SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

الهيئة الملكية لمدينة الرياض
ROYAL COMMISSION FOR RIYADH CITY

أمانة منطقة الرياض
RIYADH REGION MUNICIPALITY

وزارة الشؤون البلدية والقروية والإسكان
Ministry of Municipal Rural Affairs & Housing

Smartathon

The Smart Cities Challenge

THEME 1

- Meshal Alamr
- Abdulmajeed Almuarik
- Saud Alrasheed
- Wael Mobeirek
- Hussain Alkhayat
- Kumail Alhamoud

Table of Contents

1. Overview ... 3
2. Challenges ... 3
   2.1 Potential Approaches ... 4
3. Scalability ... 5
4. Software Used ... 5
5. Future Work ... 5
   5.1 Proposed Method ... 5
      5.1.1 Schematic Representation ... 6

1. Overview

This report will detail our efforts towards solving theme 1 of the Smartathon, which was to detect and evaluate visual pollution on street imagery taken from a moving vehicle. We developed a deep learning model based on YOLOv5's object detection architecture to localize and classify visual pollutions, such as *graffiti, faded signage, potholes, garbage, construction road, broken signage, bad streetlight, bad billboard, sand on road, clutter sidewalk, and unkept facades*.

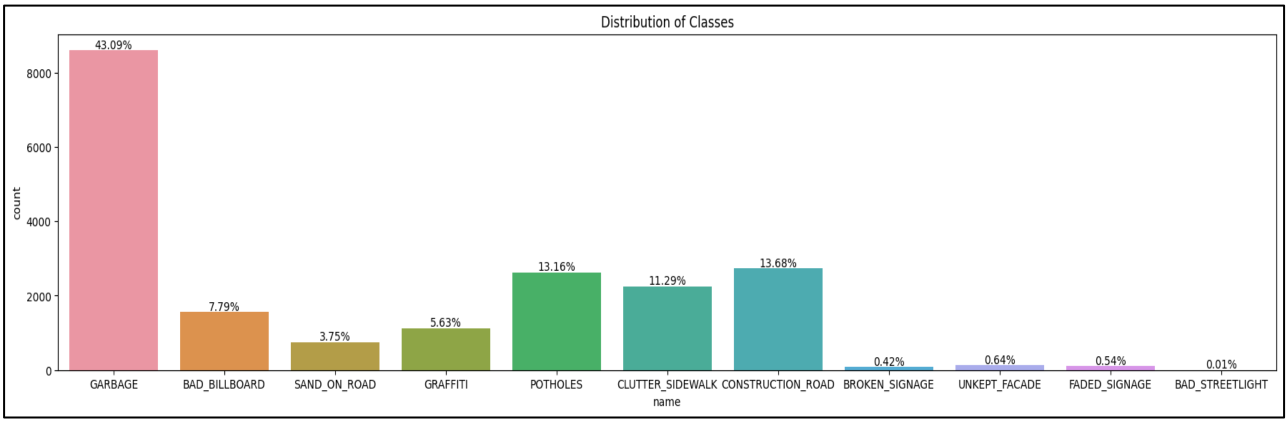
Additionally, we propose a new method to tackle this challenge, which is further discussed in Section 5.

2. Challenges

Observing the training data reveals numerous challenges such as:

- Class imbalance
- Inconsistent labels
- Mislabeled objects
- Inaccurate bounding boxes

These are major challenges for building a reliable object detection model, the class imbalance can lead to poor performance on under-represented classes. As shown below, some of the classes represent less than 1% of the data. In addition, incorrect labels can cause the model to learn the wrong associations between images and labels, further degrading the performance.



2.1 Potential Approaches

The aforementioned challenges stem from the fact that the data was labelled by human annotators which means that it is prone to error, and in this case, it is poorly labelled. Therefore, to handle this and to ease the task of labelling, here are a few approaches to consider:

Data Augmentation: Data augmentation techniques such as rotation, flipping, and random cropping are useful to create more examples of under-represented classes and reduce class imbalance.

Re-sampling: Oversampling the under-represented classes or downsampling the over-represented classes to balance the distribution of the classes.

Ensemble Methods: Combining multiple models trained on different subsets of the data, or different architectures, to improve overall performance.

Re-labeling: Manually re-label the images, ensuring that the bounding boxes are correct, and the labels match the objects in the images.

Semi-supervised Learning: Use the labeled data to train a model, and then use the model to label a large amount of unlabeled data, making it easier for the labelers.

Attention Mechanisms: Attention mechanisms can help focus the model's attention on relevant parts of the image, which can improve performance on object detection tasks.

Confident Learning: A technique that uses the model's confidence in its predictions to identify uncertain or incorrect labels in the dataset. It works by training the model on a small subset of the labeled data, and then using the model's predictions to identify uncertain or incorrect labels in the remaining data. These uncertain or incorrect labels can then be corrected by human annotators before being used to further train the model, improving its performance.

Crowdsourcing: Recruiting many labelers to label the data, and then aggregating their labels to get a more accurate label. This can help to reduce the noise and uncertainty in the labels.

Quality Control: Quality control measures can be used to check the quality of the labels provided by the labelers. This can be done by having a second labeler to check the labels provided by the first labeler.

