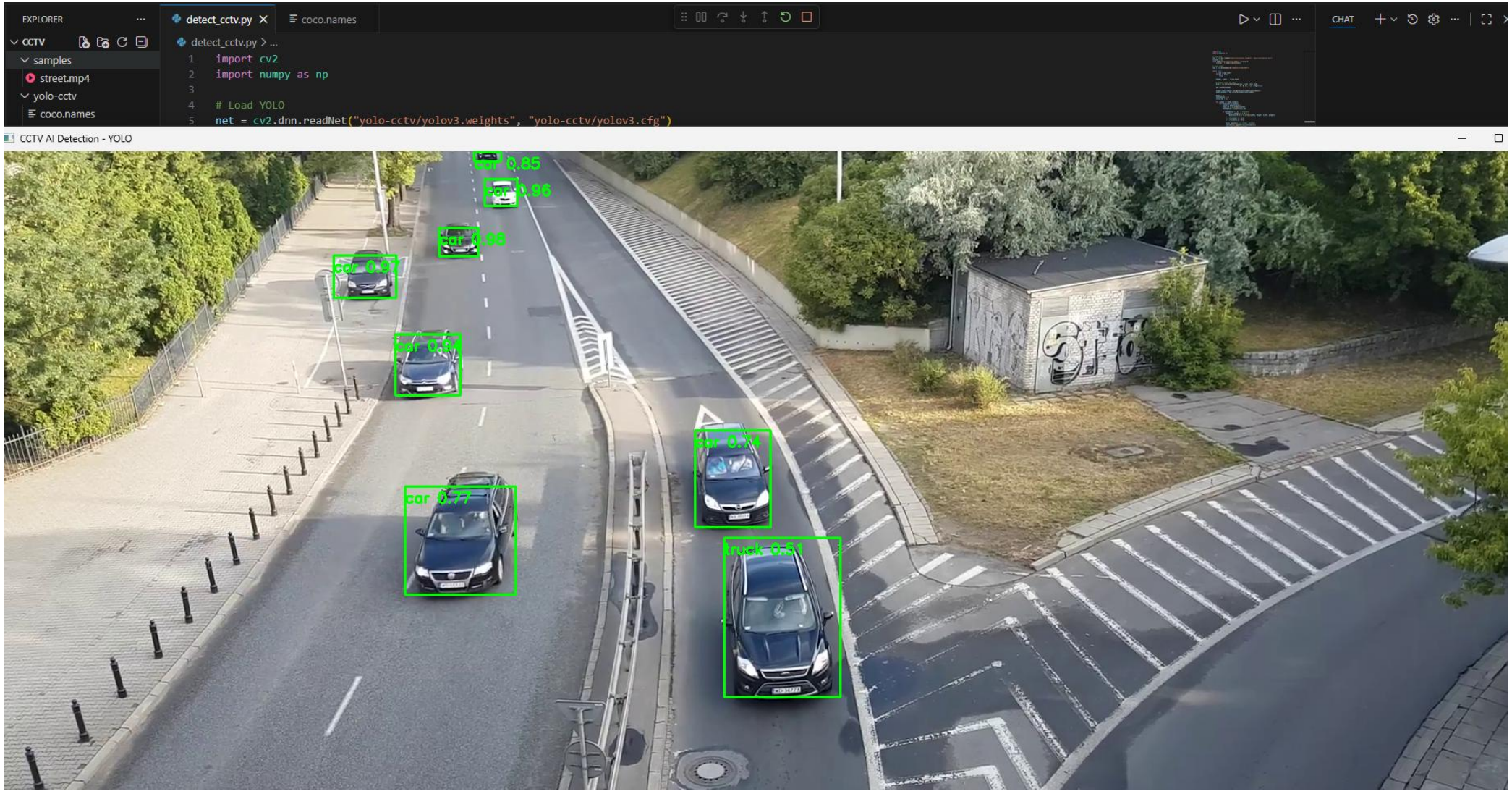
YOLO : You Only Look Once - Real Time Object Detection

