King Mongkut's University of Technology Thonburi

Real-Time Object Detection

Lecturer: Dr. Thittaporn Ganokratanaa

ตรวจจับวัตถุแบบ real-time

↳ Applied ใช้กับงานนอกจาก Image etc.

King Mongkut's University of Technology Thonburi

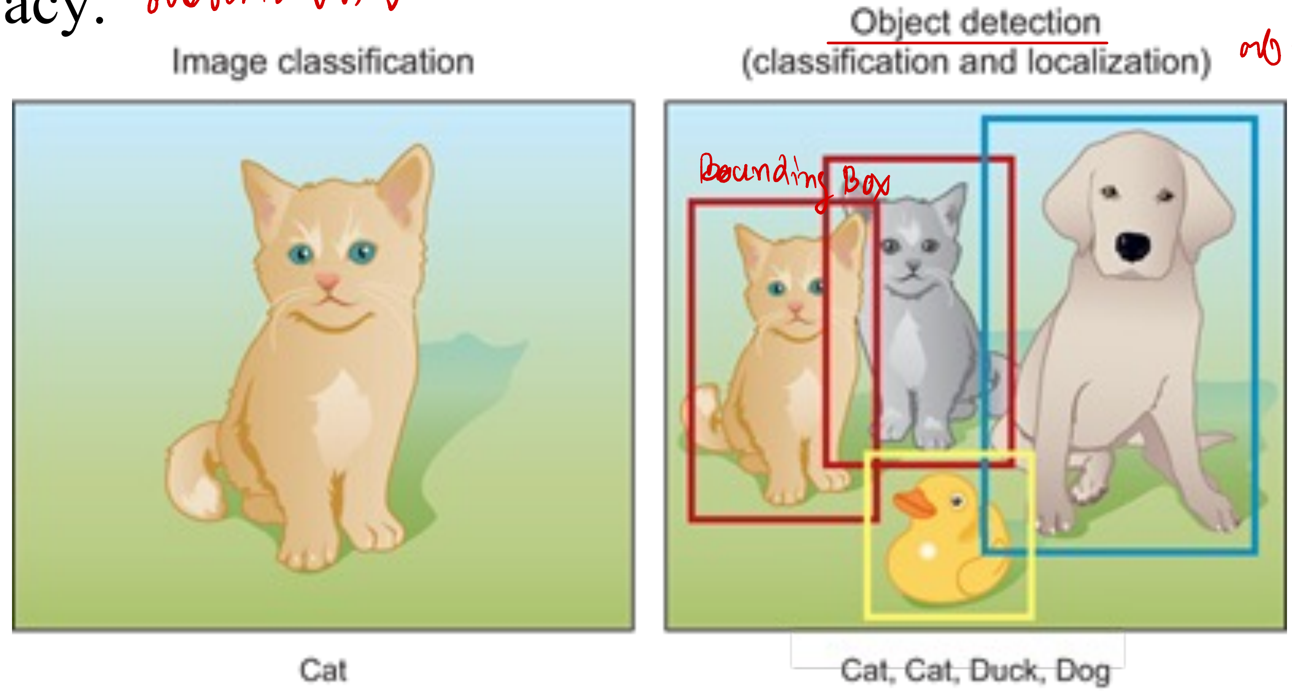
❖ Problem Addressed: Object Detection

ต้องตอบคำถาม ว่าคืออะไรและต้องหาตำแหน่ง ได้

➢ Object detection is the problem of both locating AND classifying objects

➢ Goal of object detection algorithm is to do object detection both fast AND with high accuracy. ต้องทำทั้งเร็ว ฉลาดที่ใช้ในการ ประมวลผลต้องไว

ต้อง classify ได้ ว่าสิ่งนั้นคืออะไร



→ บอกตำแหน่ง object อะไรบ้าง

Image Source [Object Detection]

King Mongkut's University of Technology Thonburi

❖ Importance of Object Detection

Yolo จะ input เป็นภาพ

- ➢ Visual modality is very powerful
- ➢ Humans are able to detect objects and do perception using just this modality in real-time (not needing radar)
- ➢ If we want responsive robot systems that work real-time (without specialized sensors), almost real-time vision based object detection can help greatly.

