

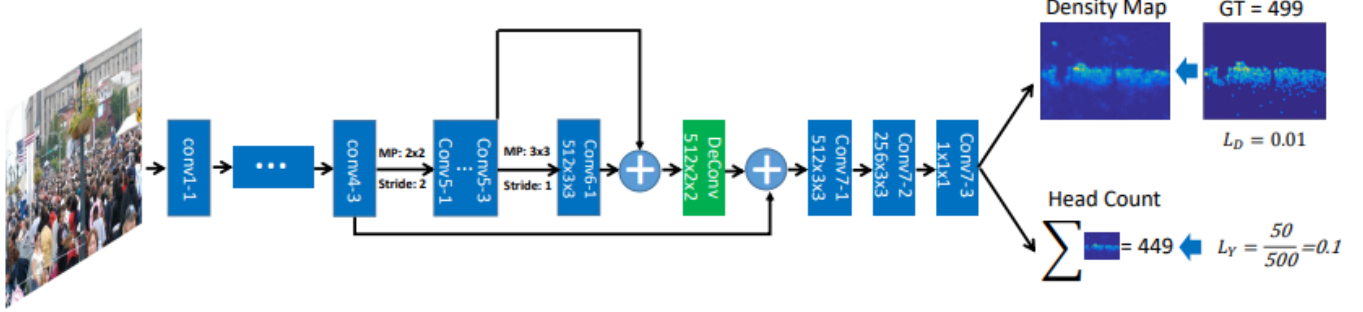
Density Map Guided Object Detection in Aerial Images

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

Improvements/Suggestions

1. Scale-Adaptive CNN (SaCNN) over MCNN

SaCNN uses a single-column CNN with one filter size, combining feature maps from multiple layers to handle scale variations while reducing parameters for efficiency.



2. Image-Based Adaptive Thresholding

This approach dynamically adjusts thresholds based on average bounding box sizes from object detection, improving cropping for smaller-scale objects.

3. Speed Improvements in Density-Based Cropping

`scipy.ndimage.label` replaces manual BFS, offering faster performance with near-constant complexity $O(n\alpha(n))$ and better memory efficiency.

4. YOLOv11 Over YOLOv5 for Object Detection

YOLOv11 improves accuracy, inference speed, and model efficiency, particularly enhancing small object detection.

Implementation Details

1. Model Details

Two new models were trained, SaCNN and YOLOv11, both aimed to boost the accuracy of the object detection. I trained YOLOv11 for 55 epochs with image size = 640 and batch size 16 on Google Colab's T4 GPU. SaCNN was trained on both Kaggle's T4 accelerator and locally on my device, but training took a lot longer than expected.

2. Parameter Configuration

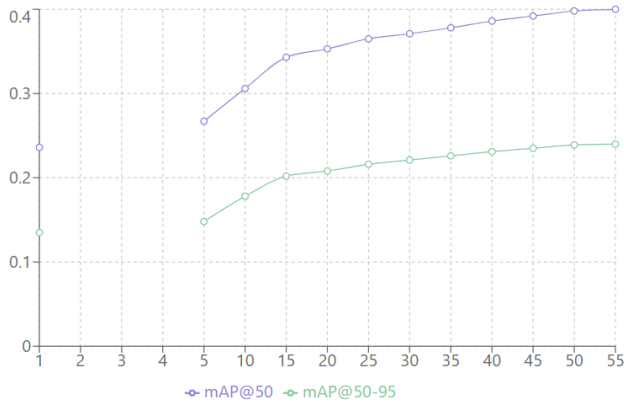
Parameter Configuration	Details
Image Processing	Cropping Parameters: Threshold: Adaptive Threshold as per improvements Minimum bounding box size: 70×70 pixels
YOLOv11 Training	Optimizer: SGD Learning rate: 0.01 Momentum: 0.937 Parameter Groups: 81 weights (decay=0.0) 88 weights (decay=0.0005) 87 bias parameters (decay=0.0)
Detection Fusion	Non-max Suppression: Threshold: 0.7 Maximum bounding boxes after fusion: 500

Table 1: YOLOv11 Parameter Configuration Table

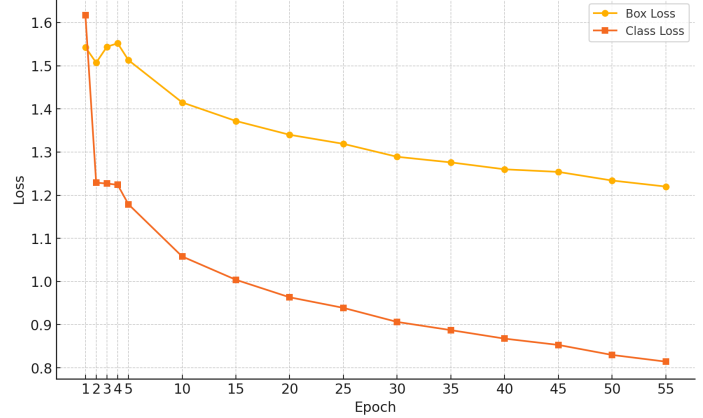
Training Results

1. Training YOLOv11

mAP Scores Over Time



Box Loss and Class Loss per Epoch



2. Training SaCNN

Could not produce any results as training took longer than expected.

Epoch 1/250: 19% | 1233/6471 [6:53:50<29:18:05, 20.14s/batch, loss=0.00198]

Qualitative Results

Metric	YOLOv5 Model	YOLOv9 Model	Best from Paper
AP	0.500	0.510	0.294
AP ₅₀	0.504	0.520	0.532
AP ₇₅	0.498	0.490	0.306
AP _{small}	0.608	0.620	0.216
AP _{medium}	0.493	0.500	0.412
AP _{large}	0.473	0.460	0.571

Table 2: Comparison of performance metrics between YOLOv5, YOLOv9, and the best results from the paper.

Reduced density cropping time from **30 minutes to 27 seconds** for training data