Density Map Guided Object Detection in Aerial Images: Implementation and Results

Changlin Li, Taojiannan Yang, Sijie Zhu, Chen Chen, Shanyue Guan

1 Introduction

This project addresses object detection challenges in high-resolution aerial images by using density map-guided image cropping to improve detection accuracy.

2 Dataset Description

2.1 Dataset Overview

I used the VisionDrone dataset as suggested and used by the paper. It has 6471 Train images, 548 Validation images, and 1610 Test images. The annotations were of the following format.

Annotation Format: bounding_box_coordinate_x, bounding_box_coordinate_y, width, height, ignored_region_flag, class, truncation, occlusion

3 Implementation Details

3.1 Model Architecture

Two models were implemented in the paper, however I could not get the 2nd model for object detection to run(faster rnn with resnet50), hence i trained YOLOv5 on the dataset.

Multi-Column Convolutional Neural Network This model was used to predict/ generate density map. I utilized pre-trained weights that were given in the github repository of the paper.

YOLOv5 Object Detector After getting the cropped out images from the MCNN, YOLOv5 is used for object detection. I trained this model for 100 epochs, with learining rate 0.0001, batch size = 16 and image size = 640. Training was done on kaggle to use the GPU (T4*2 accelerator).

3.2 Parameter Configuration

Parameter Configuration	Details
Density Map Generation	No additional parameters required
MCNN Density Map Prediction	No additional parameters required
Image Processing	**Cropping Parameters:** Threshold: 0.08 (VisDrone specific) Minimum bounding box size: 70×70 pixels
YOLOv5 Training	**Optimizer:** SGD Learning rate: 0.01 Momentum: 0.9 **Parameter Groups:** 69 weights (decay=0.0) 76 weights (decay=0.0005) 75 bias parameters (decay=0.0)
Detection Fusion	**Non-max Suppression:** Threshold: 0.7 Maximum bounding boxes after fusion: 500

Table 1: Parameter Configuration Table
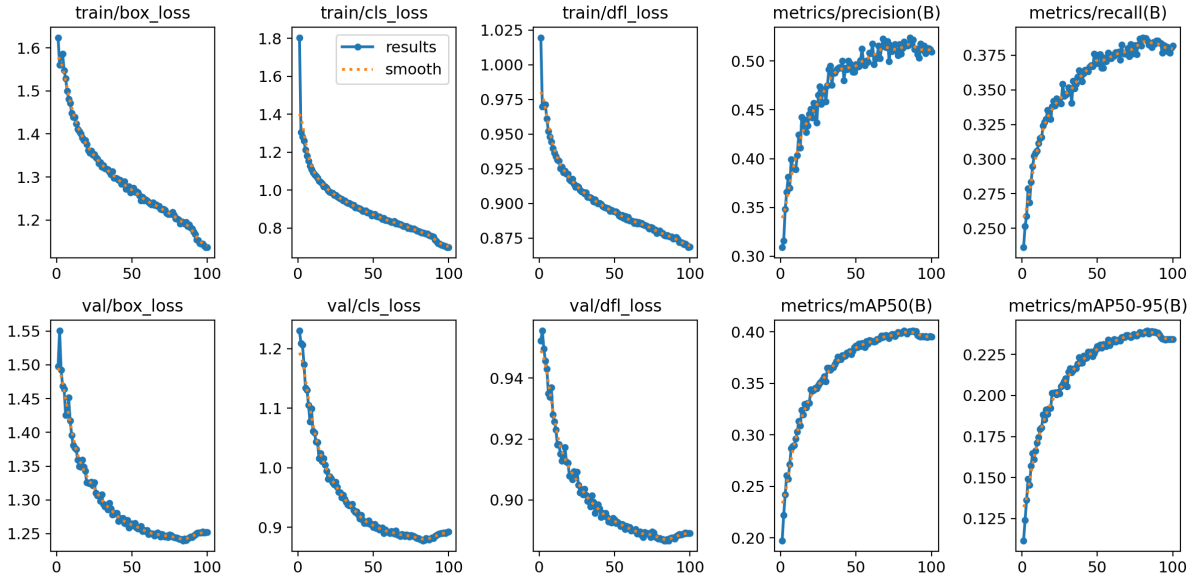
3.3 Workflow
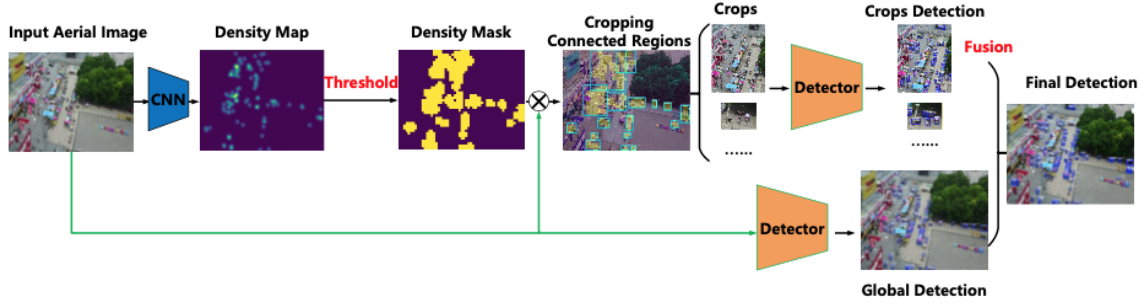
4 Results

4.1 Training Results

Following are the training results from YOLOv5:

4.2 Qualitative Results

Metric	YOLOv5 Model	Best from Paper
AP (Average Precision)	0.500	0.294
$\mathbf{AP_{50}}$	0.504	0.532
$\mathbf{AP_{75}}$	0.498	0.306
$\mathbf{AP}_{small}$	0.608	0.216
$\mathbf{AP}_{medium}$	0.493	0.412
$\mathbf{AP}_{large}$	0.473	0.571

Table 2: Comparison of Reproduced Results with the Best from the Paper