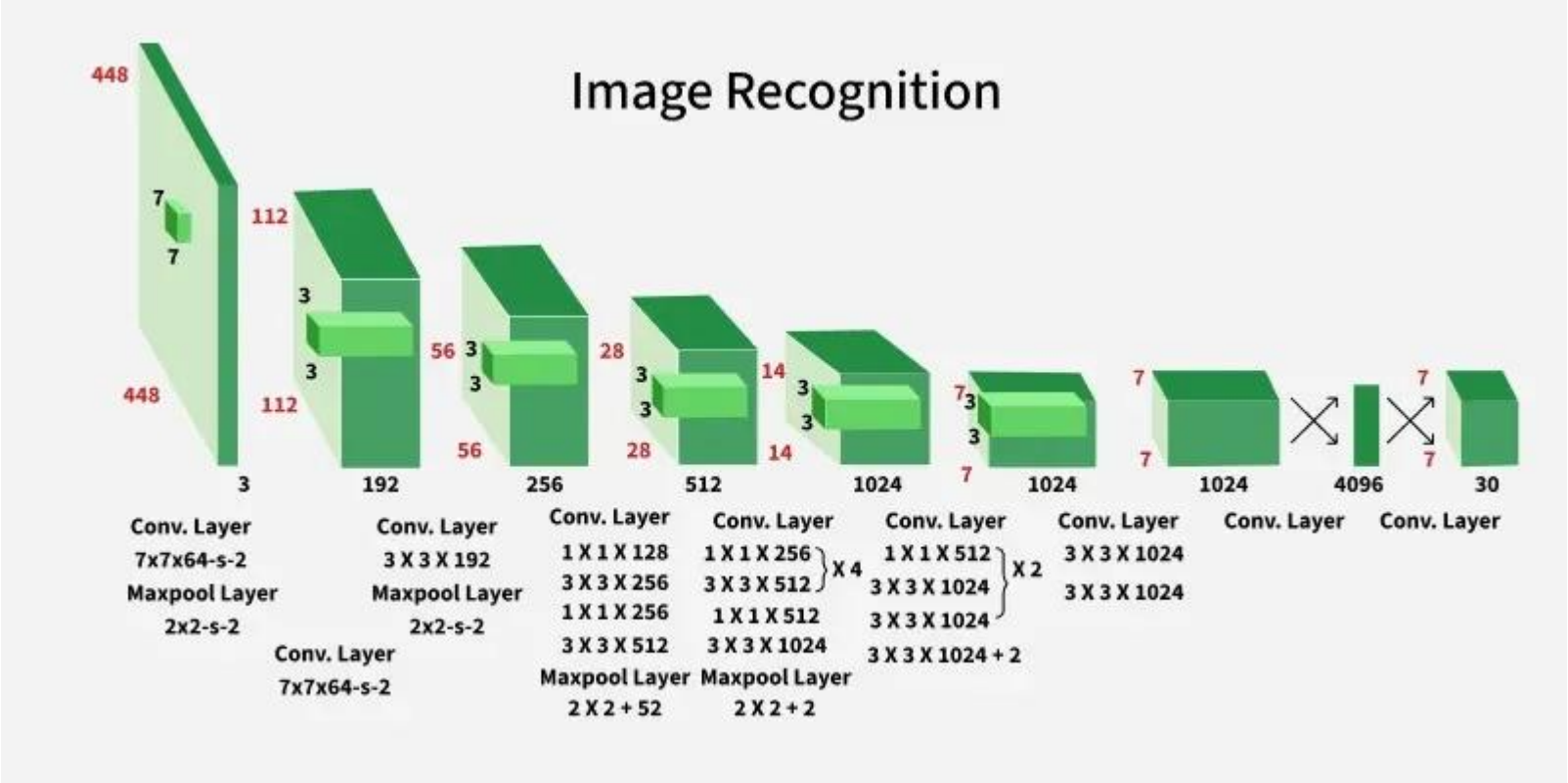
Krijimi i programit për identifikim të objekteve që përdorë modelin e trajnuar me YOLO



Komponentët kryesorë të programit

- yolov3.cfg = si ndërtohet rrjeti
- yolov3.weights = rezultati i trajnimit
- coco.names = emrat e objekteve

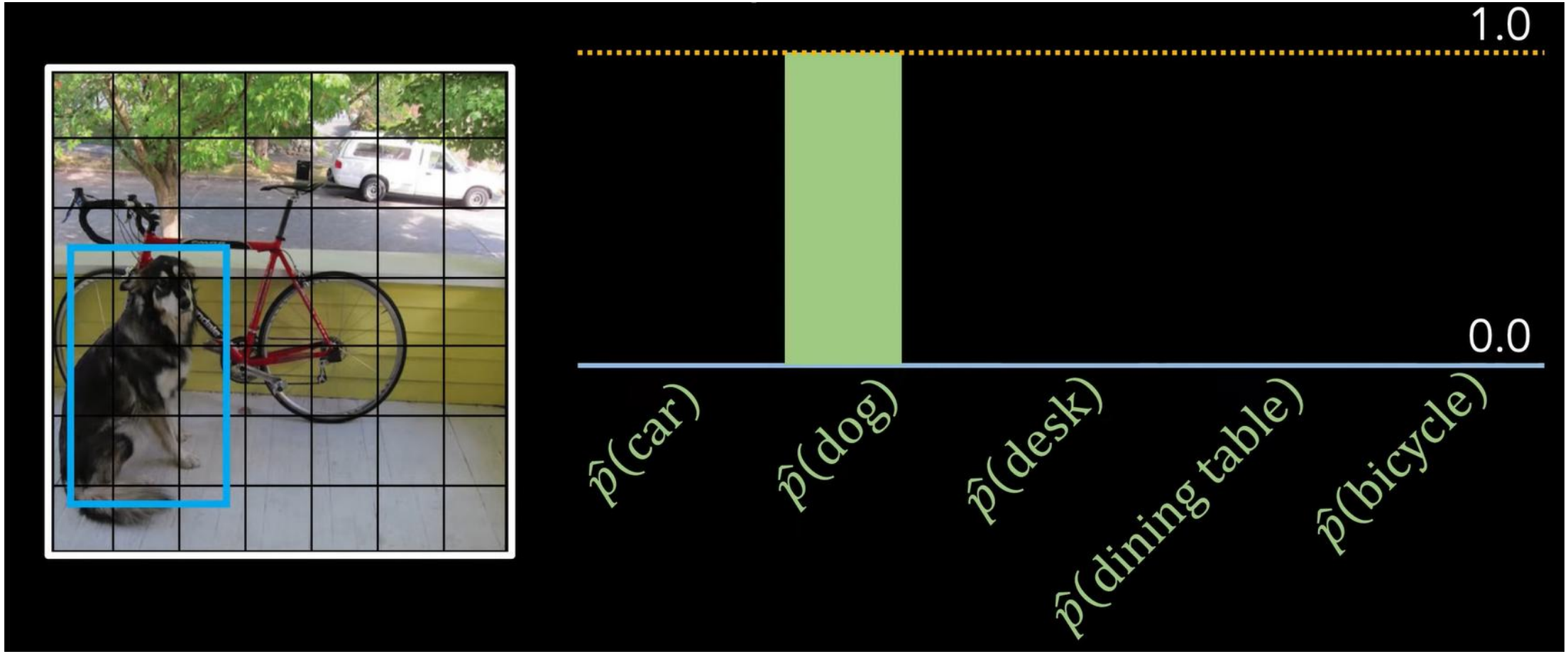
Arkitektura e YOLO – Çka ndodh brenda rrjetit?



