.

CRITICAL CONTEXT

Building damage assessment

Most works treat this task as a per-pixel semantic segmentation task and rely on CNN-based backbones. They are effective, but do not explore modern transformer-based architectures or instance-level approaches. Gupta et al. (2019) introduced the xBD dataset for assessing building damage from Satellite imagery, containing pre- and post-disaster images. It is a large scale benchmark, including over 850,000 annotated building polygons labelled with a four-level ordinal damage scale across different disasters and diverse regions. Importantly for this project, the xBD dataset also provides a baseline two-stage pipeline using ResNet-based U-Net for building segmentation followed by a ResNet50 classifier for damage assessment, highlighting early reliance on CNNs. However, the annotations of the dataset contain valuable information making it ideal for evaluating modern architectures such as Vision Transformers in an instance-level classification setting, where polygons annotations are directly leveraged.

Shen et al. (2022) proposed BDANet; a two-stage CNN-based framework for building damage assessment from the xBD dataset. Stage 1 consists of U-Net for building segmentation and Stage 2 apples a dual-branch architecture for damage classification with cross-directional attention and multiscale feature fusion. Although BDANet shows strong performance on the xBD dataset, it reinforces the dominance of segmentation-based pipelines that rely on dense pixel-level prediction. In contrast, we would focus on instance-level classification using building polygons, offering a simpler and operationally efficient alternative which would be valuable in real-world applications where building footprints are already available.

Similarly, Chen (2020) proposes a two-stage pipeline with U-Net for segmentation followed with classification applied to cropped building patches. The classification step leverages features from both the segmentation encoder and a ResNet18 model. The author evaluated different loss functions and through ablation, he found that Dice and Focal loss were the most efficient. While the method achieves strong performance (F1 = 0.63), it relies on CNNs and does not explore the potential of transformer-based models in an instance-level classification.

We observe a shift in the methods used in building damage assessment from a two-stage pipeline of building localisation and damage classification to unified semantic segmentation models that predict building presence and damage simultaneously. This shift has overlooked the simplicity and explainability of instance-level classification, especially in contexts where building footprints are available.

Weber and Kan (2020) explored a multi-temporal (using both pre and post images) damage assessment system. They experimented with Instance Segmentation and Semantic Segmentation and identified that the latter is a “more natural damage-assessment formulation without the notion of instances” as buildings are too small for instance segmentation. As most of the studies conducted, they also used a cross-entropy loss function. Their approach fuses feature before the final segmentation layer to jointly predict pixel-wise damage levels, demonstrating strong performance on the xBD dataset. However, the method remains segmentation-based and does not exploit instance-level modelling or transformer architectures. The authors also suggest the use of ordinal cross-entropy loss function which would penalise predictions heavily when the predicted damage level is further from the true class.

Neto and Dantas (2024) also implemented a one-stage CNN-based pipeline. They performed building damage assessment using segmentation architecture. Their study explores approaches for handling pre- and post-disaster satellite images, including stacked inputs, Siamese networks. The best performance was achieved using a U-Net model with a BDANet backbone. They further refined the results by applying mathematical morphology operations (dilation, erosion and combinations of both) as post-processing filters to reduce noise and fill gaps. Most of these studies utilised CNN-based architectures with mainly Focal and Cross-Entropy loss functions and Dice Loss for segmentation. In this project, we explore the viability of transformer-based architectures against different CNN in instance-level classification and implementation of different loss functions.

Classification at instance level has been explored in other research Kaur et al. (2022) proposed a CNN-based binary classifier to detect damaged vs. undamaged buildings from post-disaster satellite imagery of Hurricane Harvey. Their model achieved 95% accuracy on cropped building images, demonstrating the feasibility of instance-level classification pipelines. However, the approach is limited to binary classification and relies on a shallow CNN. This project extends this direction by exploring transformer-based architectures and ordinal damage classification.

Vision Transformers (ViT)

Dosovitskiy et al. (2021) introduced Vision Transformers (ViT), bringing major development in computer vision by directly applying transformer architectures to image analysis with great success. By dividing images into fixed patch sizes and processing them with self-attention mechanisms, ViT is capable of global receptive fields earlier in the network compared to CNNs.

