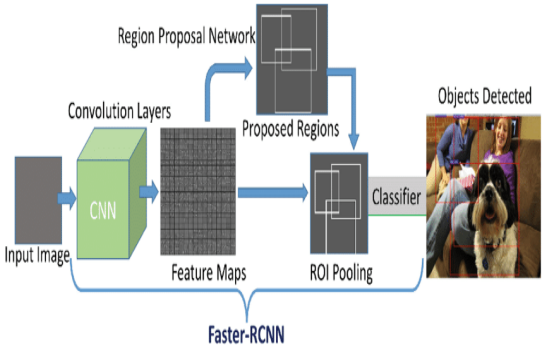
Evaluating the Performance of Faster-RCNN and its Variants for Small Object Detection

Course name:- Computer Vision

Group Members	Enrollment Number
Kaushik Gohil	AU2444022
Parth Mevada	AU2240172
Richa Saraiya	AU2444002

Problem Statement

- Faster R-CNN struggles with small objects as its RPN fails to generate good proposals. Low resolution, fewer features, and weak anchor box overlap make recognition difficult. Deep CNN layers focus on high-level features, losing fine details. As a result, small objects are often missed or misclassified.
- FPN, M2F2-RCNN, and CFINet improve small object detection using multi-scale features, better fusion, and refined proposals.



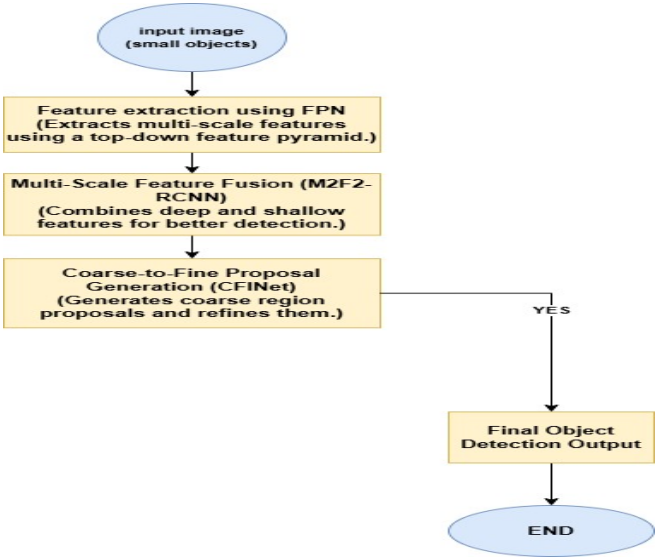
Literature survey

Ahmedabad University

Paper	Method	Key Contributions	Limitations
FPN (Lin et al., 2017)	Faster-RCNN + FPN	Multi-scale feature extraction enhances small object detection and accuracy across different object sizes.	Struggles with very small objects and depends on fixed anchor sizes.
M2F2-RCNN (Yin et al.)	Multi-scale Feature Fusion	Uses deep and shallow feature fusion with attention mechanisms to improve small object detection.	High computational cost and requires extensive fine-tuning.
CFINet (Yuan et al., 2023)	Coarse-to-Fine RPN	Refines small object proposals using a stepwise approach and feature imitation learning to improve accuracy.	Complex training and high memory usage.

Flowchart Representation

Ahmedabad University



Dataset Discussion

Ahmedabad University

VisDrone 2019/2020 Dataset

The VisDrone-VID datasetis a large-scale UAV-based dataset designed for object detection, tracking, and behavior analysis. The dataset is structured into:-

Sequences Folder, containing video frames of tracking sequences.

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.The data strcuture is:-

- **Trainset**
- **Valset**
- **Testset-Dev**
- **Testset-Challenge**

Dataset discussion [cont..]

Figure: Bounding Box annotations in Frames

Bounding Box and Object Categories:-

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. Bounding boxes are used to define object positions, making it ideal for deep learning-based object detection and tracking.

Approaches and Future work

Ahmedabad University

- The approach integrates techniques from Feature Pyramid Networks (FPN), M2F2-RCNN, and CFINet to enhance small object detection in Faster R-CNN.

1. **Multi-Scale Feature Extraction (FPN)**
   - Uses a top-down feature pyramid to improve feature representation.
   - Enhances detection for objects of varying sizes.
2. **Multi-Scale Feature Fusion (M2F2-RCNN)**
   - Combines deep and shallow feature maps to capture small object details.
   - Uses CBAM (Convolutional Block Attention Module) for better focus.
3. **Coarse-to-Fine Proposal Generation (CFINet)**
   - Generates coarse region proposals first, then refines them step-by-step.
   - Uses Feature Imitation Learning to enhance small object detection.

Approaches and Future Work [cont..]

Ahmedabad University

1. **Adaptive Anchor Boxes**
   - Traditional anchor boxes may not align well with small objects.
   - Adaptive selection improves proposal matching and localization.
2. **Lightweight Models for Real-Time Use**
   - Existing models are computationally expensive.
   - Efficient backbones (e.g., MobileNet, EfficientNet) enhance speed.

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.