2024 EY Open Science Data Challenge

Team: Double Y (Yi Jie Wong, Yin Loon Khor, Liu Ziwei)

A: Key elements:

1. Target Region (i.e., Puerto Rico)
2. Object Detection Model (i.e., YOLOv8n)
3. Microsoft Building Footprint (BF) dataset (we extract only the Puerto Rico region for usage)
4. Puerto Rico dataset (17040 unique data, **no manual annotation** required)
5. Expert dataset (28 unique data, 84 after augmentation)
6. Crowdsourced dataset (200 unique data, 600 after augmentation)

B: Assumptions:

1. Observation: When labelling multiple versions of the provided post-disaster dataset, we observed that not all annotated data aligns with the expected outcomes in the EY validation images. Some of our annotated datasets yield high mAP, some yield low mAP.
2. Consequently, we assume datasets that perform exceptionally well as the "expert dataset," annotated by "expert annotators." Conversely, the datasets that do not yield results as good as the expert dataset are referred to as the "crowd-sourced dataset".
3. We assume expert annotators could effectively differentiate damaged/undamaged commercial and residential buildings. Logically, the expert dataset is a quality > quantity dataset.
4. Meanwhile, crowd-sourced dataset is a quantity > quality dataset. We assume this is labelled by volunteer annotators in real life scenario, rather than the experts.
5. We assume all buildings in Microsoft BF dataset are undamaged residential buildings (since majority buildings are residential). The exact class is not important, since the dataset is only for pretraining.

C: Methodology:

Figure 1 visually summarizes the methodology in our study, showcasing the flow from data collection to Modelling, highlighting the pivotal role of YOLOv8 into actionable insights for building detection.

1. Pretraining

- Firstly, we identify the target region, which in this case is Puerto Rico.
- We extract the satellite images & bounding boxes of all buildings within the Puerto Rico region, from the Microsoft BF dataset. The extracted dataset is hereinafter referred to as "Puerto Rico" dataset.
- We pretrain YOLOv8n model on Puerto Rico dataset, enabling it to familiarize itself with the unique characteristics of the Puerto Rico landscape, laying the groundwork for subsequent phases.

2. Fine-tuning

- We fine-tune the pretrained model on the crowd-sourced dataset, allowing it to adapt quickly to identify various building types.
- Next, we fine-tune the model again on the expert dataset, aiming to align the actual model predictions as closely as possible with expert annotations.
- The two-step process allows us to leverage both the high-quantity crowd-sourced dataset and the high-quality expert dataset.

3. Model Retraining

- We operationalize the model through MLOps strategies. We use the fine-tuned model to self-annotate the unlabelled datasets. Minimal human intervention is required, mainly focusing on refining or adjusting pseudo-labels as needed.
- The self-annotated dataset is partitioned into multiple folds, each containing a subset of images for training and evaluation. Models trained on each fold are then assessed using expert-annotated datasets, with folds exhibiting satisfactory results being retained for further model retraining.
- Ultimately, this iterative process culminates in the development of a robust and efficient building detection model tailored to the residential and commercial structures in Puerto Rico.

