CAS2105 Homework 6: Mini AI Pipeline Project 🤗

Reece Jones (2025848556)

1 Introduction

Assessment of structural damage following natural disasters such as hurricanes, tornadoes, and typhoons is critical for coordinating emergency response. Manual inspection of satellite imagery can be slow and labour intensive. This project attempts to explore an automated AI pipeline to estimate the structural damage by comparing satellite imagery from before and after a disaster event.

Two contrasting approaches were implemented to solve this. First, a naïve baseline using basic computer vision techniques (HSV colour thresholding) to detect structures based on pixel brightness. Second, a modern AI pipeline utilising a fine-tuned RT-DTR model to detect structures based on semantic features. By comparing these two models, it can been seen that deep learning approaches significantly outperform more heuristic baselines when distinguishing structural damage from debris.

2 Task Definition

- **Task description:** The model takes a pair of satellite images (one before image and one after image) and estimates the number of intact roofs in each image. The change in roofs is then calculated, giving an estimate of the number of destroyed roofs.
- **Motivation:** Organisations providing aid or organisations planning rebuilding will need an estimation of destruction to allocate resources. Automating this process or even speeding it up can save crucial time in a crisis.
- **Input / Output:**
  - *Input:* Satellite images (jpg/png)
  - *Output:* Count of both intact roofs and difference in roofs (after - before)
- **Success criteria:** The model is considered "good" if it accurately reflects the reduction in structure count (damage trend) and avoids counting debris as intact structures.

3 Methods

3.1 Naïve Baseline

The baseline model relies on the heuristic that man made structures (roofs) are on average brighter and less colourful than the natural terrain. To do this, a colour thresholding algorithm using OpenCV is implemented.

- **Method description:** The image is first converted from BGR (Blue, Green, Red) to HSV (Hue, Saturation, Value). A binary mask is then applied to isolate pixels with low saturation and high brightness. This targets the "grey/ white" concrete or metal roofs. As the average size of an intact roof is approximately 300 pixels, the total amount of white pixels left on the screen after the mask is applied is then divided by 300 to get the estimated roof count.

- **Why naïve:** It lacks in semantic understanding. It treats any grey pixel as a ”house” including roads, driveways and even scattered debis
- **Likely failure modes:**
  - Concrete roads can be detected as a house
  - Scattered debris can be detected as a house
  - After a natural disaster, the quality of the air may be musty or a dirt cloud could have been left behind. This can cause the colours in the image it be off so the roofs may not even be detected.

3.2 AI Pipeline

A pre-trained transformer model was utilised, fine tuned specifically for remote sensing object detection

- **Model used:** `rt-detr-finetuned-for-satellite-image-roofs-detection` was selected for this project. The model is hosted on Hugging Face and is based on the RT-DTR architecture
- **Pipeline stages:**
  1. *Preprocessing:* Input images are converted to RGB and normalised to tensors using `AutoImageProcessor`
  2. *Inference:* The Transformer predicts bounding boxes and confidence scores
  3. *Post processing:* Detections are filtered using a confidence threshold of 0.3 to remove weak predictions
- **Design choices:** A Transformer based approach was chosen over a other traditional CNNs because of its ability to handle global context and varying scales of objects in satellite imagery.

4 Experiments

4.1 Datasets

To simulate realistic scanning in this scenario, a custom dataset was constructed, derived from satellite imagery

- **Dataset source**: Public databases, New York Times Website regarding a tsunami aftermath and Satellite Images Coorporation website
- **Size**: 72 distinct images (36 before images, 36 after images)
- **Splits**: No splits were performed as the model was pre trained for zero shot inference, the entire dataset acts as a test set
- **Preprocessing**: Large source images were algorithmically sliced into uniform tiles to simulate a scanning grid

4.2 Metrics

A ground truth labelling for every tile was ambiguous so a logic consistency metric was used

- **Negative damage rates:** A grim standard for a disaster is that the number of structures either decrease or stays the same. If the model predicts a negative number for the difference (increase in roofs), it becomes a logical failure caused by a misidentification of something.

4.3 Results

The AI pipeline significantly outperformed the Naïve baseline, particularly in the ”after” images where debris would be prevalent

Metric	**Naïve Baseline**	**AI Pipeline**
Total Pairs Tested	36	36
Logic Failures (Negative Damage)	15 (42%)	4 (11%)
Qualitative Performance	High Error	Robust

Table 1: Performance Comparison. The Naïve model failed in 42% of cases by predicting that the disaster ”built” new houses.

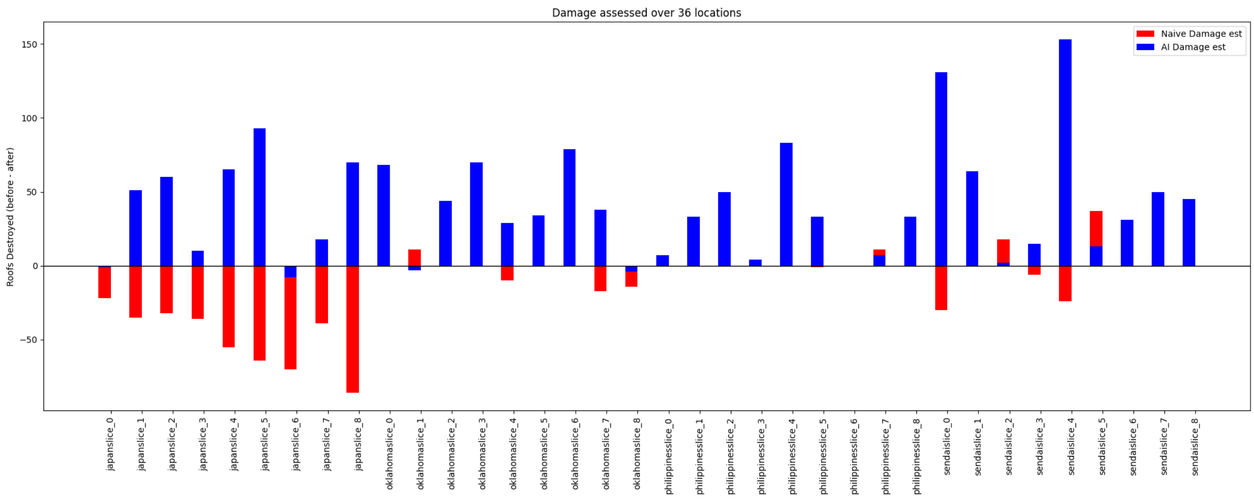

Figure 1: Damage Assessment Comparison. The Naïve model (red) consistently shows negative damage (bars going down), indicating it confused debris for houses. The AI model (blue) correctly identifies structural loss.

5 Reflection and Limitations

Reflection

This project successfully demonstrated that semantic understanding is required for this type of assessment, simple colour heuristics are insufficient. The AI pipeline worked better than expected, ignoring roads and driveways that confused the baseline. A key limitation of this project, was a lack of understanding the difference between an intact roof and one that is destroyed. Future work would involve training a multi-class model to distinguish between intact and partially destroyed.