Faculty of Engineering
University of Ruhuna

YOLO

UNIFIED, REAL-TIME OBJECT DETECTION

Conference Paper By :

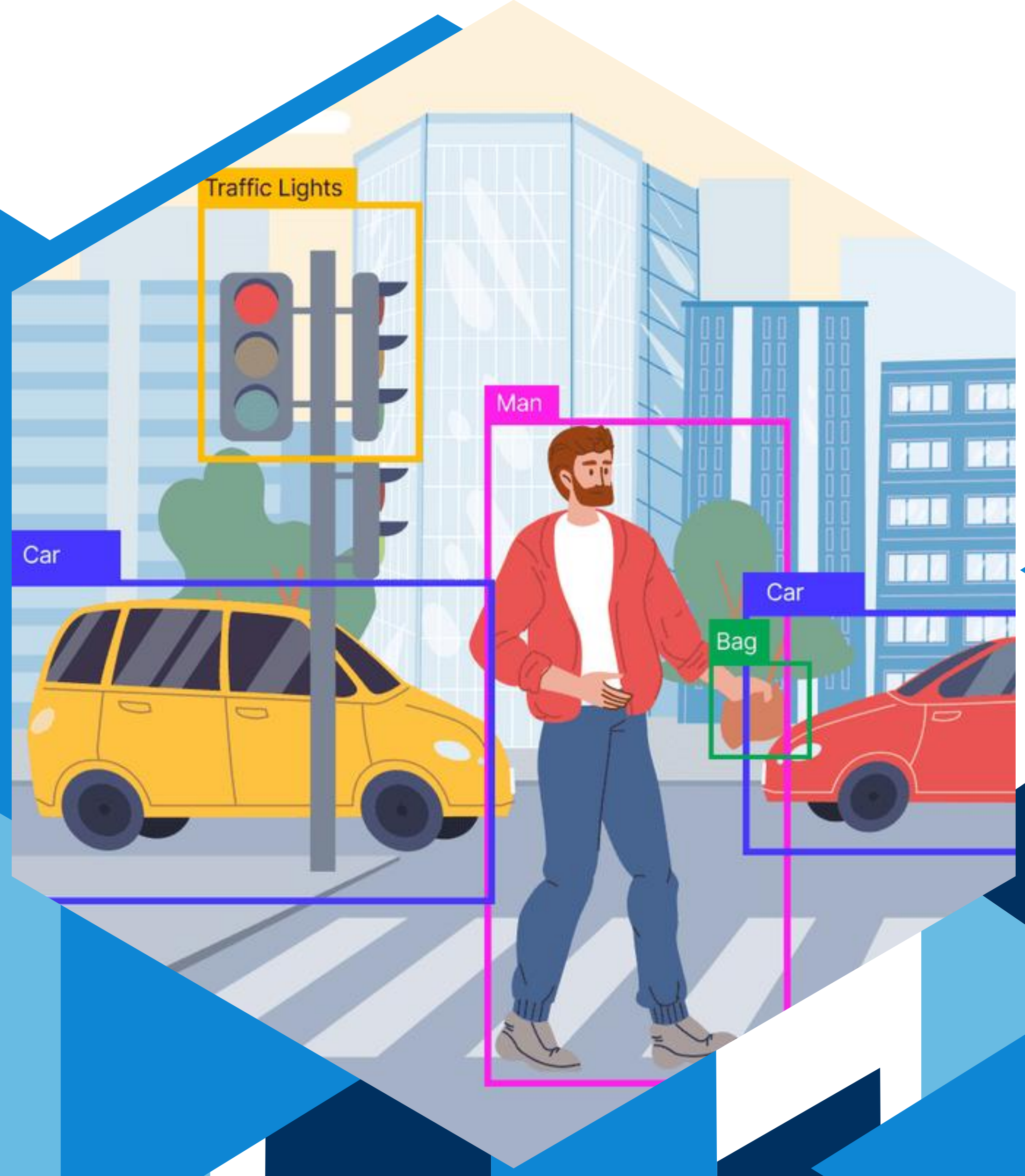
Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi (CVPR 2016)

THE CHALLENGE OF OBJECT DETECTION

- Object detection identifies what and where things are in an image.
- Traditional methods (e.g., R-CNN, DPM) use multiple stages – region proposals, classification, and bounding box refinement.
- These pipelines are slow and complex, limiting real-time performance.
- Goal: Develop a fast, unified model that detects objects in real-time using a single neural network.

Key Idea: “**You Only Look Once**” – detect everything in one pass.

