

Project Weekly Report - 3

Topic: - Evaluate the Performance of Faster-RCNN and its variants in the case of a small object detection

Group Name: Tech Trio

Project Definition: 1

Group Member's names: Kaushik Gohil, Richa Saraiya, Parth Mevada

M2F2-RCNN (Multi-functional Faster R-CNN with Multi-scale Feature Fusion):

This week, we searched for the M2F2-RCNN, a Multifunctional Faster RCNN with multi-scale feature fusion. FPN is conceptually related to multi-scale feature fusion. It uses the multi-scale feature fusion method for region searching in remote sensing images. The major idea for the M2F2-RCNN is to integrate the FPN model to improve multi-scale object detection. This model is designed to improve problems like poor object detection and poor feature representation. This model contains methods like feature pyramid networks and ResNet backbone networks, which are used for feature extraction. Also, methodologies of the M2F2-RCNN model: Usually faster RCNN contains RPN, feature extraction, Region of interest pooling, etc. At the feature extraction, ResNet converts the input image into the feature maps.

Also, RPN creates region proposals and ROI pooling is for extracting features from proposals and classification and regression of bounding boxes. Also, in the multi-scale feature extraction, connecting the feature maps from different layers and feature extraction for various resolutions. Also, ROI alignment is used for improving feature extraction and reducing errors in ROI pooling. The feature fusion at multi-scale also increases the accuracy of the model. Also, the summing up of feature maps and Loss function with smooth L1 is used for classification and bounding box regression. Overall, this model increases the small object detection process at various scales of the objects.

The research presents M2F2-RCNN, an improved Faster R-CNN model for detecting objects in satellite and aerial images with higher accuracy. Traditional models struggle with small objects, overlapping targets, and complex backgrounds, leading to errors. To solve this, M2F2-RCNN enhances feature extraction with ResNet50 and multi-scale fusion, improves region proposal using RoIAlign, and refines object classification with Soft-NMS, ensuring better detection of overlapping objects. The model was trained on Google Earth and DOTA datasets, which contain high-resolution images of objects like vehicles and planes.

Results show M2F2-RCNN outperforms existing models, achieving 65.5% accuracy for small vehicles, 86.5% for large vehicles, and 92.7% for planes. RoIAlign improved accuracy by 3%,

and Soft-NMS boosted recall. Although computationally demanding, future work will focus on real-time performance, better small object detection, and expanding datasets. This model significantly improves remote sensing image analysis, making it useful for military, disaster management, and urban planning.

Feature Pyramid Network:

Feature Pyramid Network (FPN) is a deep learning architecture designed to improve object detection, particularly for small objects. It enhances feature extraction by constructing a multi-scale feature pyramid using a top-down and lateral connection approach. High-resolution, low-semantic features are fused with low-resolution, high-semantic features, enabling better object detection across different scales. FPN is widely integrated into models like Faster R-CNN, RetinaNet, and Mask R-CNN.

The core idea of FPN is to generate a feature hierarchy that captures both fine and coarse details. It uses a bottom-up pathway to extract features, a top-down pathway to upsample, and lateral connections to merge information. This hierarchical structure ensures robust feature representation, leading to improved detection accuracy. FPN efficiently balances computational cost while enhancing performance. It is particularly beneficial in detecting small and occluded objects in complex scenes.