

Background and Motivation

- Object detection and classification play an important role in smart-city applications as they can replace human surveillance monitoring in identifying critical situations
- However it is often hard to obtain a good training dataset when an object or scenario is rare and costly
- For example: “wounded humans in disastrous situations”
- In such scenarios, artificially generated images can serve as a potential solution
 - Leveraging the advancements of GANs and other data generation techniques, we can create realistic images of rare events, thereby augmenting training datasets
 - Consequently, generated images may help fill the data gaps, ensuring a more comprehensive and diverse dataset for training ML models



- State-of-the-art generative AI models can create high-quality synthetic images fast
- Carefully crafted generation prompts can serve as labels, reducing workload in labelling and data cleaning process
- The idea: Augmenting existing datasets with synthetic images to improve the accuracy on humans in abnormal postures, particularly in earthquakes
- Primary focus:
 - Replicating post disaster scenarios
 - Injured survivors in debris
 - Damaged buildings
 - Detecting critical objects:
 - Humans, Buildings
 - Damage Assessment
 - Number of injured people
 - Severity of injuries
 - Infrastructure Damage



*Sample image created using DALL-E 2

Quality of Existing Datasets

Dataset	Description	Relevant Image Count	
South Asia dataset (SAD)	Fire Disaster, Flood Disaster, Infrastructure, Nature Disaster, Human Damage, and Non Damage	240 out of 13720 images	<p>Crisisnlp Multimodal Damage Identification Dataset</p> <p>Crisisnlp Dataset</p>
Crisis Image Benchmarks Dataset	Affected, injured, or dead people, Infrastructure, and utility damage	1375 out of 5541 images	
Indonesian Disaster Victims 50 (IDV-50)	50 images from the internet and real disaster	50 images	
MEDIC dataset	Images sourced from various disaster scenarios, including earthquake, fire, flood, hurricane, and landslide. It consists of Crisis Dataset, aidr dataset and asonam dataset	25583 out of 72380 images	
Disaster Images Dataset	Images of natural disasters like Cyclone/Hurricane, Earthquake, Flood, Wildfire	1122 out of 4428 images	

Key Observations in Existing Datasets:

- Existing datasets generalize disaster types, lacking focus on earthquakes.
- There's a significant lack of images showing individuals amid earthquake debris.
- The abundance of non-relevant images reduces the datasets' overall usefulness.
- Many photos are taken too close or includes satellite imagery, missing broader context.
- Numerous pictures are of low quality.

An enhanced, focused dataset containing images of people in the debris of earthquakes is needed to improve AI-driven response in earthquake human detection