YOLO përdor një Convolutional Neural Network si bazë të tij.
Imazhi i hyrjes, 448×448, kalon nëpër disa shtresa konvolucionale dhe pooling.
Gjatë këtij procesi, dimensionet zvogëlohen:
448×448 → 224×224 → 112×112 → 56×56 → 28×28 → 14×14 → 7×7
Në fund, rrjeti krijon një hartë 7×7, dhe për secilën qelizë parashikon:

- 2 bounding boxes
- 1 confidence score
- probabilitetet e klasave për objektin

Trajnimi i modelit : Ground truth(class probabilities)



Ground Truth është vlera e vërtetë dhe e saktë e objektit në imazh.
Bounding box-i i Ground Truth konsiderohet 100% i saktë dhe shërben si standard për të matur sa afër janë parashikimet e modelit gjatë trajnimit

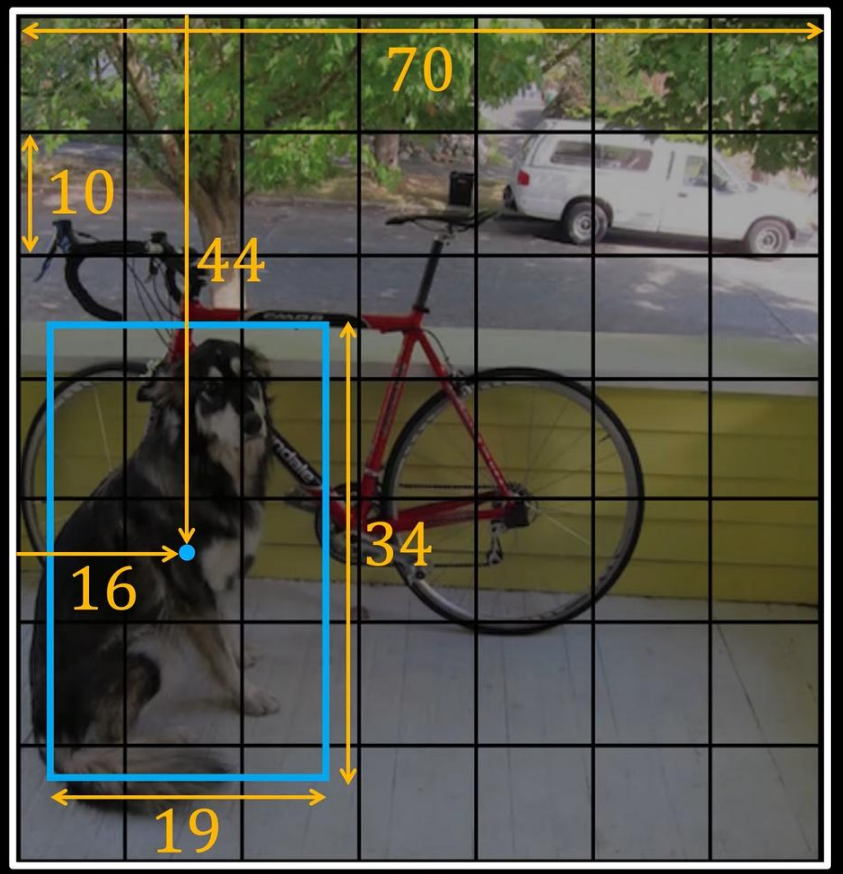
Trajnimi i modelit : Ground truth(box coordinates)