Despite ViTs lack of inherent inductive biases like translation equivariance, Dosovitskiy et al. demonstrated that "large scale training trumps inductive bias". ViT models can outperform CNNs when pretrained on sufficient data while using less computational resources. This efficiency could prove beneficial for complex satellite imagery analysis.

The attention mechanism in ViTs enables identification of image regions “semantically relevant” for classification, suggesting natural aptitude for identifying damage patterns in buildings. Furthermore, ViT shows excellent transfer learning capabilities when pretrained on sufficient data and transferred to task with fewer datapoints; which is vital for specialised domain with limited annotated data.

These advantages set ViT as a promising architecture for building damage assessment, where integrating global context and efficiently processing high-resolution imagery are essential for accurate classification across damage categories. This project aims to evaluate ViT's effectiveness for this critical application compared to established CNN approaches.

Vision Transformers in Satellite Imagery Analysis

CNN remains the cornerstone for satellite imagery analysis. However, transformer-based architectures have demonstrated higher performance recently in tasks like land cover classification by Voelsen et al. (2024) and crop segmentation by Gallo et al. (2024).

Vision Transformers (ViT) in Building damage assessment

There have been several works done where transformers have been used for building damage assessment. For instance, Chen et al. (2022) introduced DamFormer, a transformer-based framework. It employs a siamese transformer encoder built on SegFormer's architecture to learn global features. In contrast, CNN struggles with global features and would require attention mechanisms or dilated convolution which are costly. The model performs dual tasks: building localisation and damage classification within a unified end-to-end architecture. Their experiments on the xBD dataset demonstrate that transformer-based models outperform CNN approaches across all damage levels.

Another state-of-the-art building damage assessment framework is DAHiTra, a hierarchical transformer architecture, introduced by Kaur et al. (2024). The model's key innovation is its difference block mechanism, which explicitly captures temporal changes between pre- and post-disaster imagery at multiple scales. DAHiTrA achieves superior performance on both the xBD dataset and the Ida-BD dataset, demonstrating the effectiveness of transformer-based approaches.

While these transformer-based models represent significant advances in end-to-end semantic segmentation for damage assessment, they exemplify a trend toward increasingly complex architectures that simultaneously perform building localisation and damage classification. Our research takes a different approach by focusing specifically on instance-level damage classification using Vision Transformers. By utilizing pre-defined building polygons from the xBD dataset, we aim to evaluate the performance of ViT variants against traditional CNN architectures, while simultaneously exploring various loss functions to address the class imbalances. This approach would offer greater interpretability of model decisions, reduced computational requirements during inference, and direct applicability in real-world scenarios where building footprints are already available from GIS databases.

APPROACHES

- **Dataset**

We will use the xView2: Assess Building Damage (xBD) dataset, a large-scale benchmark dataset, consisting of high-resolution satellite imagery [ ]. These images are form 19 diverse natural disaster events, including hurricanes, floods, earthquakes, volcanic eruptions, wildfires and tsunamis, across six continents. The dataset contains 22,068 images and more than 800,000 building polygons, making it the most comprehensive public dataset for building damage assessment. The dataset is formally structured with pre-defined splits (training, holdout and testing). This structured split ensures robust evaluation by testing models on unseen disaster events, which is crucial for assessing generalization capabilities.

The dataset provides paired imagery for each location (pre- and post-disaster) where each image is accompanied by detailed annotations including:

1. Building polygons represented in WKT (Well-Known Text) format, consisting of vertex coordinate pairs outlining individual buildings
2. Dual coordinate systems: geographic (lat-lon) and pixel (x, y) to facilitate both geospatial analysis and computer vision tasks
3. Unique identifiers (UIDs) assigned to each building to enable tracking across pre- and post-disaster imagery
4. Damage classification labels according to the Joint Damage Scale in post-disaster images:
   No Damage (0): Building appears intact
   Minor Damage (1): Visible damage to the roof or sides, but structure remains standing
   Major Damage (2): Significant structural damage but partial roof/walls remaining
   Destroyed (3): Complete collapse or only foundation remains

The dataset exhibits significant class imbalance, with most of buildings labelled as undamaged, while destroyed buildings constitute only about 10% of the samples. This imbalance presents a key challenge that our research will address through specialized loss functions such as class weighted focal loss. Additionally, diverse disaster types and geographical regions introduce variations in building styles, sizes, and contextual features that test the generalization capabilities of damage assessment models.