เป็นหลักความรู้ของ CV และ ML ก็เพียงพอ

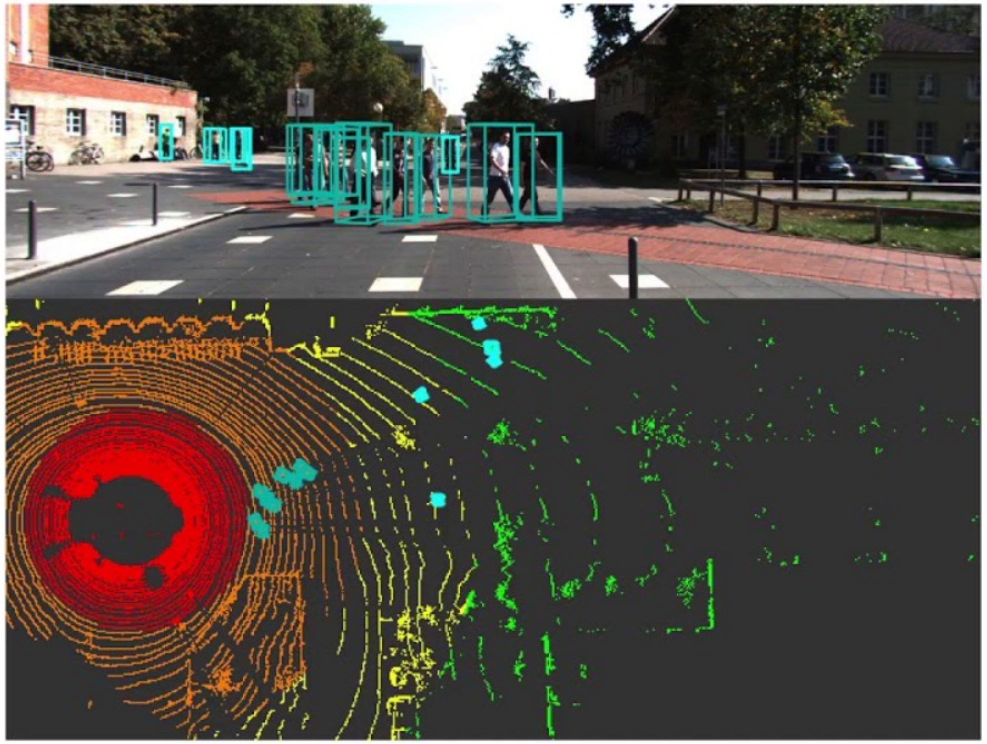


Image Source: [Object Detection and Lidar]

King Mongkut's University of Technology Thonburi

YOLO: You only look once

A timeline of YOLO versions

2015 — YOLOv1

first version

2016 — YOLO9000 v2

Image ...

Better, faster, stronger

2018 — YOLOv3

Dev ... 3 ...

Output: Logistic regression

แบ่งภาพออกเป็น grid ... center ...

Grid cell ... detect object

2020 — Scaled YOLOv4, PP-YOLO, YOLOv5, YOLOv6

Improve speed & accuracy

ตรวจจับ object ...

2021 — YOLOX, YOLOR, PP-YOLOv2

2022 — DAMO YOLO, PP-YOLOE, YOLOv7, YOLOv6

focus quantization

Improve small optimization

2023 — YOLOv8, YOLO-NAS

train บน Coco dataset

Improve Accuracy

Source: [YOLO]

