Project 1: Object detection techniques (in case of small objects) on AU Drone dataset.

Kushalkumar Suthar
Btech in CSE
Ahmedabad University
kushalkumar.s@ahduni.edu.in

Dhruvin Prajapati
Btech in CSE
Ahmedabad University
dhruvin.p@ahduni.edu.in

Rohit Rathi
Btech in CSE
Ahmedabad University
rohit.r1@ahduni.edu.in

Nish Parikh
Btech in CSE
Ahmedabad University
nish.p@ahduni.edu.in

***Abstract*—QueryDet, a novel query mechanism designed to accelerate the inference speed of feature-pyramid based object detectors. It first predicts the coarse locations of small objects on low-resolution features and then computes the accurate detection results using high-resolution features sparsely guided by those coarse positions. This method not only utilizes the benefits of high-resolution feature maps but also avoids unnecessary computation for the background area. On the VisDrone dataset, which contains more small objects, QueryDet sets a new state-of-the-art while achieving a 2.3× high-resolution acceleration on average.**

***Index Terms*—QueryDet, Small Object Detection, High-Resolution Features, Feature-Pyramid Based Object Detectors, Inference Speed, VisDrone Dataset.**

I. Introduction

THE paper discusses the challenges in detecting small objects despite the recent advances in deep learning. It highlights the performance gap between small and normal scale objects, with small objects often being more difficult to detect due to factors such as down-sampling operations in the backbone of Convolutional Neural Networks (CNN), mismatched receptive field on low-resolution features, and significant disturbance in the Intersection over Union (IoU) metric due to small perturbations of the bounding box. The paper proposes QueryDet as a solution to these challenges, offering a simple and effective method to save the detection head's computation while promoting the performance of small objects. QueryDet is evaluated on the COCO detection benchmark and the VisDrone dataset, showing significant acceleration in inference while improving the detection performance.

II. Methodology

QueryDet is a novel method aimed at improving the accuracy and speed of detecting small objects, with a specific focus on enhancing RetinaNet-based detectors. RetinaNet forms the foundation of this approach, utilizing a backbone network coupled with Feature Pyramid Network (FPN) to extract features at multiple scales. Within RetinaNet, there are two main components: classification and regression heads, responsible for predicting object categories and bounding box coordinates, respectively.

One notable observation addressed by QueryDet is the uneven distribution of computational costs across different layers of the FPN, particularly impacting the detection of small objects. To mitigate this issue, QueryDet introduces a sparse query approach. This involves predicting rough object locations on coarse feature maps and then focusing computational efforts on refining these predictions on finer feature maps. A query head is integrated into the model to predict these coarse object locations alongside the classification and regression heads.

During inference, locations with predicted scores above a set threshold are designated as queries. These queries are then mapped to key positions on the preceding layer's feature map. Sparse convolution techniques are employed to process only these selected positions, significantly reducing computational costs. Additionally, a cascade sparse query (CSQ) strategy is implemented, wherein queries for each layer are derived from key positions on the previous layer, thus avoiding exponential increases in computational complexity.

Training of the model remains consistent with RetinaNet, utilizing FocalLoss for the classification and regression heads. The query head is trained using FocalLoss as well, with binary target maps generated based on distances between feature positions and small object centers. To ensure balanced learning across all layers, the loss of each layer is appropriately weighted, particularly crucial with the addition of higher-resolution features.

In comparison to related approaches such as RPN-based detectors and PointRend, QueryDet distinguishes itself through its focus on classification in coarse prediction, sparse and selective computation, and distinct methods for query generation. Overall, QueryDet represents a significant advancement in improving the efficiency and accuracy of small object detection within RetinaNet-based detectors, offering a streamlined and effective approach to optimizing computational resources.

III. Dataset

Image 1

Image 1 is the example image from the Visdrone dataset.

ID	xc	yc	w	h	Score	Category ID	Trun.
3	777	33	53	1	10	0	0
96	783	33	37	1	2	0	0
150	827	46	57	1	1	0	1
186	833	37	60	1	1	0	2
263	697	18	56	1	1	0	1

Table 1: Given dataset

Table 1 contains 5 object details of the image 1. The Visdrone dataset contains images with annotation details shown in Table 1. In Table 1, the ID column contains object ID, xc and yc are x and y coordinates of the boundary box around that object, and w and h are the width and the height of that boundary box, and trun. refers to truncation.

IV. Results

Image 2

Image 2 is the output image after applying the QueryDet approach on the image from Visdrone dataset.

V. References

https://openaccess.thecvf.com/content/CVPR2022/papers/Yang$_{Q}ueryDe$ $Resolution_{S}mall_{O}bject_{D}etection_{C}VPR_{2}022_{p}aper.pdf$

Notebooks documentation. Kaggle. (n.d.-a). https://www.kaggle.com/docs/notebooks