$$x = \frac{16 \,\%\, 10}{10} = 0.60$$

$$y = \frac{44 \,\%\, 10}{10} = 0.40$$

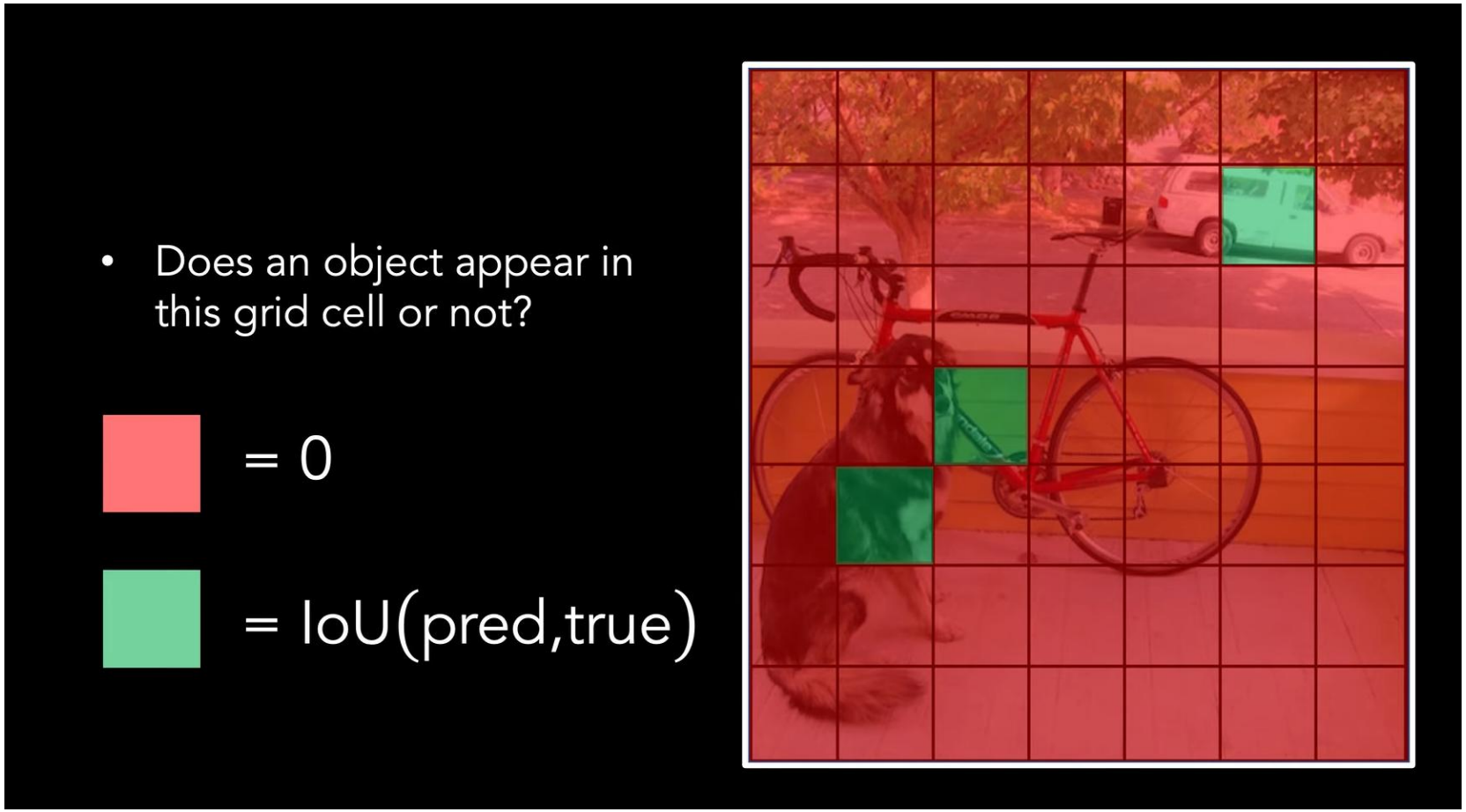
$$w = \frac{19}{70} = 0.27$$

$$h = \frac{34}{70} = 0.49$$



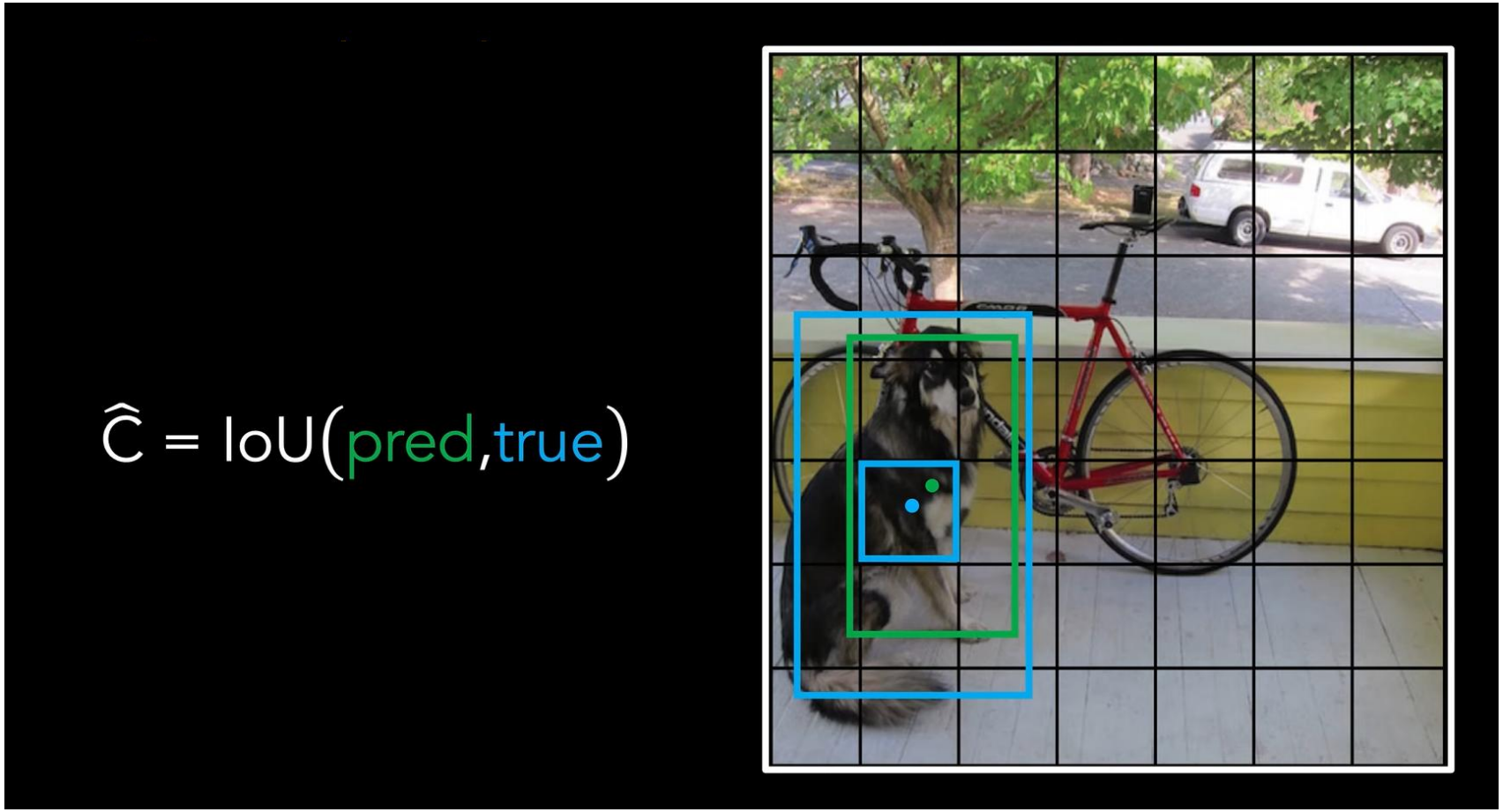
Box Coordinates janë vlerat që përshkruajnë vendndodhjen dhe madhësinë e një objekti në imazh. Ato përdoren nga YOLO për të vizatuar bounding boxes dhe për të llogaritur saktësinë e parashikimeve kundrejt Ground Truth.
Pozita dhe madhësia e objektit i lejojnë YOLO-s jo vetëm të identifikojë çfarë objekti është në imazh, por gjithashtu edhe ku ndodhet dhe sa i madh është. Kjo e bën YOLO-n shumë të vlefshëm për aplikime reale si CCTV, monitorim trafiku dhe robotikë.

