Data Science and AI - T5 Bootcamp

PolluVision

Abstract

Visual pollution poses a significant challenge to the aesthetics and sustainability of urban environments. To combat this issue, we propose the development of a smart AI-based product capable of detecting, classifying, and minimizing visual pollution.

This project aims to build a model of automated visual pollution detection and classification, utilizing the technological prowess of the twenty-first century for environmental management applications and AI.

Introduction

The term quality of life (QoL) refers to a wide range of multifaceted concepts that often involve subjective assessments of both positive and negative aspects of life. Since visual pollution is a relatively recent issue compared to other forms of environmental contamination, study is needed to define, formalize, measure, and evaluate it from many angles.

To address this, a smart AI-based product is needed to detect and classify visual pollution, Provide Coordinates of the Location and minimizing human errors.

This initiative aligns with Saudi Arabia's Vision 2030, aiming to enhance the global status of Saudi cities and promote their evolution into smart urban centers. By utilizing artificial intelligence, visual pollution can be identified, contributing to sustainable urban growth. Overall, this endeavor aims to improve urban environments and establish Saudi Arabia as a leader in AI.

Data Description and Structure

This section of the report aims to provide a comprehensive overview of the Urban Visual Pollution Dataset utilized in the project. The dataset was sourced from Kaggle and focuses specifically on visual pollution data collected from roads in the Kingdom of Saudi Arabia. The dataset encompasses various forms of visual pollution, including GRAFFITI, POTHOLES, GARBAGE, CONSTRUCTION_ROAD, BROKEN_SIGNAGE, BAD_BILLBOARD, SAND_ON_ROAD, CLUTTER_SIDEWALK, and CONSTRUCTION_WASTE.

One of the primary challenges encountered with the dataset was the issue of imbalanced classes. The original dataset consisted of 5,196 images, which exhibited an imbalance in the distribution of the different visual pollution categories. To address this, data augmentation techniques were applied, involving rotation within a range of -25° to +25° and saturation adjustments between -40% and +40%. These augmentation techniques effectively increased the dataset to a total of 12,000 images.

Another challenge encountered during the preprocessing stage was the presence of incorrect bounding box annotations. To rectify this, manual annotation and labeling of the images were performed using Roboflow. However, it is important to note that this process proved to be time-consuming due to the dataset's size and the inclusion of multiple labels and annotations for certain images.

The data structure of the Urban Visual Pollution Dataset primarily consists of image files in formats such as JPEG or PNG. These image files are organized within directories, categorizing them based on the different forms of visual pollution. Considering the dataset's size, it falls into the category of large image datasets, which can pose computational challenges.

To overcome this, we leveraged cloud resources, specifically utilizing Colab with powerful GPU, RAM, and additional computing units, to efficiently handle the computational demands of the dataset.