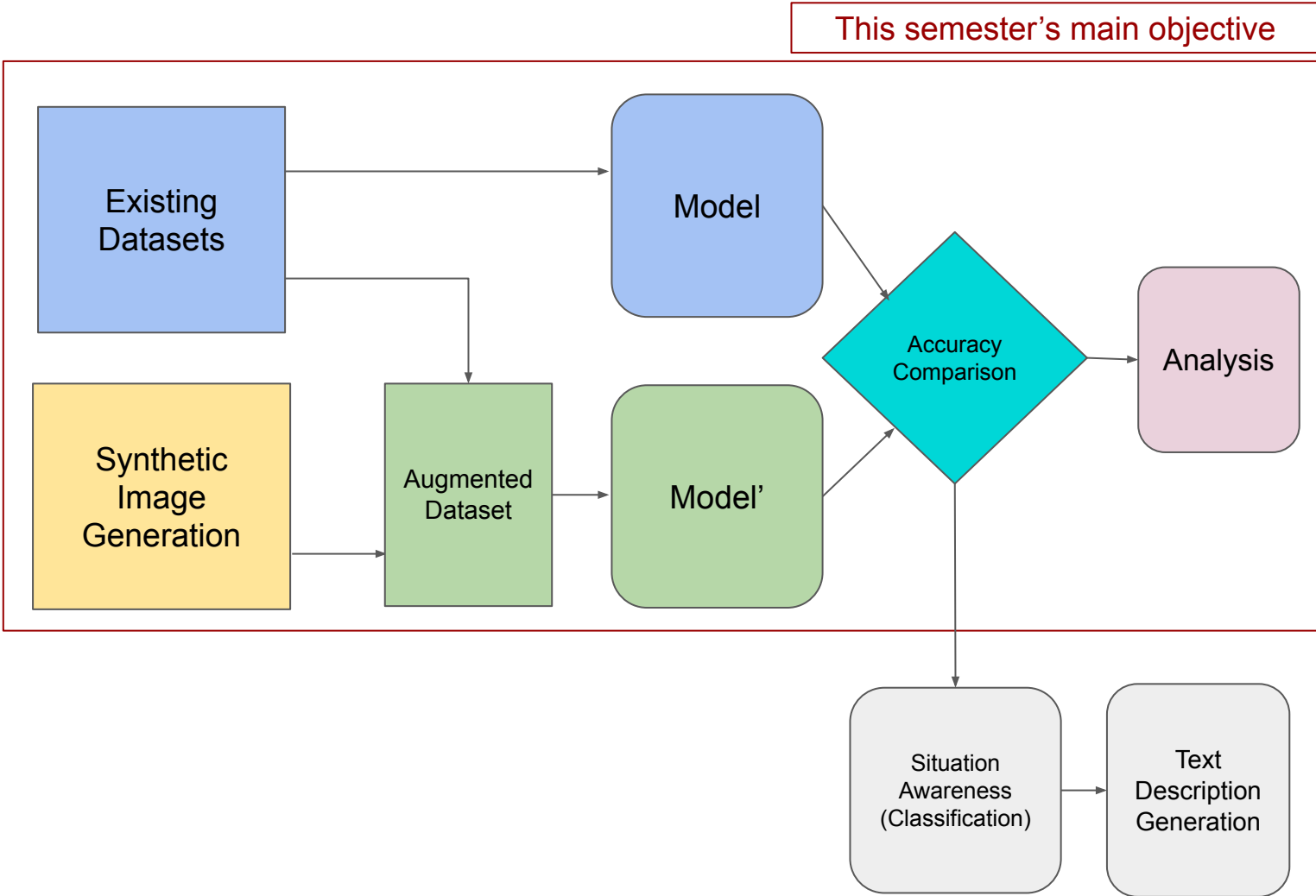
Method

Purpose	Models/Tools Used	Observations
Generate Synthetic Images	Dalle 2, Stable Diffusion	Dalle 2: <ul style="list-style-type: none">Generation time : 10 secondsIssues:<ul style="list-style-type: none">Does not generate images for injured people due to content Policy.
Image Annotation	VoTT, Labellmg, CVAT	VoTT: <ul style="list-style-type: none">manual graphical annotation tool developed by microsoftannotation data exported to JSON includes:<ul style="list-style-type: none">labelsheightcoordinates, etc
Object Detection	Yolov8, FCOS, TTFNet, PAFNet	YOLO : <ul style="list-style-type: none">Pretrained model :<ul style="list-style-type: none">Pretrained on COCO dataset.Issues:<ul style="list-style-type: none">Difficult to detect survivors who are laying flat on the ground or obscured by dust.Struggles to detect individuals from aerial or drone perspectives.Solution: Finetune Model with the current dataset.
Text Generation	CLIP-Interrogator-2, pix2struct, Mini GPT-4	CLIP-interrogator-2 <ul style="list-style-type: none">pretrained on ViT-H-14 OpenCLIP model to use to create similar image in stable diffusion model 2.0text generation time:<ul style="list-style-type: none">quality = good: 44.56 squality = poor: 17.94 sissues:<ul style="list-style-type: none">depends on quality or size of image (poorer quality → faster output times)example output:<ul style="list-style-type: none">a pile of debris sitting on top of a beach next to a body of water, wood pier and houses, wall, beaches, fallen columns

Example Prompts:

- “A photorealistic, high-resolution image of a city street after an earthquake with clear and detailed humans.”
- "With the clarity and depth seen in National Geographic photos taken with a Nikon, showcase a city street in the aftermath of an earthquake, where every human figure lying amidst the debris is as distinct and detailed as a portrait."

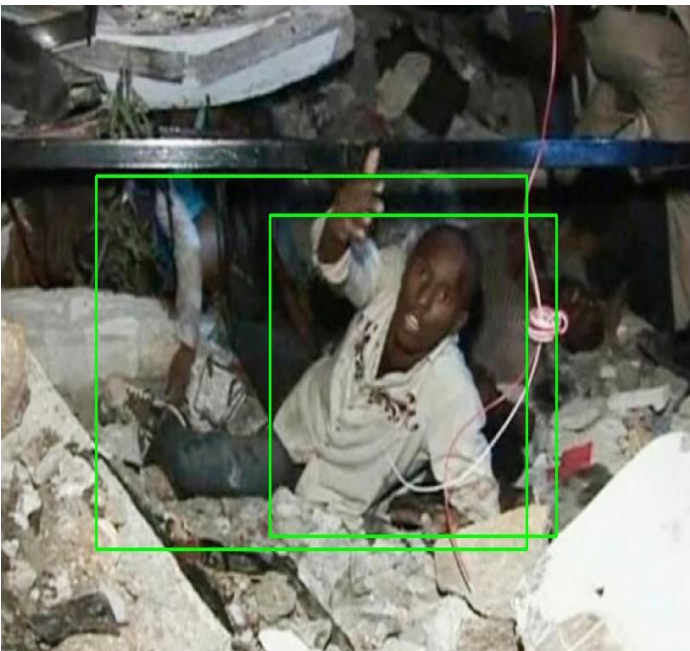
Workflow



- Model:**
 - Identifies critical objects and poses.
 - Situation awareness:**
 - Distinguishes the scenarios based on the severity
 - Text Generation:**
 - Creates text summary description of the image
 - Could generate the keywords for the scenario.
- Datasets:**
 - Existing Dataset: Preprocessed and filtered images with post earthquake scenarios
 - Synthetic Image Generation: Using Generative AI to replicate post earthquake images using prompts.

Key Observations

- Dalle 2:**
 - Dalle 2 restricts users to create some images of injured people due to the content policy.
- Pre-Trained Yolov8:**
 - Difficult to identify humans in the rubbles of the earthquake
 - Pose estimation becomes difficult when the person is not identified



Person identified in the debris



Person misclassified as bird



People not identified by Yolov8

Next Steps

- Annotate** existing dataset for human detection.
- Finetune** Yolov8 with existing Real image datasets to improve human detection in post earthquake scenarios. This will work as a benchmark for comparison.
- Merge** the existing dataset with the images generated by AI and train Yolov8 and **compare** the model performance with the benchmark based on evaluation metrics.
- Once we confirm that training dataset augmented with synthetic images improves performance in object detection, the pose estimation, text generation on parallel.
- The detections and summary would be processed to analyze the severity of the situation.