Trajnimi i modelit : Ground truth(confidence)



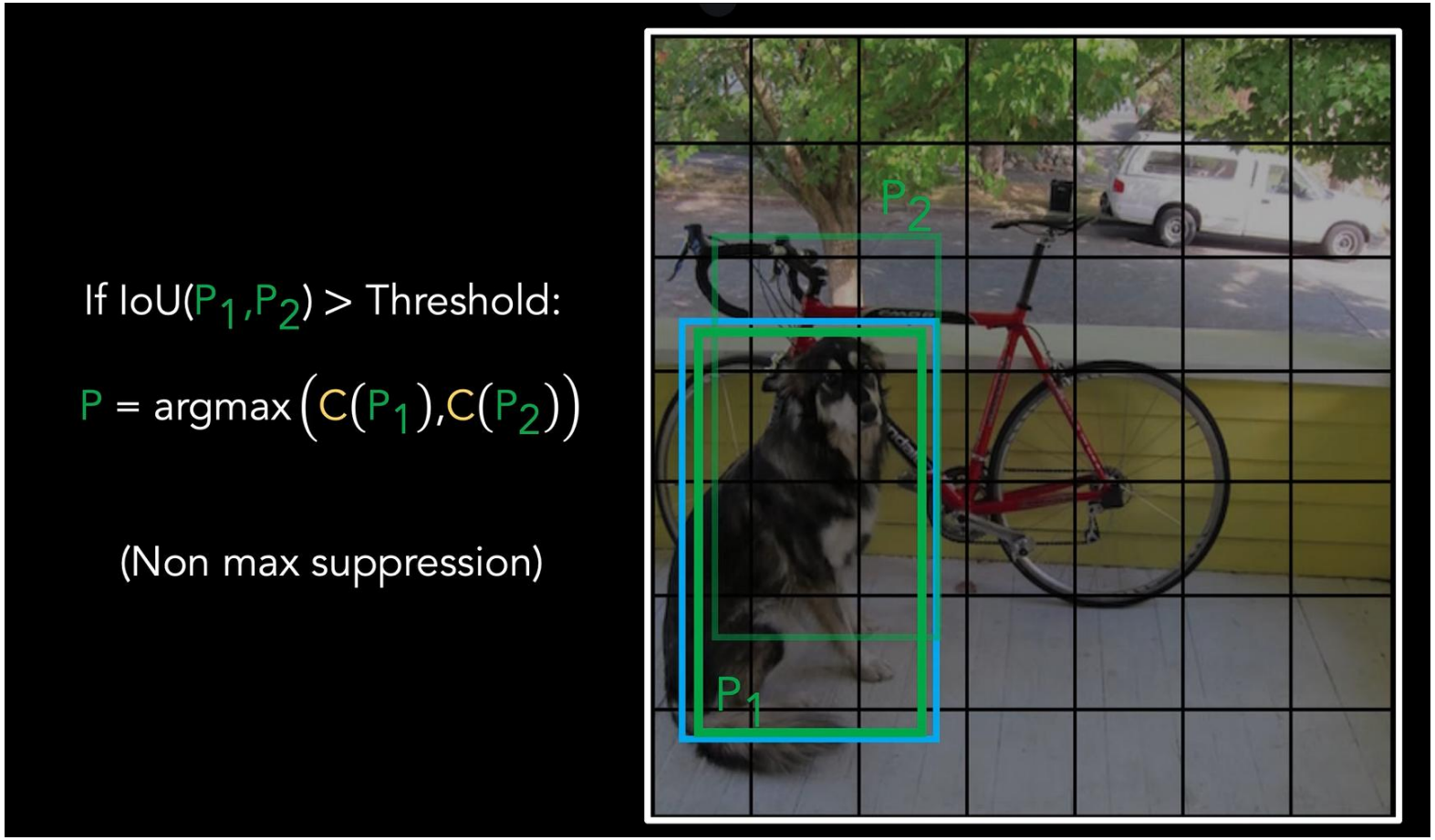
Ground Truth Confidence është vlera binare nga dataset-i që tregon nëse një objekt ekziston në një qelizë specifike të grid-it. Përdoret për të llogaritur gabimin e konfidencës gjatë trajnimit — nëse modeli thotë se ka objekt aty ku nuk ka, ai penalizohet.

Trajnimi i modelit : Ground truth(confidence)



Gjatë trajnimit, IoU mat sa përputhen kutitë e parashikuara nga modeli me Ground Truth. Sa më i lartë IoU, aq më i saktë modeli. IoU përdoret për të llogaritur humbjen dhe për të udhëhequr përmirësimin e modelit gjatë mësimit.
Gjatë trajnimit, modeli parashikon kutitë fillestare. Ne e krahasojmë secilën me Ground Truth duke përdorur IoU. Sa më shumë të përputhen kutitë, humbja është më e ulët. Modeli vazhdimisht përditëson parametrat derisa të arrijë konvergjencë — pra të mësojë të detektojë objektet me saktësi të lartë.