Prepared by: Dr. THITTAPORN GANOKRATANAA

King Mongkut's University of Technology Thonburi

❖ YOLO Overview

หลักการทำงาน

แต่ละ grid มีหน้าที่ detect object

- First, image is split into a S×S grid

  ตาม

- For each grid square, generate B bounding boxes

  ว่ามี object หรือไม่

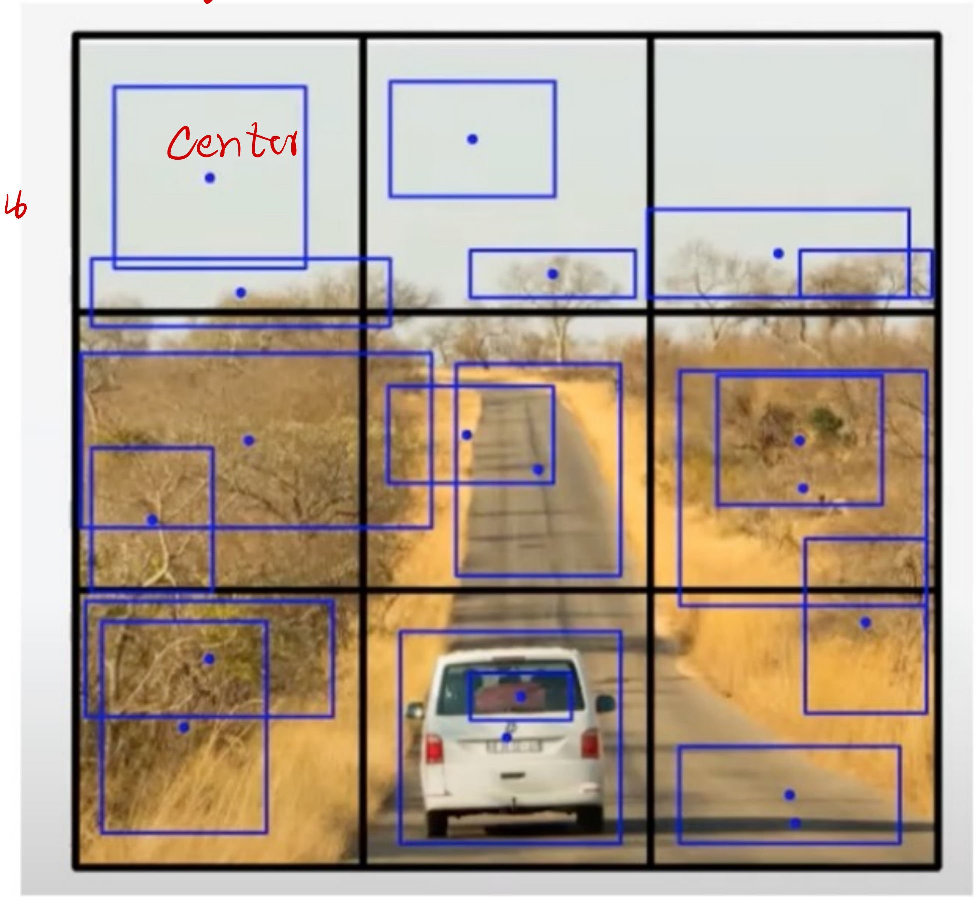
- For each bounding box, there are 5 predictions:

  กว้าง สูง ของ Bounding Box

  x, y, w, h, confidence

coordinate of Bounding box's center
↓
relate กับ grid cell ของ grid cell normalize (0,1)

บอกว่า Bounding box ถูกต้องแค่ไหน ว่ามี object หรือไม่ prob เป็น IOU ระหว่าง ground truth.



S = 3, B = 2

Source: YOLO

King Mongkut's University of Technology Thonburi

❖ YOLO Training

- ➢ YOLO is a regression algorithm. What is X? What is Y?

ใช้แก้ prob จากการทำนายค่าจริง

Input ของ Algorithm เป็นรูปภาพ ใน computer ครอบไปด้วย Array หรือ metrix

- ➢ X is simple, just an image width (in pixels) * height (in pixels) * RGB values