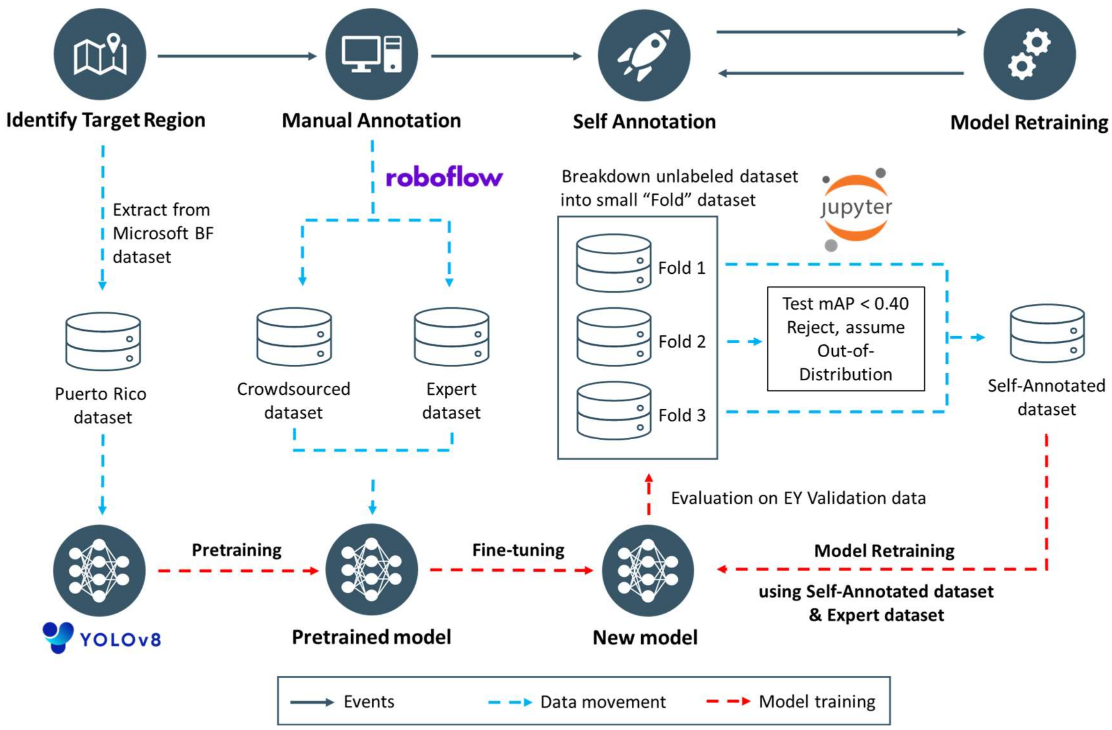

Figure 1: Advanced Methodology for Building Detection Utilizing YOLOv8 Algorithm

D: Approach (How we arrive at the proposed methodology)

Without benefit of hindsight, it is not straightforward to design the methodology. Here, we present the ablation study (Table 1) we conducted to demonstrate how we arrived at the proposed methodology in Section C.

Table 1: Comparative Evaluation of Model Performance Across Methodologies in Building Detection

Setup	Pretraining	Fine-tuning		MLOps	mAP
		Crowd-sourced	Expert		
A	√				0.10
B	√	√			0.44
C	√		√		0.39
D	√	√	√		0.50
E		√	√		0.24
F	√	√	√	√	**0.51**

- **Setup A**: We pretrained a YOLOv8n model using the Puerto Rico dataset. Surprisingly, we achieved a mAP of 0.10 on the EY validation dataset, without any manual annotation from our side!
- **Setup B**: When fine-tuning the pretrained model on the crowd-sourced dataset, we achieved an mAP of 0.44, which exceeds the completion threshold for this challenge (mAP 0.40).
- **Setup C**: When fine-tuning the pretrained model directly on the Expert dataset, we can achieve an mAP of 0.39, despite the dataset containing only 28 unique data (84 after augmentation). This shows that the quality of data is equally important, if not more important than the quantity of data.
- **Setup D**: We initially fine-tune the pretrained model using a large-scale crowd-sourced dataset to quickly warm it up. Subsequently, we fine-tune the model on the expert dataset, which has more accurate labels. With this approach, we achieved a mAP of 0.50.
- **Setup E**: We demonstrate that without pretraining, the performance is not satisfying even when both the crowd-sourced and expert datasets are utilised, only achieving mAP 0.24.
- **Setup F**: Finally, we demonstrate that by employing the proposed MLOps cycle, we can enhance the model's mAP to 0.51. Notably, the sole human intervention in this MLOps cycle involves verifying the self-labelled data using the baseline model from Setup E.

GitHub link: https://github.com/yjwong1999/EY-challenge-2024