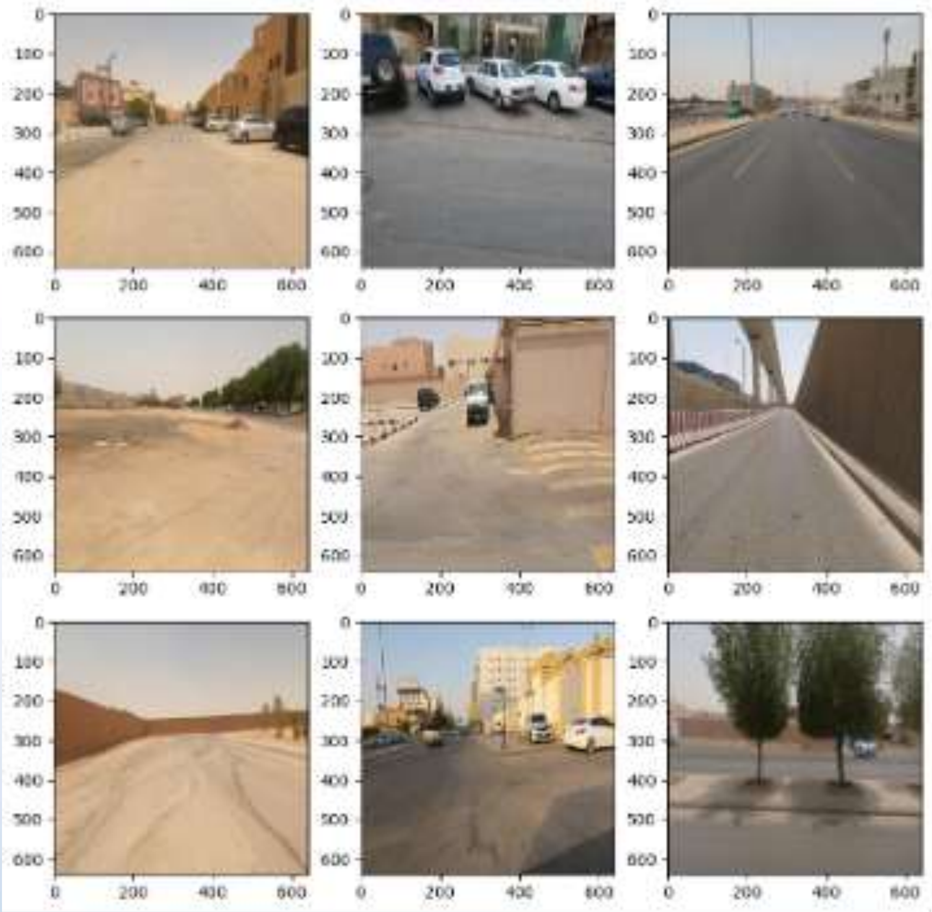

Figure 1 Sample of datasets

Methodology

CNN

For the task of detecting visual pollution in Riyadh, Saudi Arabia, a Convolutional Neural Network (CNN) model is employed to analyze images and identify instances of environmental clutter. The data collection process involves assembling a dataset containing diverse examples of visual pollution, with subsequent annotation of bounding boxes to establish ground truth information. Prior to training, the images are pre-processed by resizing them to a standardized format suitable for the CNN model's input requirements. The dataset is then meticulously split into training and validation sets, ensuring a balanced representation of visual pollution instances in both subsets.

The CNN model, tailored specifically for object detection, is initialized with pre-trained weights from a relevant source, providing a robust starting point for learning hierarchical features. Adjustments are made by incorporating appropriate convolutional layers and fine-tuning the fully connected layers to adapt the model to the intricacies of visual pollution detection. Training involves exposing the model to the annotated training images, enabling it to learn to identify and localize visual pollution objects within the images.

Post-training, the model's performance is evaluated using the validation set, allowing for an assessment of its object detection capabilities. Subsequent iterations of fine-tuning and evaluation may be performed to refine the model further. To apply the trained CNN model on unseen test images depicting visual pollution in Riyadh, the model's predictions are utilized.

These predictions facilitate the identification and classification of visual pollution objects by delineating bounding boxes around them, aiding in the comprehensive assessment and monitoring of environmental clutter in the region.

Methodology

VGG16

The VGG16 model is utilized for object detection on images depicting visual pollution in Riyadh, Saudi Arabia.The collected images are pre-processed by resizing them to a uniform size suitable for the model's requirements. The dataset is divided into training and validation sets, ensuring a balanced distribution of visual pollution instances.

The VGG16 model is initialized using pre-trained weights from ImageNet and modified for object detection by adding suitable fully connected layers. During training, the model is fed with the training images and their annotated bounding boxes. After training, the model is evaluated using the validation set to assess its object detection performance. Further iterations of fine-tuning and evaluation may be conducted to improve the results.

To classify the objects in unseen test images depicting visual pollution in Riyadh, the trained model is applied. The model's predictions are used to identify and classify visual pollution objects by drawing bounding boxes around them.

In the post-processing stage, the results are analyzed to gain insights into the prevalence and distribution of visual pollution in Riyadh. While the model achieves successful classification, occasional misplacements of bounding boxes are carefully examined and addressed to enhance accuracy and reliability.

Overall, this approach leverages the VGG16 model to detect and analyze visual pollution in Riyadh, contributing to understanding and addressing the issue.

Methodology

YOLOv8

After gather our comprehensive dataset of images depicting various forms of visual pollution, such as graffiti, cluttered sidewalks and more. Ensure the dataset represents the diverse range of visual pollution encountered in urban environments. Also, we annotate the collected images using a suitable annotation tool, such as roboflow and providing bounding box coordinates. We divide the annotated dataset into training, validation, and testing sets. We experimented with YOLOv5 and YOLOv7 but found that YOLOv8m produced the best results. Number of Epochs: 100 epochs were used for training.

We use image size of 640 during training. A confidence threshold of 0.45 was chosen. This threshold determines the minimum confidence score required for a detection to be considered valid. Intersection Over Union (IoU): An IoU threshold of 0.4 was used. The training process show that the model's training and validation losses and metrics generally improve as the number of epochs increases. However, the validation mAP50 (mean average precision at 50% intersection over union) metric is at 0.45.

