Team 3

ODLC implementation for aerial images

Introduction

The goal of this project is to design and implement an end-to-end pipeline for detecting, localizing, and classifying objects in aerial images.
The pipeline aims to solve the Object Detection and Localization Challenge (ODLC) by automatically identifying standard objects that consist of a colored alphanumeric character placed on a colored shape.

Standard object specifications:

- **Shapes:** Circle, semicircle, quarter circle, triangle, rectangle, pentagon, star, cross
- **Colors:** White, black, red, blue, green, purple, brown, orange
- **Alphanumeric Characters:** Uppercase English letters and numbers

The implemented system will include:

1. Data gathering and preprocessing
2. Object detection model
3. Object localization
4. Alphanumeric and color classification

Data Gathering

The datasets used in this project were collected from the following sources:

- Roboflow-suas-shape-detection-2(for shape detection)
- Kaggle-color-classification-11-categories (for color classification)
- Kaggle-AlphaNum (for alphanumeric classification)

Object detection model

For the object detection stage, the chosen model was YOLOv12 was trained on the prepared dataset to detect and identify the target shapes in aerial images. The dataset was divided into training, validation, and testing sets to ensure fair evaluation and to prevent overfitting.

During training, the model learned to identify different shapes and objects within the images by adjusting its parameters to minimize detection errors. Unfortunately, due to some code errors and model requiring to be trained over and over (which takes at least 17 hours) to adjust to changes, the process could not be fully completed as planned. While the model did make some progress, the final results were limited, and further work is needed to fully train and optimize the detection pipeline.

Object Localization

Once an object was detected, the next step was to determine its real-world position by calculating its x and y coordinates in meters relative to the center of the image. This process involved extracting the object's pixel coordinates and converting them into real-world distances using camera specifications and flight altitude. The SIYI A8 Mini camera, with a focal length of 21 mm, was assumed to be mounted facing straight downward at a 90-degree angle, and the images were taken at an altitude of 80 meters. By treating the image center as the origin, the detected object's position could be accurately mapped to a coordinate system. Additionally, as a bonus step, these local coordinates were converted into latitude and longitude values using a given set of home coordinates, allowing the results to be expressed in a format compatible with mapping and navigation systems.

Alphanumeric and Colour classification

In addition to detecting and localizing objects, a classification system was developed to identify the alphanumeric characters placed on the detected shapes. This involved training a separate model to recognize letters and numbers, enabling the system to fully interpret the targets. For color classification, a simple method was implemented to determine the color of each shape using its image data, ensuring accurate differentiation between the defined color classes. However, similar to the object detection stage, errors occurred during the training process, which prevented the models from fully completing their training. As a result, the final performance of these components was limited and requires further refinement to achieve reliable results.

Conclusion

This project aimed to design and implement a complete ODLC pipeline capable of detecting, localizing, and classifying objects in aerial images. While progress was made in preparing the dataset and setting up the models for detection, localization, and classification, technical errors during the training process prevented the models from being fully trained and optimized. As a result, the final pipeline was only partially functional and could not achieve its full potential.

Despite these challenges, the work completed provides a strong foundation for future development. With further troubleshooting, and additional time for training, the system can be enhanced to accurately detect and classify objects, as well as reliably provide real-world coordinates. This project highlighted the complexities of building an end-to-end computer vision solution and provided valuable learning experiences that will guide future improvements.