For the instance-based classification approach, we will extract individual building patches using the provided polygon coordinates, maintain the same split organization of the dataset to ensure no data leakage between training, holdout and test evaluations. Varying building sizes would be addressed through adaptive cropping strategies.

- **Model architectures** (the models we are planning to evaluate and compare)

Convolutional Neural Network (CNN)

1. Baseline (ResNet)

We select ResNet as our traditional CNN baseline due to its widespread use in building damage assessment tasks and proven effectiveness in handling the complexity of satellite imagery in Weber and Kan (2020) and Chen (2020). The residual blocks address the vanishing gradient problem, allowing for deeper networks while maintaining computational efficiency.

2. Lightweight CNN (EfficientNet)

EfficientNet is considered as the majority of satellite imagery analysis implemented CNN such as VGGNet and ResNet. To our best knowledge, lightweight models have not been explored on the xBD dataset. The lightweight model would also be a good comparison with ViT models, as Dosovitskiy et al. (2021) mentioned that ViT showed efficient performance results. The model would also be relevant in cases where computational resources are limited as it would require significantly less parameters compared to ResNet.

Vision Transformers (ViT)

1. ViT-B/16 (Vision Transformer Base 16x16)

ViT-B/16 serves as our transformer baseline, representing the foundational ViT architecture. This model offers global receptive field from the first layer through self-attention mechanisms and can capture long-range dependencies. Furthermore, ViTs are relatively new in the computer vision field and as such these architectures have not been explored much in the field of satellite imagery. We may experiment with other ViT variants such as the 'big' and 'huge' ViT models.

2. DeiT (Data-Efficient Transformer)

DeiT is being considered in our experimental framework as a more practical approach. While standard ViTs require massive datasets and computational resources, DeiT achieves competitive performance through knowledge distillation with less requirements. This makes it particularly relevant for real-world deployment scenarios where computational constraints exist. The success of DeiT in remote sensing applications (Bashmal et al., 2021) demonstrates its transferability to satellite imagery tasks, making it an ideal candidate for evaluating whether the efficiency gains can be maintained while achieving competitive accuracy in building damage classification.

The models would be implemented through transfer learning and will be modified for four levels of damage classification. Additional models may be explored such as Swin and MobileNet.

- **Methodology**

1. CNN implementation

Using PyTorch or Keras framework, we would implement our base CNN model using transfer learning with ImageNet pre-trained weights. We will apply a fine tuning strategy where we initially freeze all convolutional layers and only train the classification layer for a specific number of epochs. In the classification block, the output is replaced to be four damage classes. We will progressively unfreeze deeper layers, starting from the last convolutional block and use different learning rates for the classification and convolutional layer. We will compare performance.

We would follow a similar approach for the lightweight CNN model (EfficientNet) and monitor improvements in performance.

2. ViT implementation

For ViT (ViT-B/16 and DeiT), we will adapt pre-trained models originally trained on ImageNet for our building damage classification task. The final classification head will be replaced to output four damage classes while maintaining the core transformer architecture.

We will similarly apply a staged fine-tuning approach where we initially freeze the patch embedding layers and transformer blocks, training only the classification(MLP) layer. We will progressively unfreeze transformer blocks starting from the last layer and implementing different learning rate where deeper transformer blocks receive lower learning rates.