ทั้งหมด pixel ที่มีอยู่ทั้งหมด

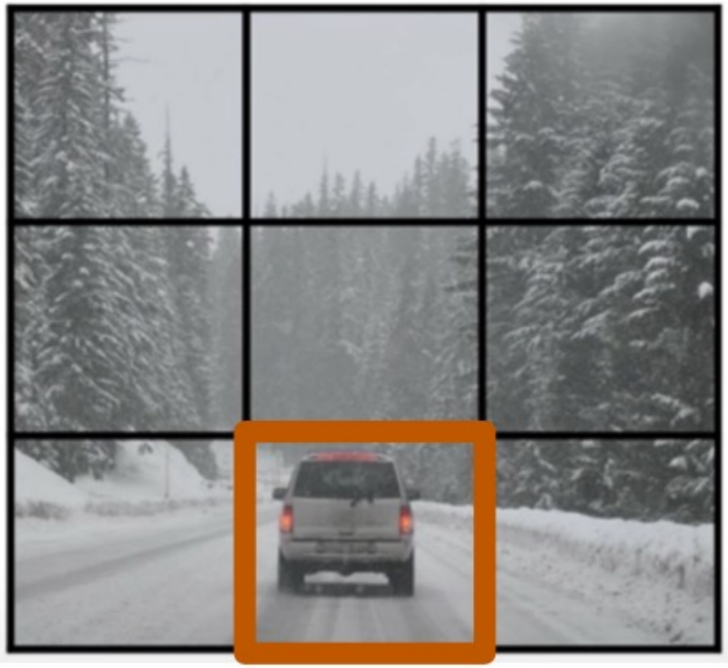
output

- ➢ Y is a tensor of size S * S * (B * 5 + C)
- ➢ B*5 + C term represents the predictions + class predicted distribution for a grid block

For each grid block, we have a vector like this. For this example B is 2 and C is 2

p_1
b_x_1
b_y_1
b_h_1
b_w_1
p_2
b_x_2
b_y_2
b_h_2
b_w_2
c_1
c_2

Split image into SxS grid
แต่ละ grid cell รับผิดชอบ predict Bounding Boxes



GT label example:

1
b_x_1
b_y_1
b_h_1
b_w_1
0
?
?
?
?
c_1 = 1
c_2 = 0

Source: YOLO

King Mongkut's University of Technology Thonburi

❖ YOLO Architecture

- ➢ Now that we know the input and output, we can discuss the model

Image size : 448 x 448 x 3 — 3 ช่องสี RGB

CNN : Capture ข้อมูลที่สำคัญ

แต่ละ layer detect feature ที่สำคัญ ต่างๆ feature ที่ผ่าน extract มันมีข้อมูล

- ➢ We are given 448 by 448 by 3 as our input.

ทำ convo กับ image 7 layer

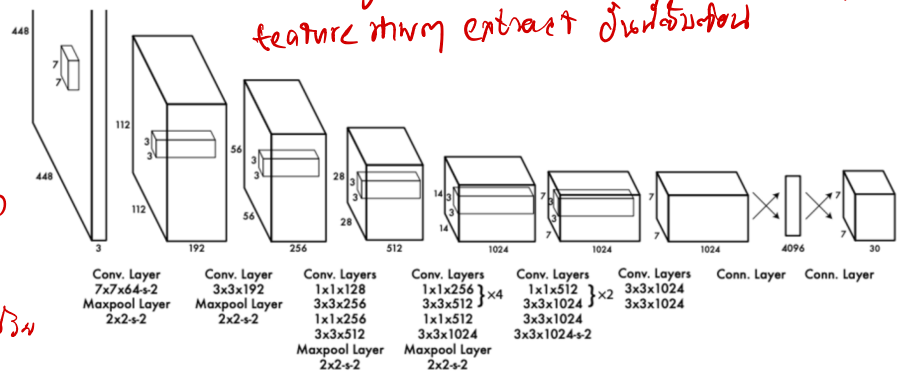
- ➢ Implementation uses 7 convolution layers

แบ่งภาพเป็น grid 7x7 — จำนวน class ที่ model สามารถ detect ได้

- ➢ Paper parameters: S = 7, B = 2, C = 20

แต่ละ grid cell Predict 2 Bounding Boxes → บอก position & dimension

- ➢ Output is S*S*(5B+C) = 7*7*(5*2+20) = 7*7*30



Multiple bounding Boxes
คือ box ที่ครอบวัตถุที่ต้องการ Detect
Object 1 Object อาจได้ หลาย Boxes

