

Criminal

Problem

- การระบุตำแหน่ง และแยกประเภท
- ต้องมีความรวดเร็ว และแม่นยำ (Realtime)

YOLO (You Only Look Once)

YOLO Overview

- แบ่งภาพออกเป็นตาราง $S \times S$
 - สร้าง Bounding boxes สำหรับแต่ละช่องตาราง ขึ้นมา B boxes
 - ในแต่ละ Bounding boxes จะมีการทำนาย 5 ตัวแปร ($x, y, w, h, \text{confidence}$)
- ทำ IOU เกือบกับ ground truth

YOLO Training

เป็น Regression โดย x คือ รูปภาพ (Array) $w \times h \times \text{RGB values}$

y คือ tensor ขนาด $S \times S \times (B \times 5 + C)$

จำนวน class ที่ detect ได้

YOLO Archi.

- ประกอบด้วย 7 convolution layers
- ยิ่ง layer ลึกเท่าไรยิ่ง detect รูปร่างที่ซับซ้อนขึ้น

Non-maximal Suppression

เมื่อมี Bounding boxes มากกว่า 1 จะตั้ง Threshold เพื่อไม่ให้มี Bounding box ซ้ำกันอันเดียวที่ดีที่สุด (วัดตาม prediction score)

- ลด Redundant
- ลด multiple bounding boxes

YOLO Objective fcn.

ประกอบด้วย 3 พังก์ชัน

- 1) Localization loss : ทำให้มั่นใจว่า BB ที่ predict ได้ ตรงกับ Ground truth หรือไหม ทั้ง x, y และ w, h

2) Confidence loss : เปรียบเทียบค่าความมั่นใจ

3) Classification loss : เปรียบเทียบ Probability