Discussion and Results

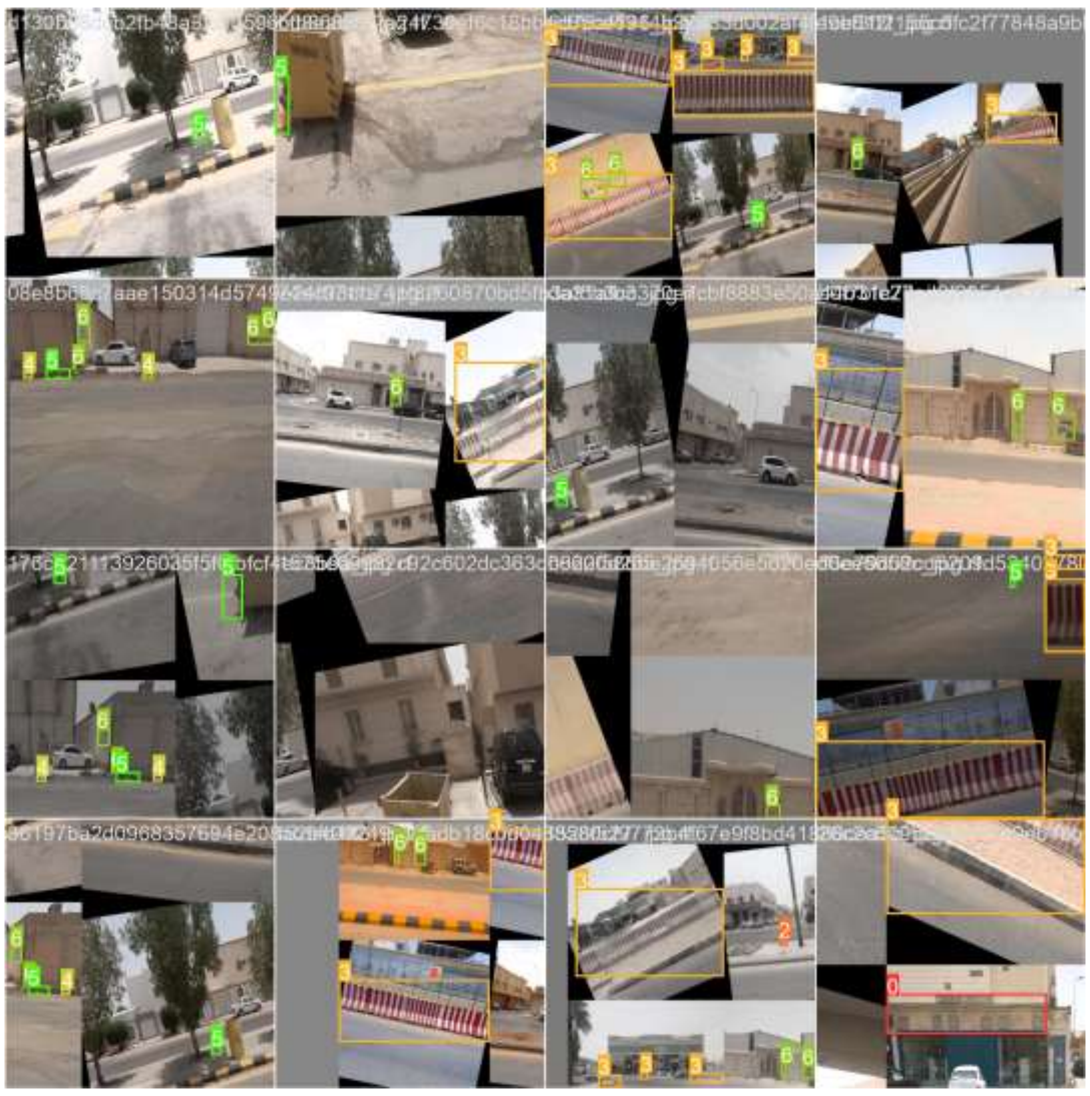

Figure 2 Sample after training

The graphs show the prediction with data augmentation result. The model's training and validation losses and metrics generally improve as the number of epochs increases. However, the validation mAP50 (mean average precision at 50% intersection over union) metric is still relatively low, at 0.45.

Conclusion and Future Work

In conclusion, PolluVision successfully detects and classifies visual pollution in the environment. By utilizing the YOLO8 transfer learning model for object detection, we have developed an AI-based solution capable of accurately detecting and classifying various forms of visual pollution, including graffiti, potholes, garbage, and broken signage.

The significance of this project lies in its potential to contribute to the aesthetics and sustainability of urban areas, aligning with Saudi Arabia's Vision 2030. The practical implications are vast, as the developed product can provide coordinates of the location of the visual pollution to assist the authorities in addressing visual pollution and minimizing human errors.

The results obtained from this project provide valuable insights into the prevalence and distribution of visual pollution, enabling the development of effective strategies for its reduction and mitigation. The future work for this project involves enhancing the model to work in real-time and incorporating image captioning capabilities. These extensions will contribute to the ongoing efforts to reduce visual pollution and promote sustainable urban growth.

Team

Abeer Alharbi

Arjwan Alqarni

Batool Alghamdi

Nouf Aldawsari

Rawabi Algethami

Project Repositorie