Finalize : ตัด Box ที่ไม่จำเป็นออกไป

❖ Non-maximal suppression

เลือกตัดตัว Redundant

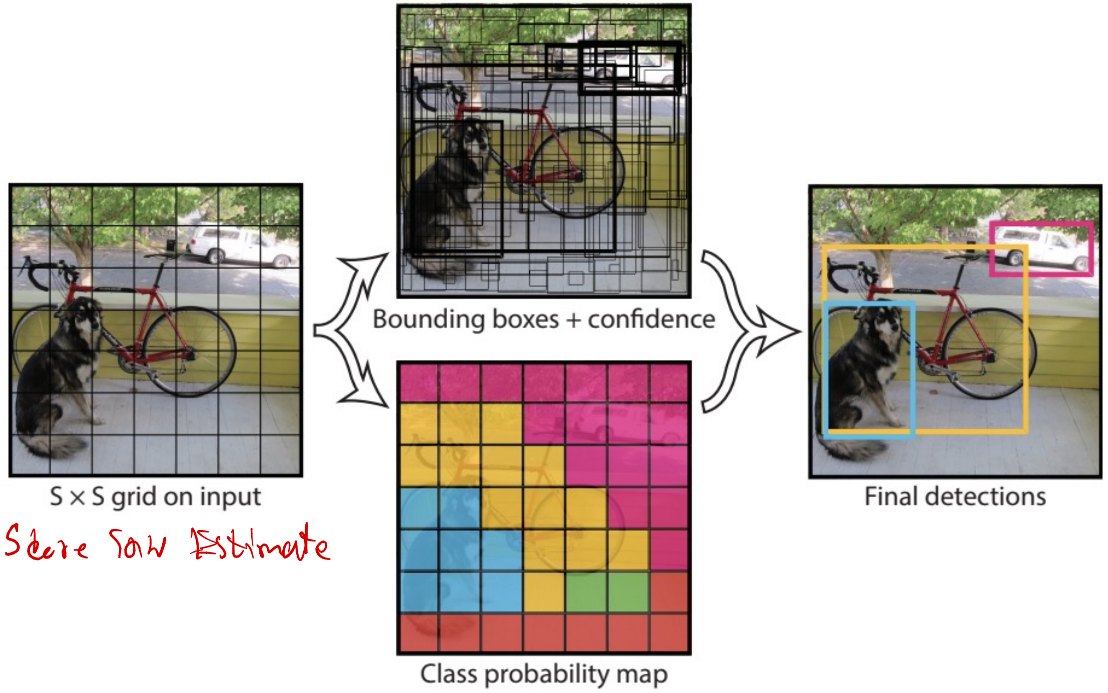
- We then use the output to make final detections

Screen confidence score ที่ต่ำกว่า threshold ทิ้ง

confidence score ต่ำ filter ออก

- Use a threshold to filter out bounding boxes with low P(Object)
- In order to know the class for the bounding box compute score take argmax over the distribution Pr(Class|Object) for the grid the bounding box's center is in

คำนวณ class score โดย Estimate จาก model



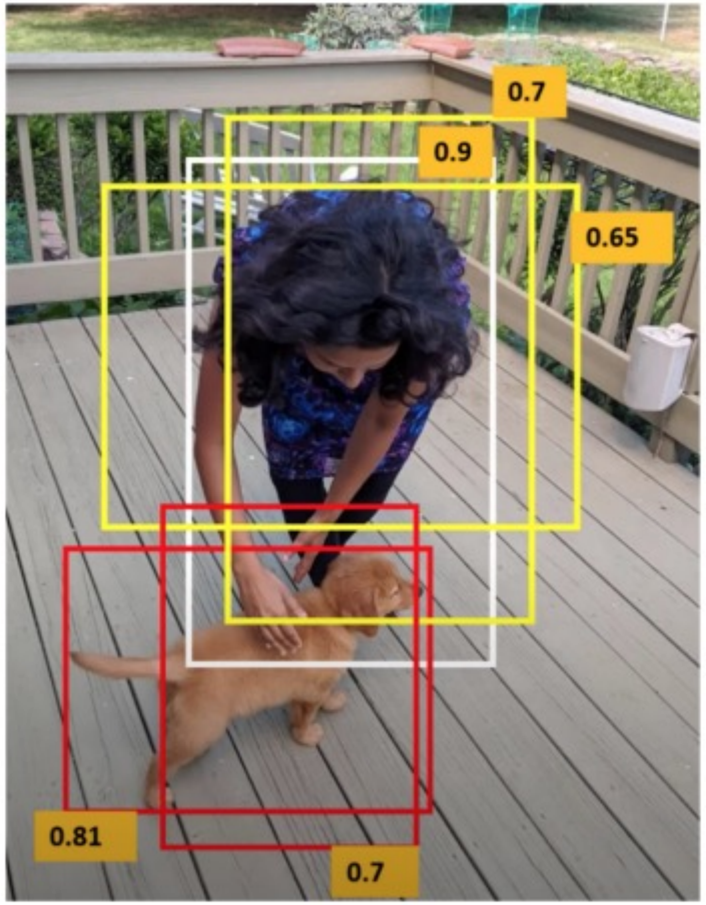
$$\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}}$$

