Evaluating the Performance of Faster-RCNN and Its Variants for Small Object Detection

Kaushi Gohil, Parth Mevada, Richa Saraiya
AU2444022, AU2240172, AU2444002

***Abstract*—Small object detection is challenging due to low resolution, weak feature representation, and poor localization accuracy. While Faster R-CNN performs well for general object detection, it struggles with small objects due to inadequate region proposals and feature loss. This project evaluates Faster R-CNN and its advanced variants—FPN for multi-scale feature extraction, M2F2-RCNN for feature fusion, and CFINet for refined proposal generation—to enhance small object detection. The models are compared based on accuracy, recall, and efficiency, aiming to identify the best approach. Future improvements include adaptive anchor selection, lightweight architectures, and transformer-based enhancements for better small object detection.**

I. Introduction

- Smallobject detection remains a critical challenge in computer vision, especially in applications like aerial imagery, autonomous driving, and medical imaging. Traditional object detection models, such as Faster R-CNN, struggle with detecting small objects due to their low resolution, weak feature representation, and inadequate overlap with anchor boxes.
- Traditional object detection models primarily focus on extracting high-level features, often discarding fine-grained details necessary for identifying small objects. As a result, small objects are frequently misclassified, poorly localized, or entirely missed by standard detection frameworks.
- Several advanced techniques have been introduced to enhance small object detection performance: Feature Pyramid Networks (FPN) improve feature extraction by introducing a multi-scale, top-down feature pyramid. M2F2-RCNN enhances detection by fusing deep and shallow feature maps while utilizing attention mechanisms. CFINet refines object proposals step by step using coarse-to-fine proposal generation and feature imitation learning. This project evaluates the performance of Faster R-CNN and its advanced variants—FPN, M2F2-RCNN, and CFINet—by analyzing their accuracy, recall, and computational efficiency. The study also explores future improvements such as: Adaptive anchor selection for better proposal matching. Lightweight architectures for real-time inference. Transformer-based enhancements to optimize feature learning.

II. Literature Survey

The following table summarizes key small object detection methods, highlighting their approaches, contributions, and limitations.

TABLE I
Summary of Literature on Small Object Detection

Paper	Method	Key Contributions	Limitations
FPN (Lin et al., 2017)	Faster-RCNN+FPN	Improves small object detection using a feature pyramid.	Struggles with small objects, fixed anchors.
M2F2-RCNN (Yin et al.)	Multi-scale Feature Fusion	Combines deep and shallow features with CBAM attention.	High computational cost, complex training.
CFINet (Yuan et al., 2023)	Coarse-to-Fine Proposal	Uses refined proposals and feature imitation learning.	Complex training, high memory usage.

III. Dataset discussion

- The VisDrone 2019/2020 dataset is a large-scale UAV-based dataset designed for object detection, tracking, and behavior analysis in both urban and rural environments. Captured using drones under varying conditions such as different lighting, weather, and camera angles, it provides high-resolution aerial imagery for real-world applications like traffic monitoring, pedestrian tracking, and autonomous navigation. The dataset is structured into a Sequences Folder, containing video frames of tracking sequences, and an Annotations Folder, which provides bounding box annotations for each detected object, including tracking IDs, occlusion levels, and truncation indicators. Additionally, metadata files offer extra details like timestamps, drone flight parameters, and environmental conditions, aiding in more context-aware object detection and tracking.
- Objects in the dataset are categorized into five size-based classes: Small, Medium-Small, Medium, Medium-Large, and Large, ensuring robust detection across various scales. It is widely used for vehicle and pedestrian tracking, UAV-based traffic analysis, occlusion handling, and trajectory reconstruction, making it a benchmark dataset for deep learning and computer vision research. The dataset helps in improving multi-object tracking (MOT)

models, autonomous navigation, and AI-powered drone surveillance by addressing challenges such as small object detection, motion blur, and occlusion handling. With its structured annotations and real-world applicability, VisDrone continues to advance AI-driven object tracking and recognition technologies.

IV. Methodology

- The proposed methodology enhances small object detection by integrating techniques from Feature Pyramid Networks (FPN), M2F2-RCNN, and CFINet into the Faster R-CNN framework. The process follows a structured approach, consisting of the following key stages:

1) **Feature Extraction using FPN**
   - A top-down feature pyramid is used to generate feature maps at multiple scales.
   - Helps capture finer details of small objects by utilizing feature representations from different layers.
   - Ensures better spatial resolution for small object detection.
2) **Multi-Scale Feature Fusion (M2F2-RCNN)**
   - Deep and shallow feature fusion is performed to enhance small object localization.
   - CBAM (Convolutional Block Attention Module) is used to focus on the most relevant object regions.
   - Reduces the loss of crucial information that typically occurs in deep convolutional layers.
3) **Coarse-to-Fine Proposal Generation (CFINet)**
   - A coarse-to-fine region proposal approach refines object proposals step by step.
   - Feature Imitation Learning helps enhance small object representations by mimicking larger objects.
   - Improves recall and precision for small objects in cluttered environments.
4) **Detection and Refinement**
   - The Region Proposal Network (RPN) generates candidate object regions.
   - If the detection confidence is low, refinement is performed by re-evaluating features and adjusting proposal boxes.
   - Iterative refinement ensures better accuracy in detecting small objects with minimal misclassification.

- This methodology overcomes Faster R-CNN's limitations by combining multi-scale feature extraction, improved feature fusion, and stepwise proposal refinement to achieve higher detection accuracy for small objects in complex scenarios.

V. Future scope

- Future improvements in small object detection can benefit applications like surveillance, navigation, and medical imaging. Research can focus on adaptive anchor boxes for better small object alignment and lightweight models like MobileNet for real-time use.
- Self-supervised learning can enhance recognition using unlabeled data, while Vision Transformers (ViTs) improve feature extraction. GAN-based data augmentation can create synthetic datasets, making models more robust and accurate.

VI. References

1) Lin et al. (2017) introduced Feature Pyramid Networks (FPN) to improve small object detection using a top-down feature pyramid, enhancing multi-scale feature representation.
2) Yin et al. (2022) proposed M2F2-RCNN, which combines deep and shallow feature fusion with CBAM attention to improve small object detection but requires high computational power.
3) Yuan et al. (2023) developed CFINet, a coarse-to-fine inference network that refines object proposals and uses feature imitation learning to improve detection accuracy, though it demands high memory and complex training.