Trajnimi i modelit : Box selection(inference)



Non-Max Suppression është një teknikë që heq kutitë e tepërta të detektuara për të njëjtin objekt, duke zgjedhur vetëm kutinë me probabilitetin më të lartë dhe përputhjen më të mirë

Trajnimi i modelit : Loss function

$$L =$$

Loss function

$$\sum_{i=0}^{S^2} \mathbb{I}_i^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2$$

$$+\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{I}_{ij}^{\text{obj}} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2$$

(Object detection as regression)

$$+\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{I}_{ij}^{\text{obj}} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2$$

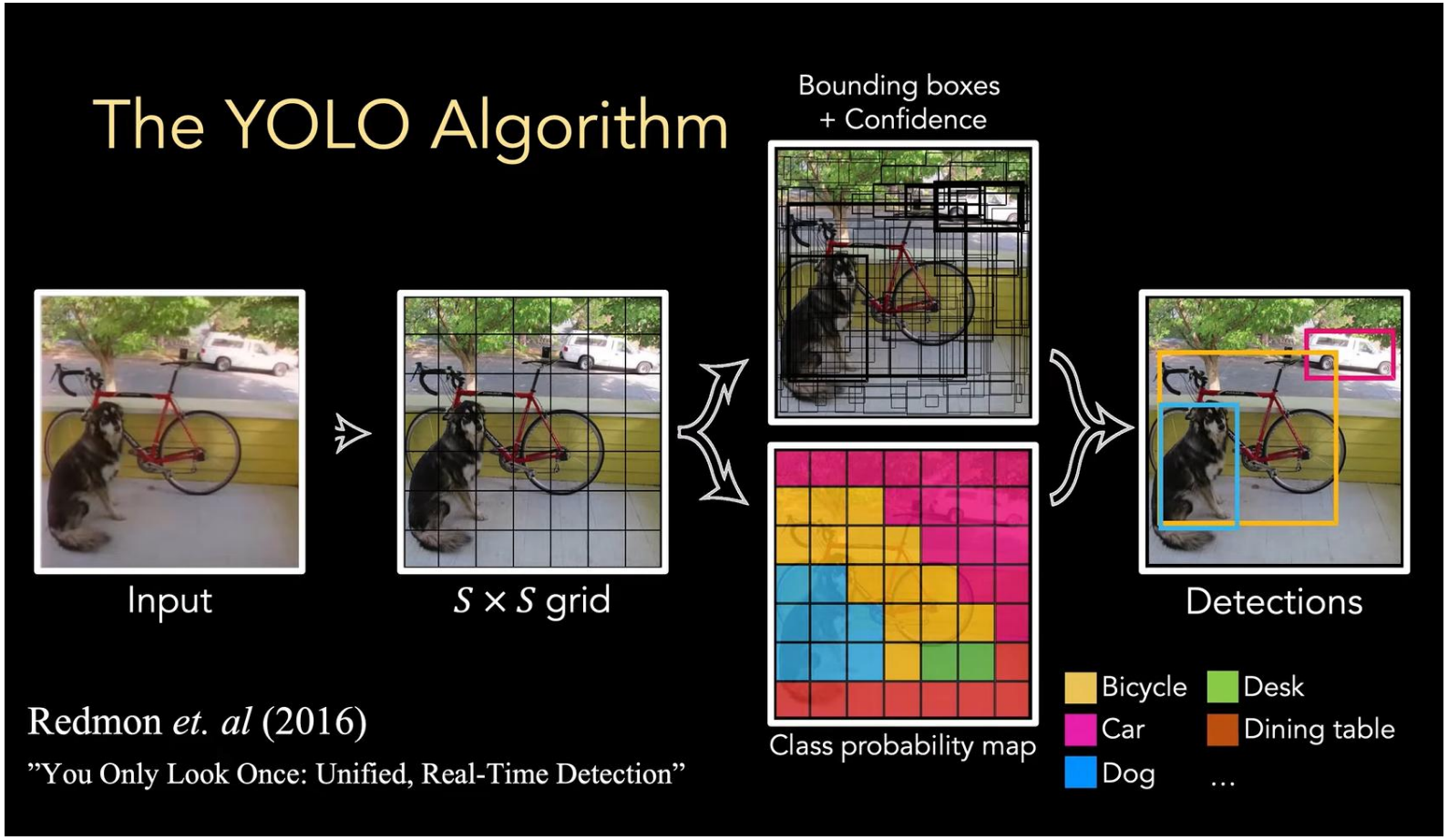
$$+\lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} [\mathbb{I}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2] + \sum_{i=0}^{S^2} \sum_{j=0}^{B} [\mathbb{I}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2]$$

Loss Function në YOLO përbëhet nga tre komponentë kryesorë:

- Localization Loss për pozicionin dhe përmasat e kutive,
- Confidence Loss për vlerësimin e pranisë së objektit dhe
- Classification Loss për të identifikuar saktë klasën e objektit.