EVOLUTION OF OBJECT DETECTION



- DPM (Deformable Parts Model): Uses sliding windows; accurate but slow.
- R-CNN → Region proposals + CNN + SVM (40 sec/image)
- Fast R-CNN → Shared computation, but still needs region proposals.
- Faster R-CNN → Region Proposal Network (7 FPS).
- These methods rely on multi-stage pipelines and local reasoning.

YOLO VS PREVIOUS METHODS

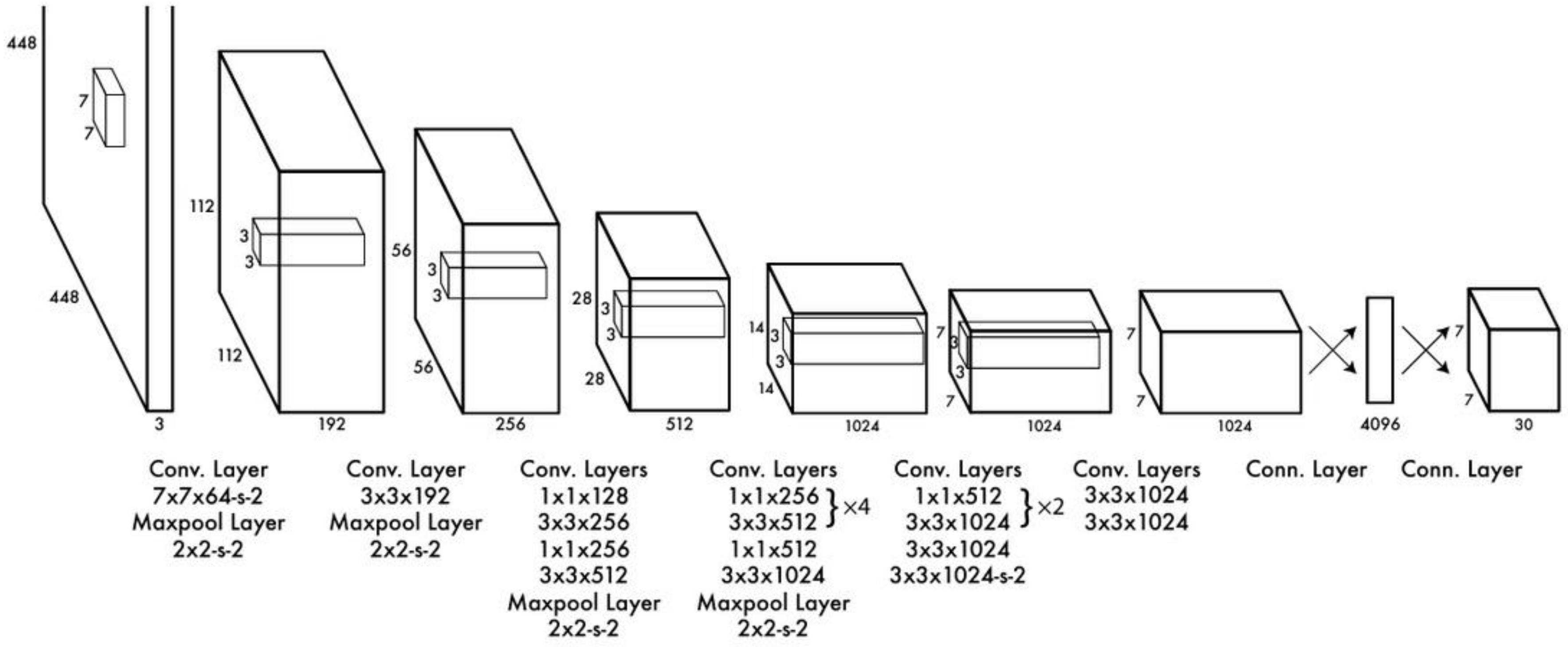
Method	Speed	Pipeline
R-CNN	Slow	Multi-stage
Fast R-CNN	Moderate	Multi-stage
Faster R-CNN	Faster	Multi-stage
YOLO	**Real-Time (45–150 FPS)**	**Single Network**

YOLO reframes detection as a single regression problem from image

pixels → bounding boxes + class probabilities.

YOLO MODEL OVERVIEW

- Input image divided into S × S grid (S=7).
- Each cell predicts:
  - B bounding boxes (B=2)
  - Confidence scores
  - C class probabilities (C=20 for PASCAL VOC)
- Output: 7 × 7 × 30 tensor → class & location predictions.



