Session 3: Object Detection Techniques for Earth Observation

From Faster R-CNN to YOLO/DETR: Counting and Localizing Objects

Stylianos Kotsopoulos
EU-Philippines CoPhil Programme

Session Overview

Duration: 1.5 hours
Type: Theory + Discussion
Goal: Understand modern object detection for EO

You will learn: - Detection vs classification vs segmentation - Two-stage vs single-stage vs transformer-based detectors - Anchor boxes, NMS, IoU, mAP - EO applications: ships, vehicles, buildings, oil tanks, aircraft - Architecture selection for Philippine use cases

Prerequisites: - Day 2 CNNs

Resources: - Reference PDF: "Day 3, Session 3_ Object Detection for Earth Observation Imagery.pdf"

What is Object Detection?

Definition & Output

- Predicts class + bounding box for each object instance
- Outputs: [x_min, y_min, x_max, y_max], label, confidence

Task	Output	Use
Classification	Single label	Scene type
Object Detection	Boxes + labels	Counts & locations
Segmentation	Pixel mask	Precise shapes/areas

When to use detection vs segmentation

- Detection: count and localize distinct instances (ships, vehicles, aircraft)
- Segmentation: delineate precise boundaries (floods, land parcels)

Architecture Families

Two-Stage (Faster R-CNN)

- RPN proposes ROIs; second stage classifies + regresses boxes
- Pros: high accuracy, good for small objects
- Cons: slower inference

Single-Stage (YOLO/SSD/RetinaNet)

- Direct dense predictions in a single pass
- Pros: fast (real-time), simple deployment
- Cons: historically lower accuracy (modern versions close gap)

Transformer-Based (DETR)

- Set prediction without anchors; bipartite matching
- Pros: elegant, end-to-end
- Cons: needs data, longer to converge

Key Concepts

Anchor Boxes

- Predefined sizes/ratios; model predicts offsets
- Anchor-free: YOLOv8/CenterNet predict centers + sizes directly

Non-Maximum Suppression (NMS)

- Remove duplicate overlapping detections
- Threshold trade-off: 0.3 strict vs 0.7 lenient (typical 0.4–0.5)

Intersection over Union (IoU)

- Overlap between predicted and ground-truth boxes
- Used in training, NMS, and evaluation

mAP (mean Average Precision)

- AP per class from precision-recall curve; mean over classes
- Variants: mAP@0.5, mAP@0.75, mAP@[0.5:0.95]

EO Applications

Maritime: Ship Detection

- Data: Sentinel-1 SAR / Sentinel-2 optical
- Challenges: small targets, waves → FPs
- Models: YOLOv8 (monitoring), Faster R-CNN (inventory)

Urban: Vehicles & Buildings

- Vehicles: very small, dense → Faster R-CNN + FPN
- Buildings: YOLOv5/8 or Faster R-CNN depending on speed vs accuracy

Infrastructure: Oil Tanks & Aircraft

- Distinct shapes and contexts; moderate object sizes

EO-Specific Challenges

Small Objects

- Solutions: FPNs, higher resolution inputs, tiling, multi-scale training

Scale & Orientation Variation

- Solutions: multi-scale anchors, rotation augmentation, oriented boxes

Complex Backgrounds

- Solutions: context windows, stronger backbones, hard-negative mining

Limited Labels

- Solutions: transfer learning, SSL pretraining, active learning, synthetic data

Atmospheric/Quality Issues

- Solutions: preprocessing, SAR fusion, multi-temporal inputs, domain adaptation

Architecture Selection Guide

Application	Recommended	Rationale
Ship detection (EEZ)	YOLOv8 / Faster R-CNN	Speed vs accuracy trade-off
Vehicle counting	Faster R-CNN + FPN	Very small, dense
Building detection	YOLOv5/8	Balance for urban monitoring
Aircraft monitoring	YOLO series	Medium targets, fast inference
Oil tank inventory	Faster R-CNN	Precision matters
Disaster assessment	YOLOv8	Large-area rapid scans

Summary & Prep for Lab

Key Takeaways

- Detection = classification + localization via boxes
- Families: two-stage, single-stage, transformer-based
- Concepts: anchors, NMS, IoU, mAP
- EO challenges + mitigation strategies

Next (Session 4)

- Hands-on: fine-tune a pre-trained detector for Metro Manila buildings
- Evaluate with mAP; visualize detections