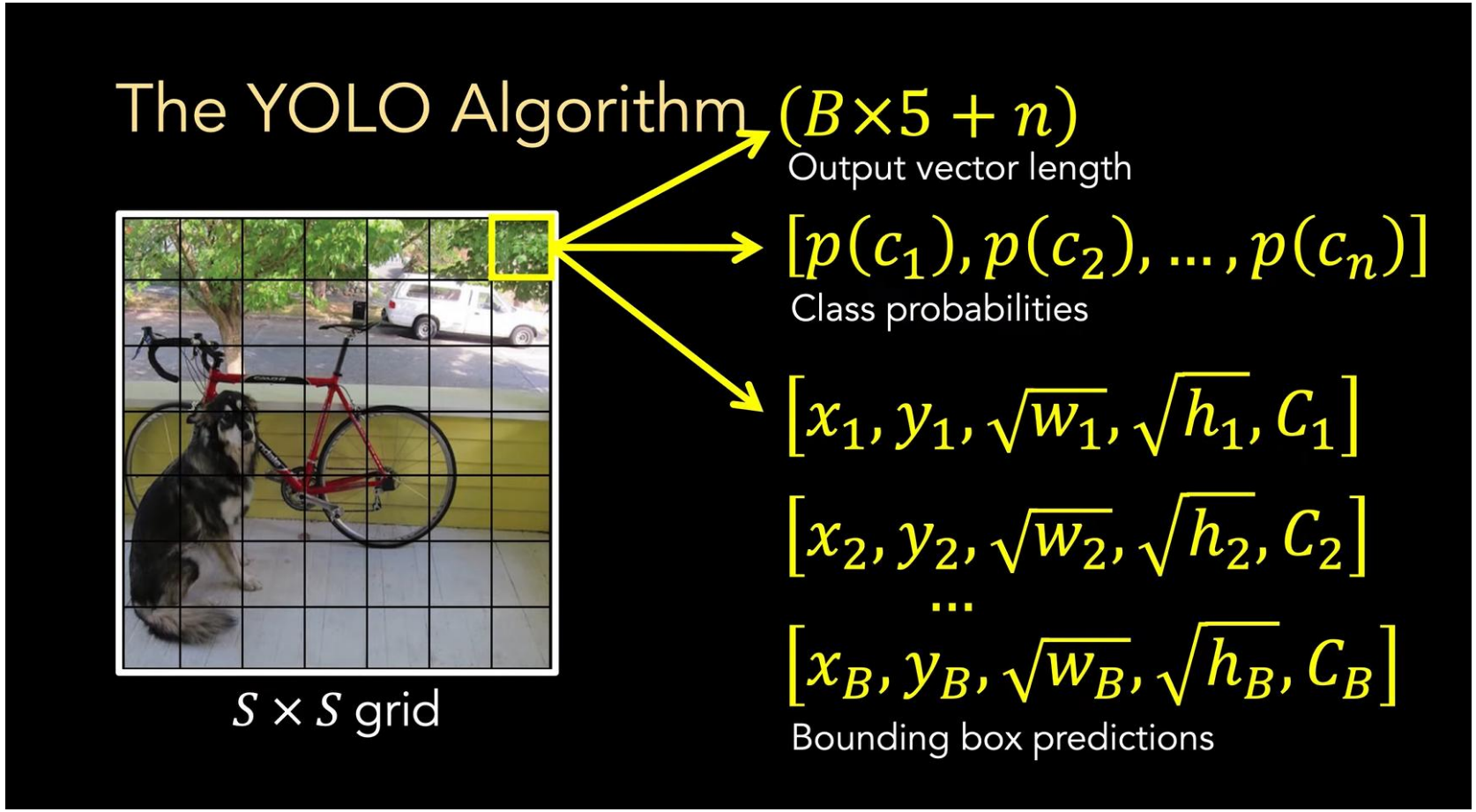
Gjatë trajnimit, modeli përditëson parametrat në mënyrë që të minimizohet vlera totale e humbjes dhe të përmirësojë saktësinë e detektimit.