3. Training Configurations

To optimise the models, we may implement learning rate scheduling strategies such step or change on plateau. To mitigate overfitting, we may use early stopping, gradient clipping and other regularization techniques (weight decay, dropout). Batch size will vary depending on memory constraints. All experiments will be conducted with fixed random seeds to ensure reproducibility, and the hardware configurations will be provided. We will explore different loss functions. We will initially run our baseline models with cross-entropy and through ablation, we would find the optimal loss function for each model. Weber and Kan (2020) observed that class weighted loss functions and ordinal loss functions could potentially improve the models. We would, therefore, explore class weighted cross-entropy, focal loss and ordinal loss to address class imbalance.

4. Evaluation

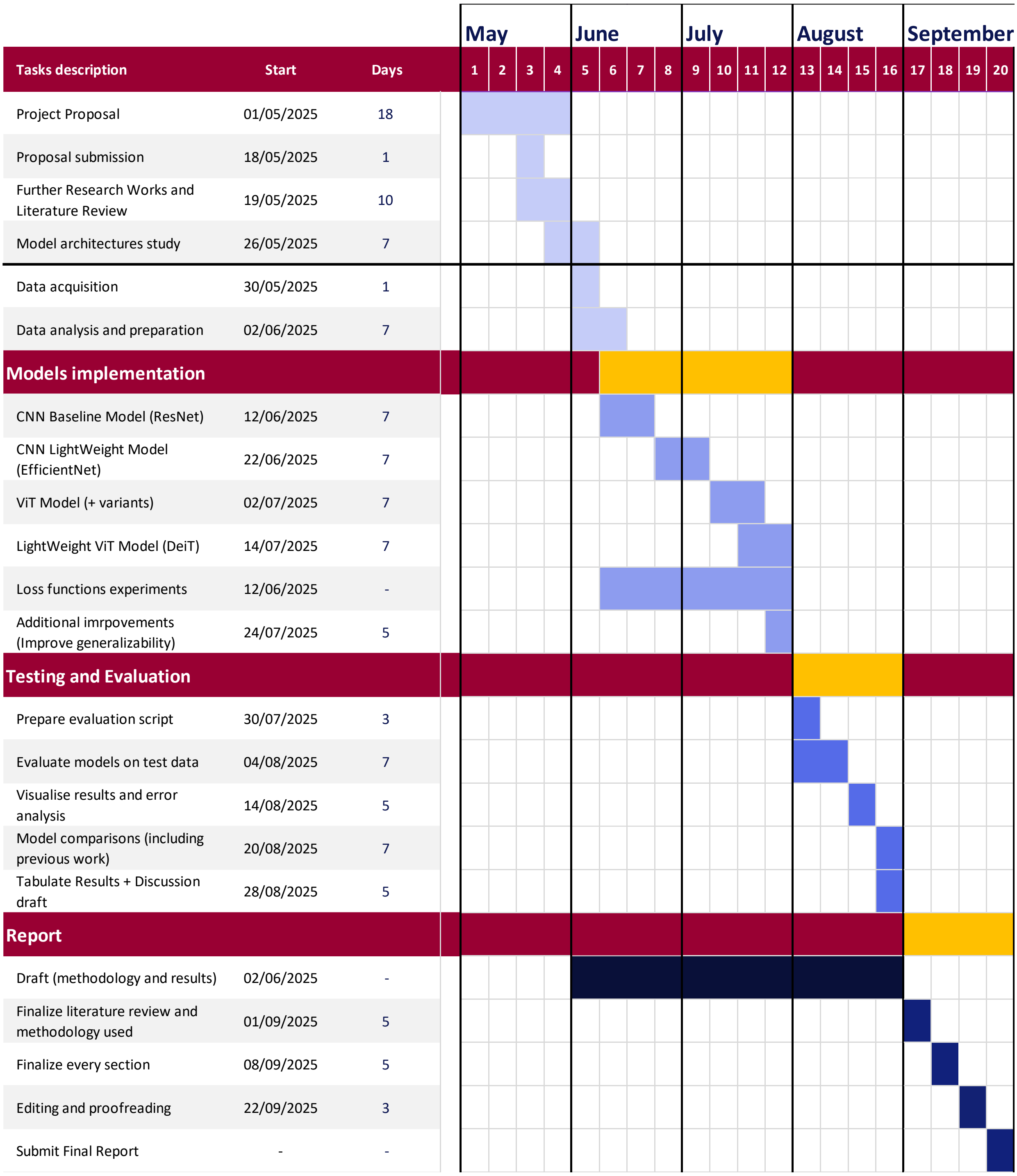
The performance evaluation metrics that we plan to use are overall accuracy and F1-score mainly. Accuracy is to identify the percentage of correctly classified buildings across all damage categories, and F1-score is used due to class imbalance. Confusion matrix would provide a detailed breakdown of predictions and would help identify common misclassification patterns. We would also record training time for each model for comparison. Statistical analysis will be conducted to evaluate model performance differences and ensure reliable comparison across architectures. Appropriate statistical measures will be applied to assess the significance of our results.

