Object Detection - Evaluation layer

Core metrics for Object Detection (images)

COCO-style Average Precision (AP / mAP)

- **AP@[.5:.95] (primary):** mean AP over IoU thresholds 0.50:0.95 (step 0.05). *Use when you want a single, strict quality number balancing classification + localization.*
- **AP50 / AP75:** AP at IoU=0.50 (lenient) and IoU=0.75 (strict). *AP50 = sensitivity to detection presence; AP75 ≈ localization precision.*
- **APS / APM / APL:** AP on Small/Medium/Large objects. *Reveals scale-specific strengths/weaknesses (e.g., tiny object recall).*

Average Recall (AR)

- **AR@1/10/100:** max recall with at most N detections per image; also **ARS/ARM/ARL**. *Use to diagnose missed detections independent of precision.*

Precision–Recall (PR) curves (per-class)

- *Use when you need threshold selection and class-wise behavior.*

Latency/Throughput

- **ms/image, FPS, memory** *Operational KPI; evaluate at the same input size used in production.*

When to use which

- **Model selection/reporting:** AP@[.5:.95] + AP50/AP75 + APS/APM/APL.
- **Debugging misses:** AR@100 + per-class PR curves (find low-recall classes).
- **Deployment tuning:** Choose confidence/NMS thresholds from PR; report latency/FPS.
- **Open-vocabulary/grounded:** Standard AP on held-out “novel” classes; phrase grounding uses IoU ≥ 0.5 between predicted box for a phrase and the GT region.
- **Video detection / tracking:** Evaluate per-frame with AP, and if tracking IDs are produced, add **MOTA/MOTP/IDF1** (not shown here; image OD focus).

Visualization Methods

Bounding boxes & diagnostics

- Draw predicted vs GT boxes, color by TP/FP/FN to spot failure modes (occlusion, small objects, duplicates).

Grad-CAM / CAM (for CNN backbones)

- Heatmap indicating spatial importance for a prediction (helps explain false positives/negatives).

Attention maps (Transformers / DETR)

- Visualize **cross-attention** of decoder queries over image features to understand object localization behavior.