Inference

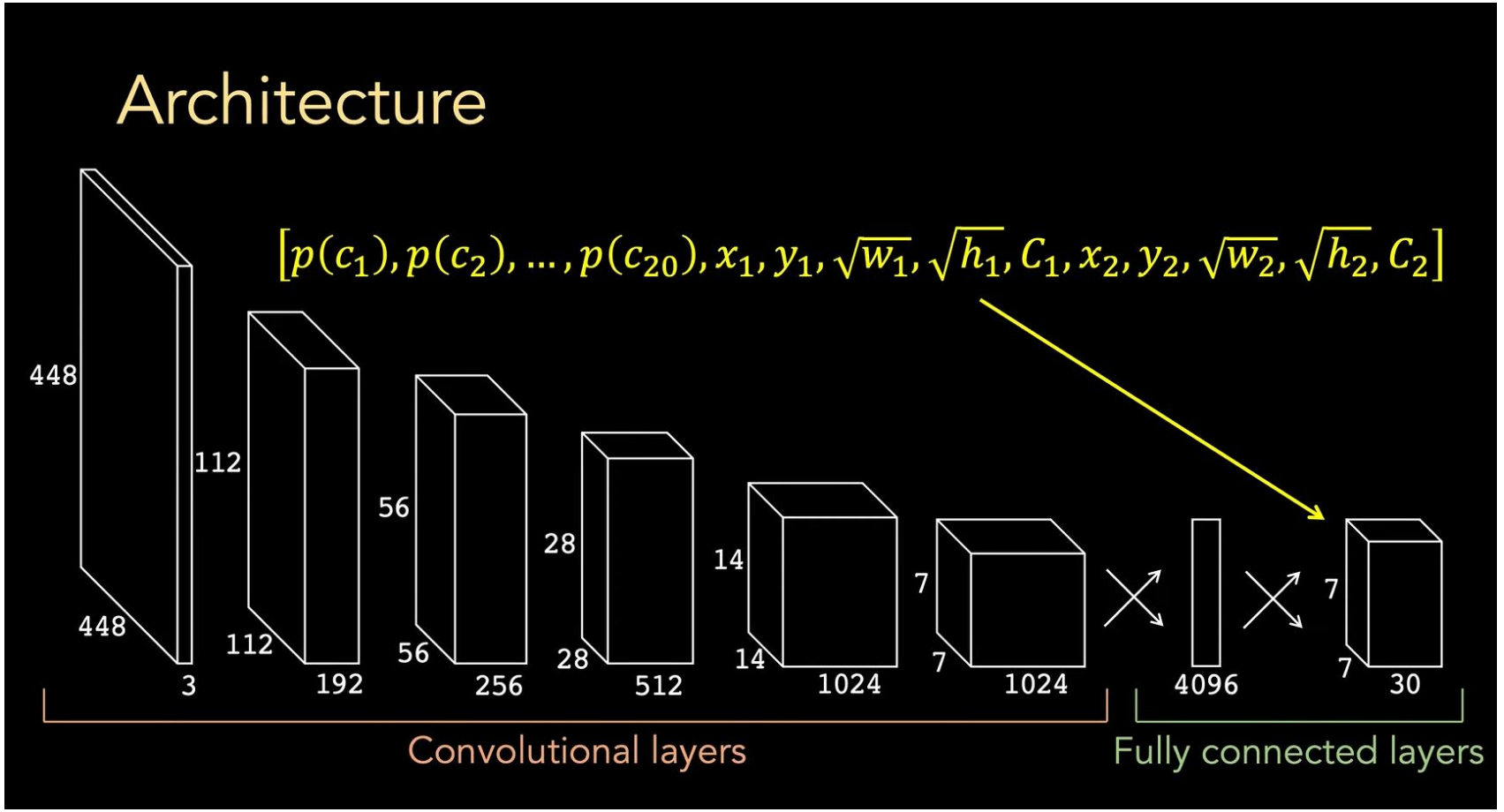


Imazhi ndahet në 7×7 grid që çdo qelizë të jetë përgjegjëse për detektimin e objektit në zonën e saj. Kjo e thjeshton detektimin, e ul kompleksitetin dhe e bën YOLO-n jashtëzakonisht të shpejtë për aplikime në kohë reale si CCTV.
Bounding boxes tregojnë vendndodhjen e objektit, ndërsa confidence score tregon sa i sigurt është modeli në këtë parashikim.
Class Probability Map është grupi i probabiliteteve për secilën klasë objekti në çdo qelizë të grid-it. Ai i tregon modelit se çfarë objekti është

Inference - vektorizimi

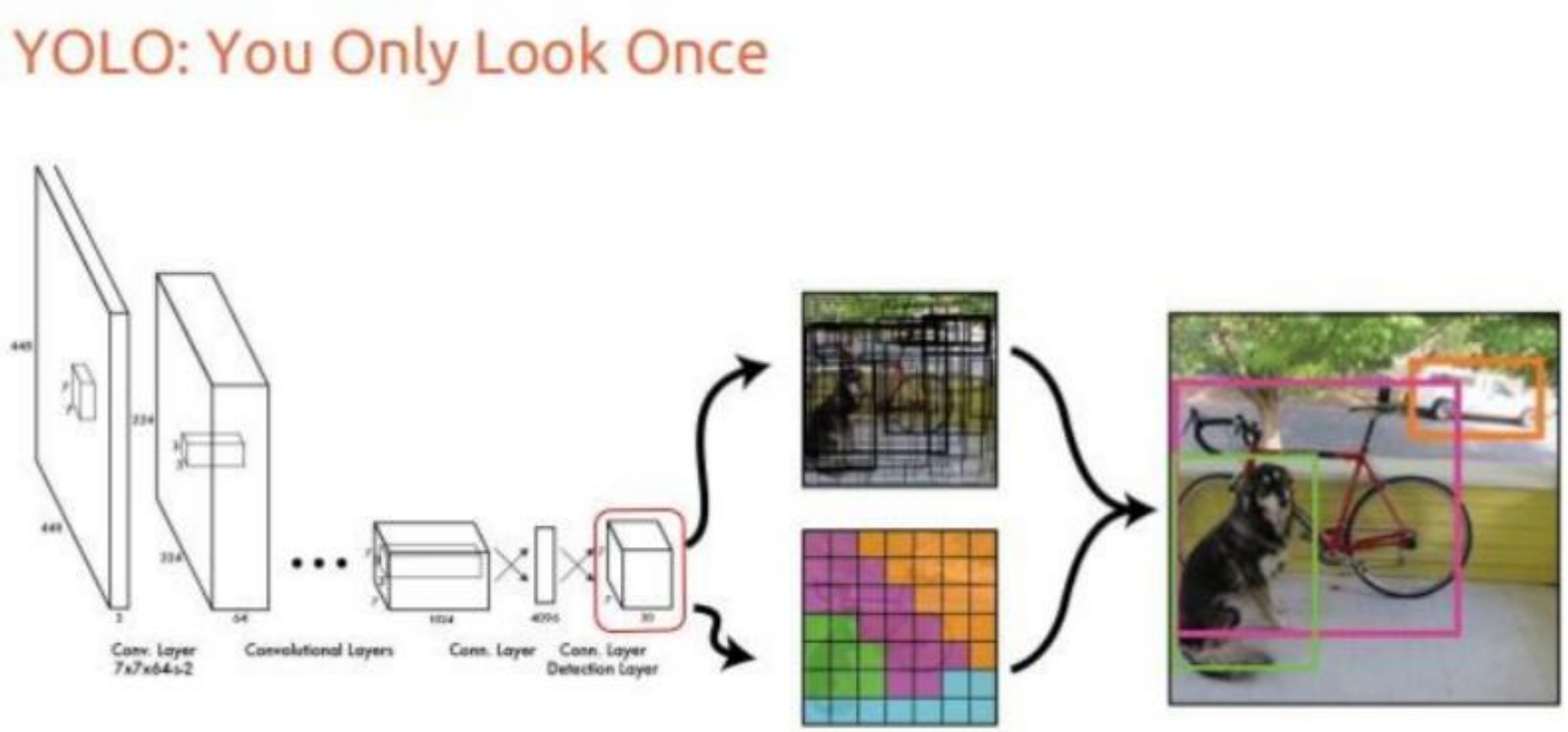


Inference - CNN



YOLO transformon imazhin fillestar përmes shtresave konvolucionale për të nxjerrë tipare, dhe më pas përmes fully connected layers prodhon vektorët e parashikimit të objekteve për secilën nga 49 qelizat e grid-it (7×7), duke përfshirë klasa, pozicion dhe madhësi të bounding box-eve.

Infernce - CNN



Imazhi përpunohet nga CNN për të nxjerrë tipare vizuale.
Për secilën qelizë të grid-it, modeli parashikon coordinates e bounding box-it, confidence score dhe probabilitetet e klasave.
Pas filtrimit të kutive të tepërta, YOLO identifikon objektet finale dhe pozicionin e tyre në imazh në kohë reale.