Generalizability assessment

To ensure the practical applicability of our findings, we will evaluate the cross-domain generalizability of our trained models. This will involve testing the CNN and ViT architectures on datasets from different geographical regions, disaster types, or imaging conditions than those present in the xBD training set. This cross-domain evaluation will help assess which architecture demonstrates better robustness.

WORK PLAN

Figure 1 provides an estimate of the duration of each task of the project. Data preprocessing must be completed before model training can begin. CNN baseline and ViT baseline training can proceed in parallel once preprocessing is finished. Comparative evaluation depends on completion of all model training phases.



Tasks description	Start	Days
Project Proposal	01/05/2025	18
Proposal submission	18/05/2025	1
Further Research Works and Literature Review	19/05/2025	10
Model architectures study	26/05/2025	7
Data acquisition	30/05/2025	1
Data analysis and preparation	02/06/2025	7
Models implementation		
CNN Baseline Model (ResNet)	12/06/2025	7
CNN LightWeight Model (EfficientNet)	22/06/2025	7
ViT Model (+ variants)	02/07/2025	7
LightWeight ViT Model (DeiT)	14/07/2025	7
Loss functions experiments	12/06/2025	-
Additional imrpovements (Improve generalizability)	24/07/2025	5
Testing and Evaluation		
Prepare evaluation script	30/07/2025	3
Evaluate models on test data	04/08/2025	7
Visualise results and error analysis	14/08/2025	5
Model comparisons (including previous work)	20/08/2025	7
Tabulate Results + Discussion draft	28/08/2025	5
Report		
Draft (methodology and results)	02/06/2025	-
Finalize literature review and methodology used	01/09/2025	5
Finalize every section	08/09/2025	5
Editing and proofreading	22/09/2025	3
Submit Final Report	-	-

Figure 1: Gantt Chart

RISKS

In table 1 below, we have identified a set of potential risks for the project, using Dawson's (2006) framework. We have outlined the risks and alleviating strategies. The table will be updated as the project progresses.

Risk Description	Likelihood	Consequence	Impact(LxC)	Mitigation and Contingency
Tasks take longer than estimated	2	3	6	Conservative planning and weekly checkpoint using Gantt Chart. Reallocate hours from optional tasks to prioritize core experiments, if behind.
Class imbalances affect performance	4	3	12	Implement weighted loss functions, data augmentation and oversampling. If performance remains low, we will focus on the results for each class (using recall, precision)
Insufficient GPU resources	3	4	12	Consider Google Colab or Cloud services. Reduce batch sizes. We can also consider reducing image resolution and switching to more efficient models.
Poor performance of satellite imagery domain	2	4	8	Extensive fine-tuning or identify potential domain specific pre-trained models and data augmentation techniques. We can also consider further customisation of the models.
Limited datasets with both building masks and damage labels for generalizability assessment	4	3	12	Identify potential datasets during the 'Further Research Work', exploring other publicly available datasets (ABCD, DesignSafe). If we have not been able to find any, we could potentially create small manually annotated validation set from available building masks.
Illness or personal circumstances	2	4	8	Maintain detailed documentation and plan buffer time in schedule. Prioritize core tasks and seek extension in necessary. If prolonged absence, we would consider reassessing the scope of the project.
Loss of data	1	5	5	Establish a cloud back up strategy. If data loss occurs, we would need to redownload the dataset and attempt to restore most recent version of the work.