King Mongkut's University of Technology Thonburi

mean Average precision

เป็นเรื่อง Enhance M.AP ส่วน impact
ของการเพิ่ม Accuracy

❖ YOLO Prediction



- Most of the time objects fall in one grid, however it is still possible to get redundant boxes (rare case as object must be close to multiple grid cells for this to happen)
- Discard bounding box with high overlap (keeping the bounding box with highest confidence)
- Adds 2-3% on final mAP score

ทำให้ได้ผลลัพธ์ออกมาดีขึ้นกว่า

Source: [YOLO]

King Mongkut's University of Technology Thonburi

ลด Error
→ ต้อง optimize model

YOLO Objective Function

ใช้ 3 function

Ground truth

Ensure ที่ Bounding Boxes

Localization loss

Set to 5 to increase the loss of bounding box predictions

Position loss → MSE

Position & size ของ Bounding Boxes

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

→ location → x, y

GT bbox x-coordinate in the ith cell

GT bbox y-coordinate in the ith cell

Predicted bbox x-coordinate in the ith cell

Predicted bbox y-coordinate in the ith cell

Sum-squared error

For each grid cell

For each grid box

'1' if object appears in the ith cell and the jth box detect it, '0' otherwise

Classification

$$+\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \right]$$

→ size → w, h

GT bbox height in the ith cell

Square root to reduce the range of the values

GT bbox width in the ith cell

Predicted bbox width in the ith cell

Predicted bbox height in the ith cell

loss Error

Source: YOLO

Prepared by: Dr. THITTAPORN GANOKRATANAA

คือการ penalize model ที่ predict ผิด

YOLO Objective Function (Cont.)

Confidence loss

$$+\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[\left(C_i-\hat{C}_i\right)^2\right]$$

GT confidence score

เทียบกับ.

Predicted confidence score

Confidence error when an object is detected in the ith cell

Set to 0.5 to decrease the loss for empty boxes

$$+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{noobj}\left[\left(C_i-\hat{C}_i\right)^2\right]$$

'1' if there is no object in the ith cell, '0' otherwise

Confidence error when an object not detected in the ith cell

Clasification loss

Predicted conditional probability of an object of class c appearing in the ith cell

$$+\sum_{i=0}^{S^2}\mathbb{1}_{i}^{obj}\sum_{c\in classes}\left[\left(p_i(c)-\hat{p}_i(c)\right)^2\right]$$