The Architecture

NETWORK ARCHITECTURE

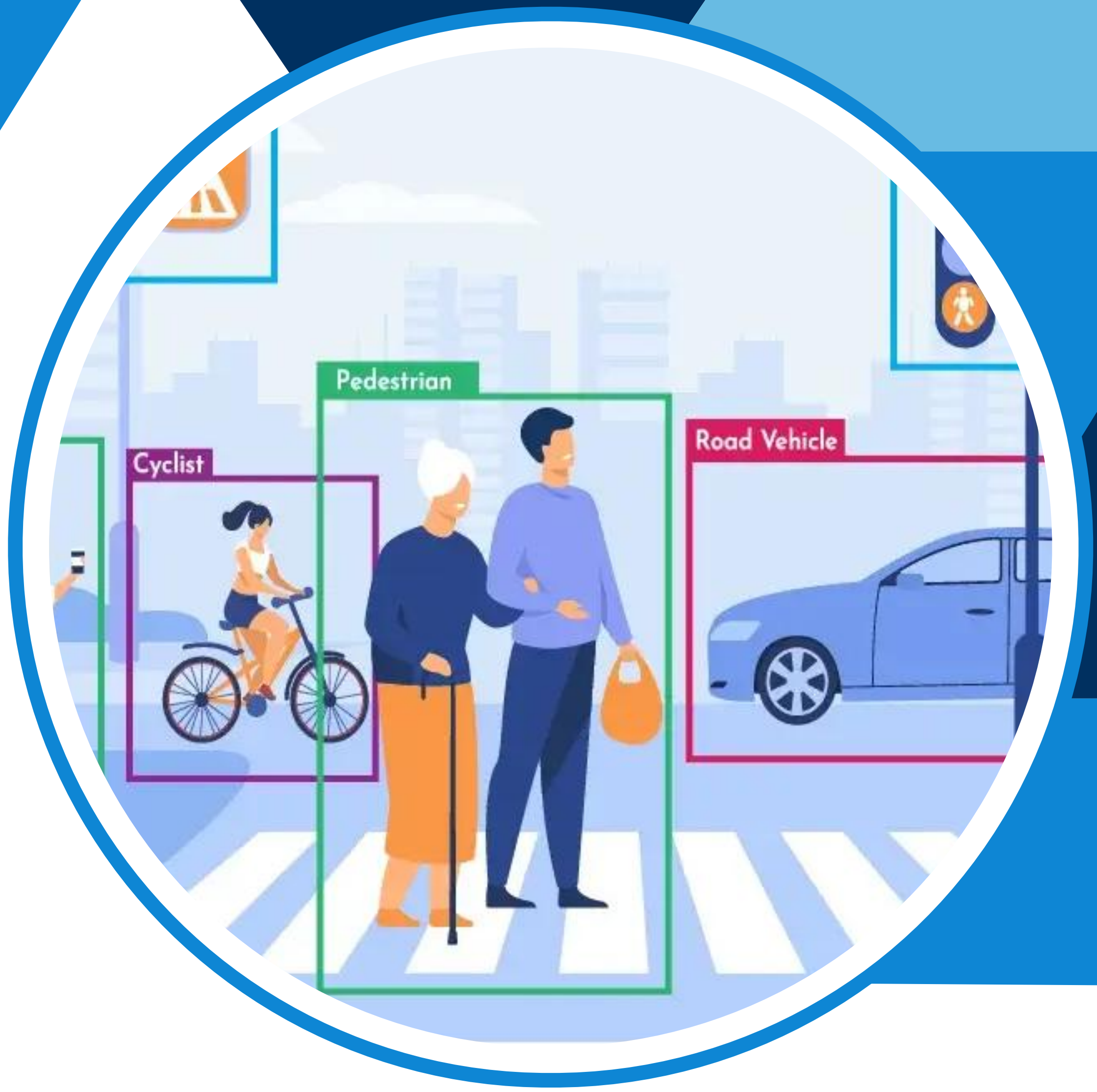
- 24 convolutional + 2 fully connected layers
- Inspired by GoogLeNet but simplified (uses 1×1 and 3×3 conv layers).
- Trained on ImageNet, fine-tuned for detection (448×448 images)
- Loss Function: Sum-squared error with weighting for
  - Localization (λcoord = 5)
  - No-object confidence (λnoobj = 0.5)
- Activation: Leaky ReL

TRAINING & INFERENCE

- **Training:**
  - 135 epochs on PASCAL VOC 2007+2012
  - Data augmentation (scaling, translation, color shift)

- **Inference:**
  - One forward pass → 98 boxes/image
  - Non-Max Suppression removes duplicates

- Fast YOLO: Smaller network (9 conv layers), 155 FPS.