Table 1: Risks breakdown

References

Bashmal, L., Bazi, Y. and Al Rahhal, M., 2021. *Deep vision transformers for remote sensing scene classification*. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp.2815–2818. doi:10.1109/IGARSS47720.2021.9553684

Chen, X., 2020. *Building Damage Detection from Satellite Imagery*. [online] Stanford University https://web.stanford.edu/~markcx/sample-project/IISE_Building_Damage_Detection_from_Satellite_Imagery.pdf [Wednesday 30/05/2025]

Chen, L., Lin, H., Yang, H., Xu, X. and Chen, Z., 2022. *DamFormer: Dual-task Siamese Transformer Framework for Building Damage Assessment*. arXiv preprint arXiv:2211.06842. doi.org/10.48550/arXiv.2201.10953

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N., 2021. *An image is worth 16x16 words: Transformers for image recognition at scale*. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2010.11929

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E. and Choset, H., 2019. *Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery*. arXiv preprint arXiv:1911.09296. https://arxiv.org/abs/1911.09296

Kaur, G., Chu, W.T., Singh, A., Kumar, D., Qadir, J. and Khattak, A.M., 2022. *A deep learning-based approach for post-disaster damage assessment using satellite imagery*. Computers, Materials & Continua, 73(1), pp.1453–1467. doi:10.1007/s00500-022-06805-6

Kaur, P., Kaur, P., Jain, S., Kaul, A., Patel, V.M. and Ramachandra, B., 2024. *DAHiTrA: Damage assessment using hierarchical transformer architecture*. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2345–2354.

Neto, P.R. and Dantas, R.A.T., 2024. *Building damage segmentation after natural disasters in satellite images with mathematical morphology and convolutional neural networks*. Computers, Environment and Urban Systems, 102, p.101001.

Shen, C., Li, L., Li, D. and Li, S., 2022. *BDANet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images*. Remote Sensing, 14(4), p.808 10.1109/TGRS.2021.3080580

Voelsen, J., Heyer, T., Tuia, D. and Persello, C., 2024. *Transformer models for land cover classification with satellite image time series*. Computers & Geosciences, 183, p.105482.

Weber, E. and Kan, H., 2020. *Building disaster damage assessment in satellite imagery with multi-temporal fusion*. arXiv preprint arXiv:2009.12556. https://arxiv.org/abs/2004.05525

Gallo, I., Capobianco, R., Marozzo, F., Salvi, M. and Roverelli, M., 2024. *Self-attention-based transformer networks for semantic segmentation of crops from satellite imagery*. Remote Sensing, 16(3), p.451.

Braik, M. and Koliou, M., 2024. *Automated building damage assessment and large-scale mapping by integrating convolutional neural networks with GIS building data*. Computer-Aided Civil and Infrastructure Engineering, 39(3), pp.466–483.

Dawson, C. W. (2009). *Projects in Computing and Information Systems: A Student's Guide (2nd ed.)*. London: Addison Wesley, 304pp.

Research Ethics Review Form for MSc Projects

Computer Science Research Ethics Committee (CSREC)

https://www.city.ac.uk/about/governance/committees/cs-research-ethics

Postgraduate students undertaking their final project in the Department of Computer Science must consider the ethics of their project work and ensure that it complies with research ethics guidelines and the law for data protection. In some cases, a project will need approval from an ethics committee before it can proceed. Usually, but not always, this will be because the student is involving other people ("participants") in the project.

To ensure that they give appropriate consideration to ethical issues, all students must complete this form and attach it to their project proposal document. There are two parts:

PART A: Ethics Checklist. All students must complete this part.
The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval.