3. Scalability

In Section 2.1, we went over some potential approaches to mitigate the issues that stem from the labelling. These approaches will also work when considering scalability. For example, the Confident Learning approach, would greatly ease the labelling effort and it will further improve the performance of the model and this can go back and forth until the labelling is almost not needed.

In addition, we can use models that can detect noisy labels within the data to further mitigate the labeling effort by using automated tools such as Cleanlab.

4. Software Used

The following open-source software was used to develop and study the model.

- Yolov5
- Weights and Biases

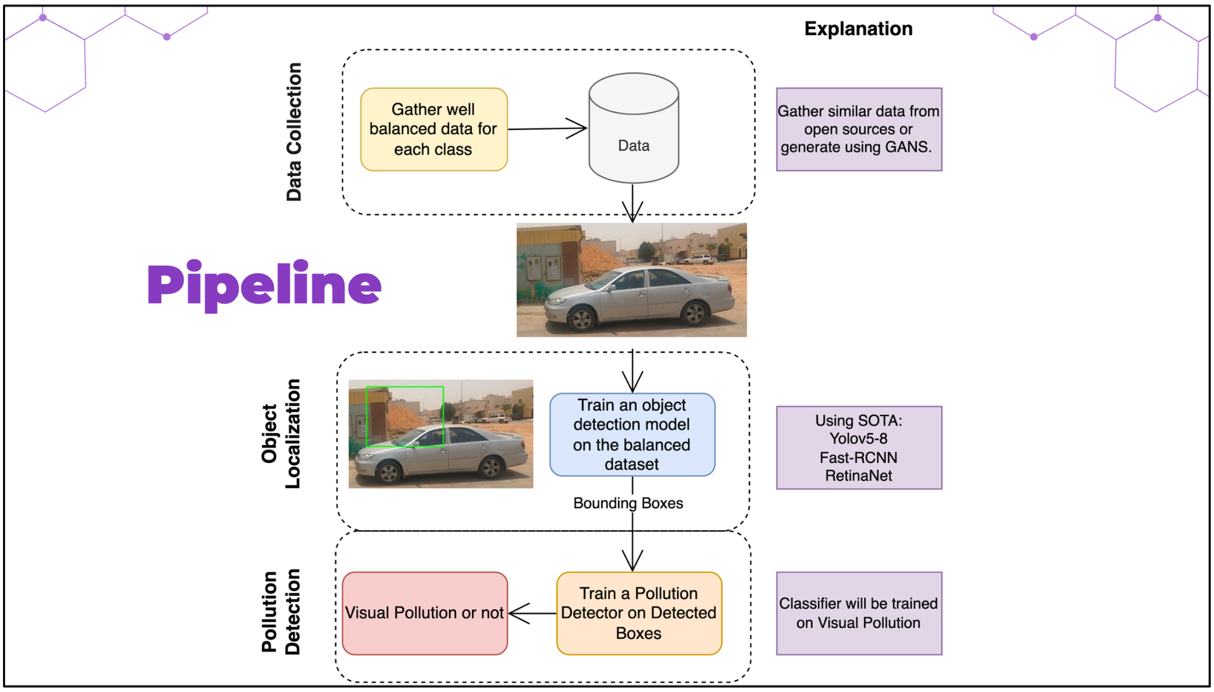
5. Future Work

In this section, we'll explain a new proposed method that we would try if we had more time and resources.

5.1 Proposed Method

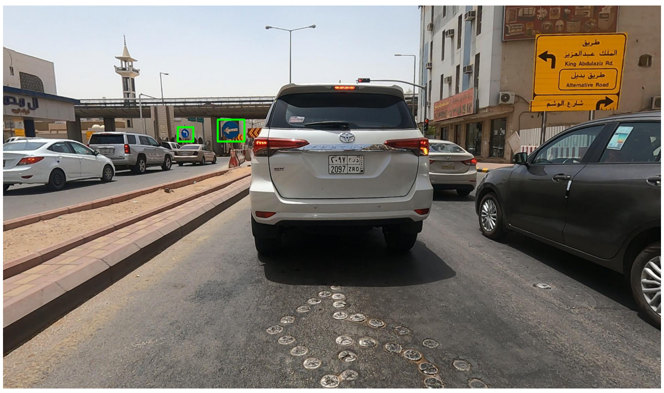
- We aim to split the task to into two stages object detection and classification which we think can result in higher accuracy compared to doing both using the same model.
- SOTA object detection models such as YOLOv8, Fast R-CNN, and RetinalNet **will be trained on detecting the object regardless whether the object is a visual pollution or not**. For example, the model will return the bounding boxes of any traffic SIGN detected in the image without determining if it is faded or not.
- A **dedicated classifier** will be trained to identify if the given object is **a visual pollution or not.**
- This approach can help reduce false positives and false negatives, resulting in improved accuracy.

5.1.1 Schematic Representation

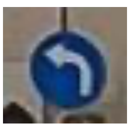


Raw Image Input:

Output of the object detection model (Stage 1)



Classification on Visual Pollution (Stage 2)



Classifier output:
NOT FADED

Classifier output:
FADED