

Deep Learning Evaluation Metrics

Classification Metrics

1. Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

2. Precision & Recall

- Precision** (Accuracy of positive predictions):

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall** (Sensitivity, True Positive Rate):

$$\text{Recall} = \frac{TP}{TP + FN}$$

3. F1-Score

Harmonic mean of precision and recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. ROC Curve & AUC

- ROC Curve:** Plots TPR (Recall) vs FPR at various thresholds

$$\text{FPR} = \frac{FP}{FP + TN}$$

- AUC:** Area Under Curve (0.5 = random, 1.0 = perfect)

Object Detection Metrics

1. IoU (Intersection over Union)

Measures overlap between predicted and ground-truth boxes:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

2. mAP (mean Average Precision)

- For each class:

$$AP = \int_0^1 p(r)dr$$

(Area under precision-recall curve)

- **mAP**: Average AP over all classes (COCO standard: mAP@[0.5:0.95])

YOLO (You Only Look Once)

Definition

A real-time object detection system that predicts bounding boxes and class probabilities **in a single pass** through a CNN.

Key Innovations

1. **Unified Architecture**: Combines detection and classification in one network.
 2. **Grid-based Prediction**: Divides image into $S \times S$ grid cells (e.g., 7×7).
 3. **Anchor Boxes**: Predefined box shapes for different object aspect ratios.
-

Mathematical Formulation (YOLOv1)

For each grid cell:

$$\text{Output} = [p_c, b_x, b_y, b_w, b_h, C_1, C_2, \dots, C_k]$$

- (p_c): Object confidence
 - (b_x, b_y): Box center relative to grid cell
 - (b_w, b_h): Box width/height relative to image
 - (C_k): Class probabilities
-

Loss Function (Simplified)

$$\lambda_{\text{coord}} \sum \text{Localization Loss} + \lambda_{\text{obj}} \sum \text{Confidence Loss} + \sum \text{Class Loss}$$

Evolution

Version	Key Features	Speed (FPS)	mAP (COCO)
v1	Single CNN, grid-based	45	63.4
v3	Darknet-53 backbone, multi-scale prediction	30	55.3
v5	PyTorch implementation, auto-learning anchors	140	50.7

Version	Key Features	Speed (FPS)	mAP (COCO)
v8	Anchor-free, SOTA performance	85	53.9

Why YOLO Matters

- 1. **Speed:** Real-time detection (>30 FPS).
- 2. **Accuracy:** Competitive mAP on COCO/PASCAL VOC.
- 3. **Versatility:** Supports detection, segmentation, pose estimation.

Example Workflow

- 1. Resize input to 416×416
- 2. Pass through CNN backbone (e.g., Darknet)
- 3. Predict boxes at 3 scales
- 4. Apply Non-Max Suppression (NMS) to remove duplicates

Non-Max Suppression (NMS)

- 1. Discard boxes with confidence < threshold (e.g., 0.5)
- 2. Select box with highest confidence
- 3. Remove boxes with IoU > threshold (e.g., 0.45)
- 4. Repeat until no boxes remain