PART B: Ethics Proportionate Review Form. Students who have answered "no" to all questions in A1, A2 and A3 and "yes" to question 4 in A4 in the ethics checklist must complete part B as well. The project supervisor has delegated authority to provide approval in such cases that are considered to involve MINIMAL risk. The approval may be ***provisional*** – *identifying the planned work with human end user participants as likely* to involve MINIMAL RISK. In such cases you must additionally seek ***full approval*** from the supervisor as the project progresses and details are established. ***Full approval*** must be acquired in writing, before recruiting and engaging with human end users participants for your project.

A.1 If you answer YES to any of the questions in this block, you must apply to an appropriate external ethics committee for approval and log this approval as an External Application through Research Ethics Online - https://researchmanager.city.ac.uk/. This type of research is not covered by City's process, and external approval from an appropriate institution is required.		*Delete as appropriate*
1.1	Does your research require approval from the National Research Ethics Service (NRES)? *e.g. because you are recruiting current NHS patients or staff?* *If you are unsure try -* https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/	**NO**
1.2	Will you recruit participants who are covered by the Mental Capacity Act 2005? *Such research needs to be approved by an external ethics committee such as NRES or the Social Care Research Ethics Committee -* http://www.scie.org.uk/research/ethics-committee/	**NO**
1.3	Will you recruit any participants who are covered by the Criminal Justice System, for example, people on remand, prisoners and those on probation? *Such research needs to be authorised by the ethics approval system of the National Offender Management Service.*	**NO**

A.2 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee, you must apply for approval from the Senate Research Ethics Committee (SREC) through Research Ethics Online - https://researchmanager.city.ac.uk/		*Delete as appropriat e*
2.1	Does your research involve participants who are unable to give informed consent? *For example, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf.*	**NO**
2.2	Is there a risk that your research might lead to disclosures from participants concerning their involvement in illegal activities?	**NO**
2.3	Is there a risk that obscene and or illegal material may need to be accessed for your research study (including online content and other material)?	**NO**
2.4	Does your project involve participants disclosing information about protected characteristics (as identified by the Equality Act 2010)? *For example, to: racial or ethnic origin; political opinions; religious beliefs; trade union membership; physical or mental health; sexual life; criminal offences and proceedings*	**NO**
2.5	Does your research involve you travelling to another country outside of the UK, where the Foreign & Commonwealth Office has issued a travel warning that affects the area in which you will study? *Please check the latest guidance from the FCO - http://www.fco.gov.uk/en/*	**NO**
2.6	Does your research involve invasive or intrusive procedures? *These may include, but are not limited to, electrical stimulation, heat, cold or bruising.*	**NO**
2.7	Does your research involve animals?	**NO**
2.8	Does your research involve the administration of drugs, placebos or other substances to study participants?	**NO**
A.3 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee or the Senate Research Ethics Committee (SREC), you must apply for approval from the Computer Science Research Ethics Committee (CSREC) through Research Ethics Online - https://researchmanager.city.ac.uk/**. Depending on the level of risk associated with your application, it may be referred to the Senate Research Ethics Committee.**		*Delete as appropriat e*
3.1	Does your research involve participants who are under the age of 18?	**NO**
3.2	Does your research involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)?	**NO**

	This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people.	
3.3	Are participants recruited because they are staff or students of City, University of London? *For example, students studying on a particular course or module.* *If yes, then approval is also required from the Head of Department or Programme Director.*	**NO**
3.4	Does your research involve intentional deception of participants?	**NO**
3.5	Does your research involve participants taking part without their informed consent?	**NO**
3.5	Is the risk posed to participants greater than that in normal working life?	**NO**
3.7	Is the risk posed to you, the researcher(s), greater than that in normal working life?	**NO**

A.4 If you answer YES to the following question and your answers to all other questions in sections A1, A2 and A3 are NO, then your project is deemed to be of MINIMAL RISK. **If this is the case, then you can apply for approval through your supervisor under PROPORTIONATE REVIEW. You do so by completing PART B of this form.** **If you have answered NO to all questions on this form, then your project does not require ethical approval. You should submit and retain this form as evidence of this.**		*Delete as appropriate*
4	Does your project involve human participants or their identifiable personal data? *For example, as interviewees, respondents to a survey or participants in testing.*	**NO**