For each grid cell

For each class

GT conditional probability of class c appearing in the ith cell

เซลล์ ลำดับที่ i

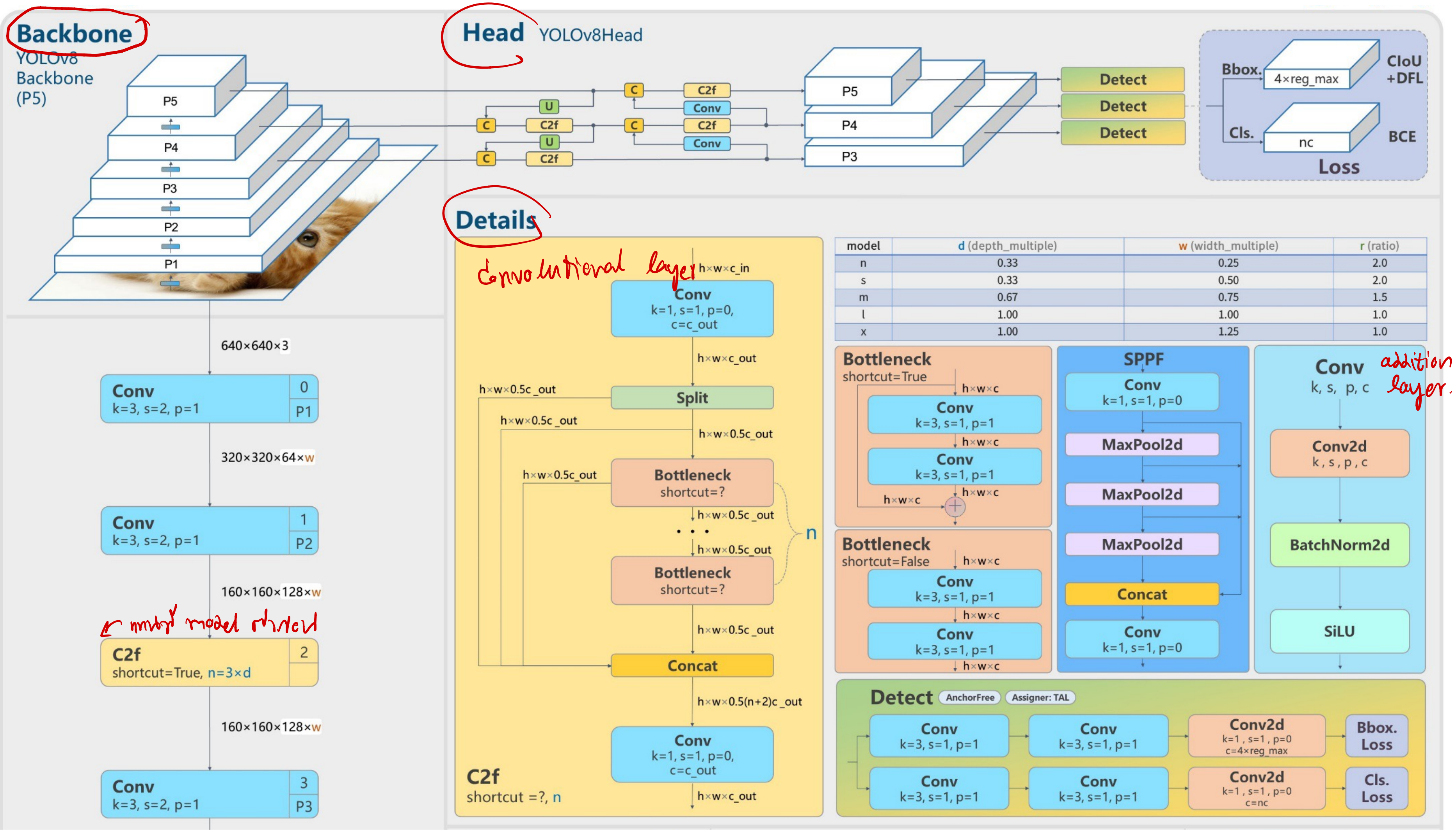
❖ YOLO V8

- ➢ YOLOv8 uses a similar backbone as YOLOv5 with some changes on the CSPLayer, now called the C2f module.

- ➢ The C2f module (cross-stage partial bottleneck with two convolutions) combines high-level features with contextual information to improve detection accuracy

❖ YOLO V8 Architecture

ใช้ 3×3 ใน Backbone



model	d (depth_multiple)	w (width_multiple)	r (ratio)
n	0.33	0.25	2.0
s	0.33	0.50	2.0
m	0.67	0.75	1.5
l	1.00	1.00	1.0
x	1.00	1.25	1.0