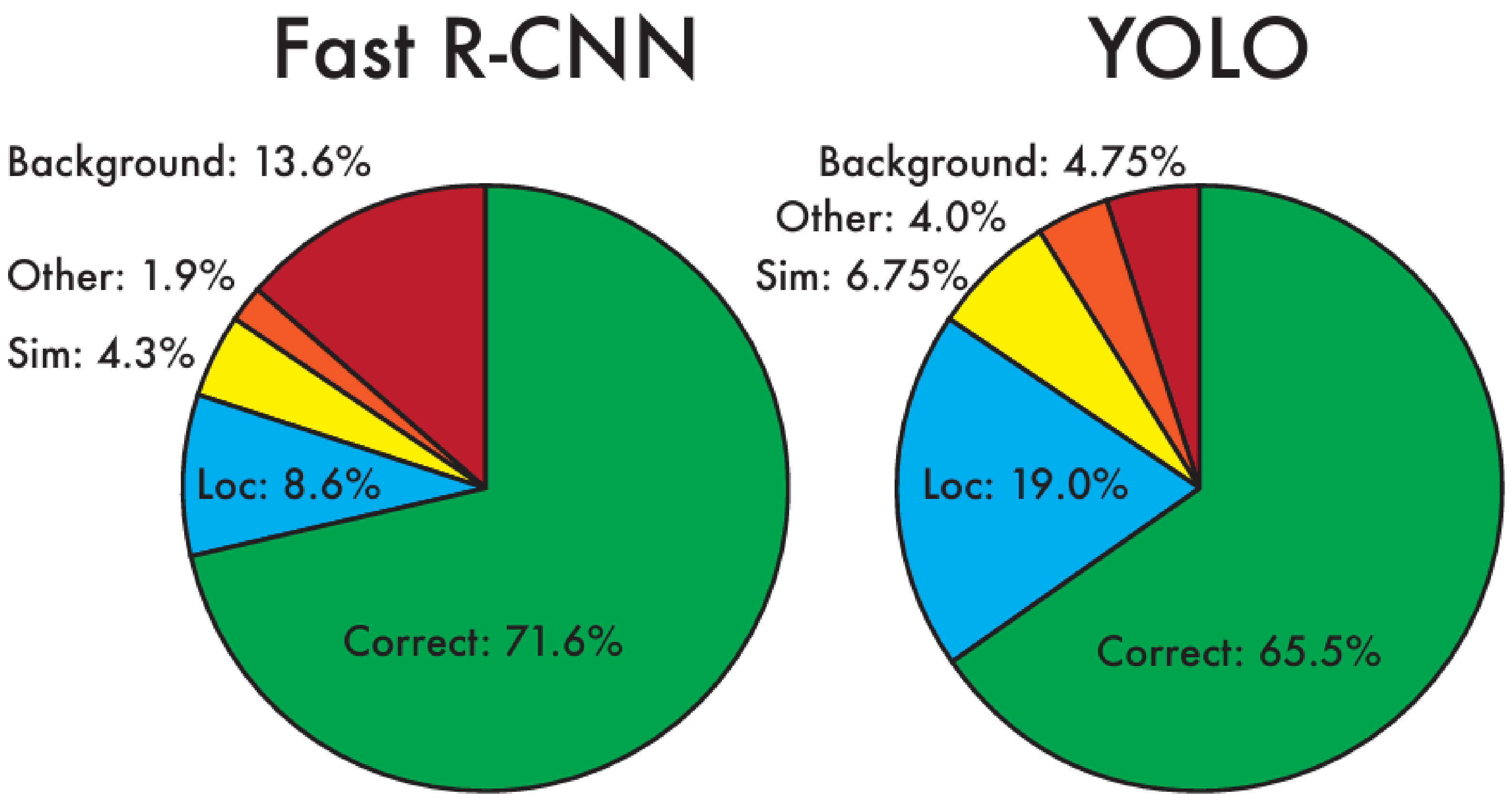
REAL–TIME DETECTION RESULTS

Detector	mAP	FPS
100Hz DPM	16.0	100
Fast YOLO	52.7	155
YOLO	63.4	45
Faster R-CNN	73.2	7

YOLO achieves real-time speed with competitive accuracy.