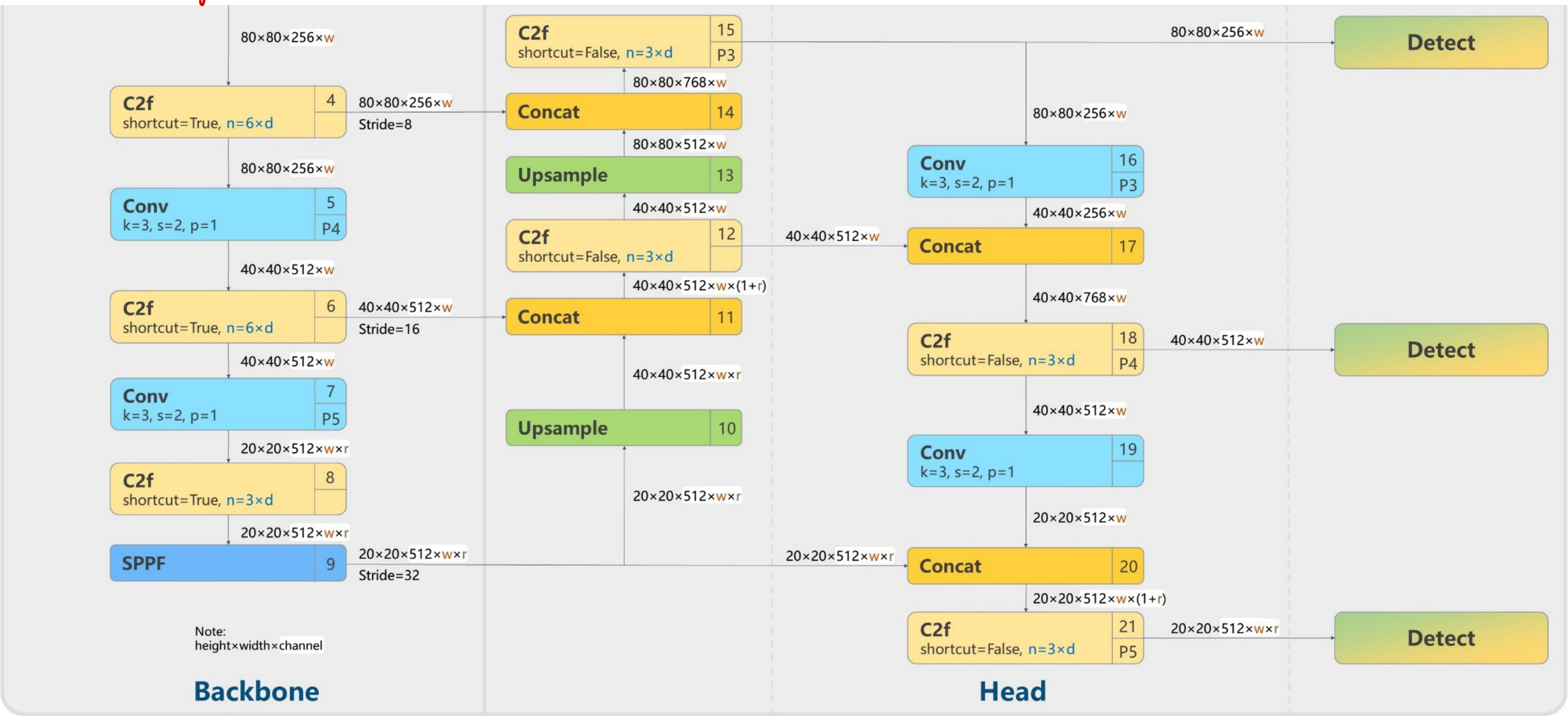
Source: YOLO

King Mongkut's University of Technology Thonburi

YOLO จะเพื่อเพิ่มประสิทธิภาพ และความแม่นยำ
↳ Community พัฒนา

❖ YOLO V8 Architecture (Cont.)

first layer จะ ใช้ 3x3



Source: YOLO

❖ YOLO V8 Experiment

➢ Using this Google Colab:

https://colab.research.google.com/drive/14x7_B44tBvAe8RzuETDVJ14cYWstnT2D?usp=sharing

Exercise

Extract this video into frame and label it into four classes (bus, taxi, car, and pedestrian), then generate the model to classify those four classes using yolov8

King Mongkut's University of Technology Thonburi

Conclusion

ผลของการพัฒนาระบบที่ช่วยติดตามจับกุมคนร้าย ? เทคนิคที่สำคัญ : ตรวจจับวัตถุ

- The research focused on utilizing AI technology to augment police efficiency in Thailand.
- We aimed to enhance law enforcement capabilities and bolster public trust in crime prevention measures.
- By employing AI in crime data analysis, leveraging intelligent CCTV technology for crime monitoring, and integrating real-time alerts for suspicious activities to police.

ข้อเสนอแนะต่อ : เพิ่มในส่วนของการวิเคราะห์เหตุการณ์ผิดปกติ (หรือ Anomaly Detection)

→ DL gen AI, gen adversarial Network เพื่อ gen ภาพ หา scene ที่ผิดปกติ
ที่ผ่านมาการ learn มาก่อน ทดสอบด้วย scene จากที่ยังไม่ learn.

→ neural network อื่นๆ

King Mongkut's University of Technology Thonburi

Q&A