ERROR ANALYSIS

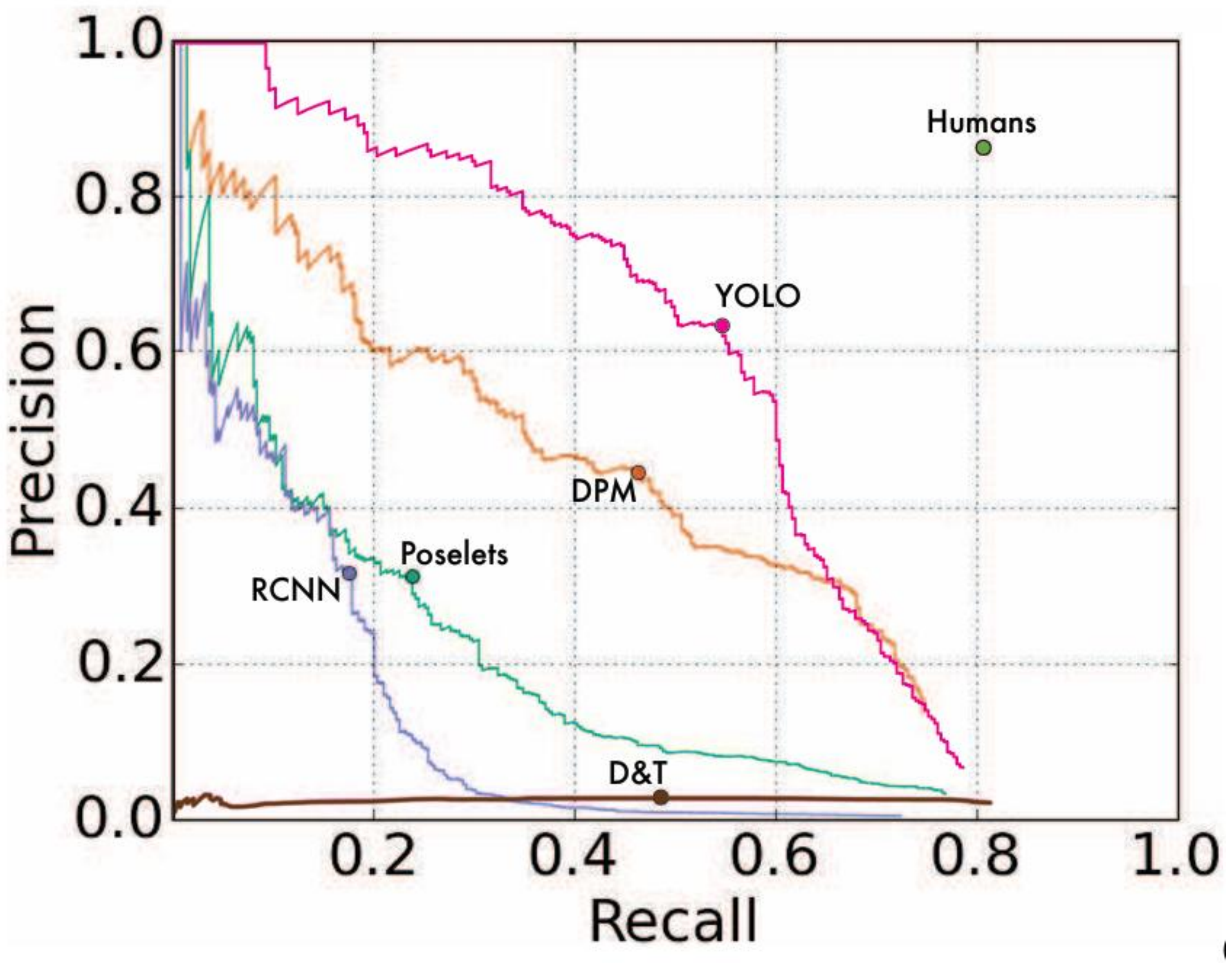
- YOLO makes fewer background (false positive) errors.
- Main weakness: Localization errors (inaccurate bounding boxes).
- Combining YOLO + Fast R-CNN improves mAP from 71.8 → 75.0%.



ErrorAnalysis: FastR-CNNvs. YOLO

GENERALIZATION TO NEW DOMAINS

- Tested on artwork datasets (Picasso, People-Art).
- YOLO performs better than R-CNN and DPM.
- Shows strong generalization ability across visual styles.



Generalization results on Picasso

CONCLUSION & FUTURE WORK

Achievements:

- Introduced unified detection framework – single CNN, real-time speed.
- Improved generalization beyond natural images.

Limitations:

- Struggles with small and overlapping objects.

Future Work:

- Improve localization accuracy.
- Refine model for small object detection.
- Extend to multi-scale detection (→ YOLOv2, v3, v4, YOLOv5+).

THANK YOU

Presented By :

Dissanayake D.K.R.C.K - EG/2020/3910
Samaraweera R.P.R.T.N - EG/2020/4180
Weerasinghe L.W.S.T. - EG/2020/4271
Wickramage W.D.M. - EG/2020/4278