

ADARLAB AI Training Course

Lec8 Homework Report

110511118 陳孟頌

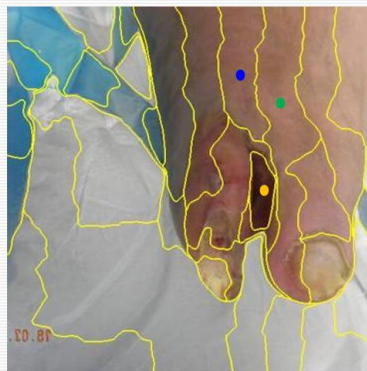
Part I. The most interesting topic in the three project

- **Topic:**

我最感興趣的 project 是 region-based image segmentation。

Projects (3/3)

- Region-based image segmentation



- **Reason:**

對此 project 感興趣的原因有三:

1. 可行性高:

如課程中介紹，現今機器學習在 object detection 任務上已有卓越表現，加上 HAC algorithm 等圖片分割演算法層出不窮，此題目在可行性上高於另外兩個主題。

2. 研究價值:

Region-based image segmentation 不只可以用在課程中所提到的疾病偵測，在一般的 object detection 任務如自駕車系統也能有所發揮，讓此題目增添實用研究價值。

3. 興趣:

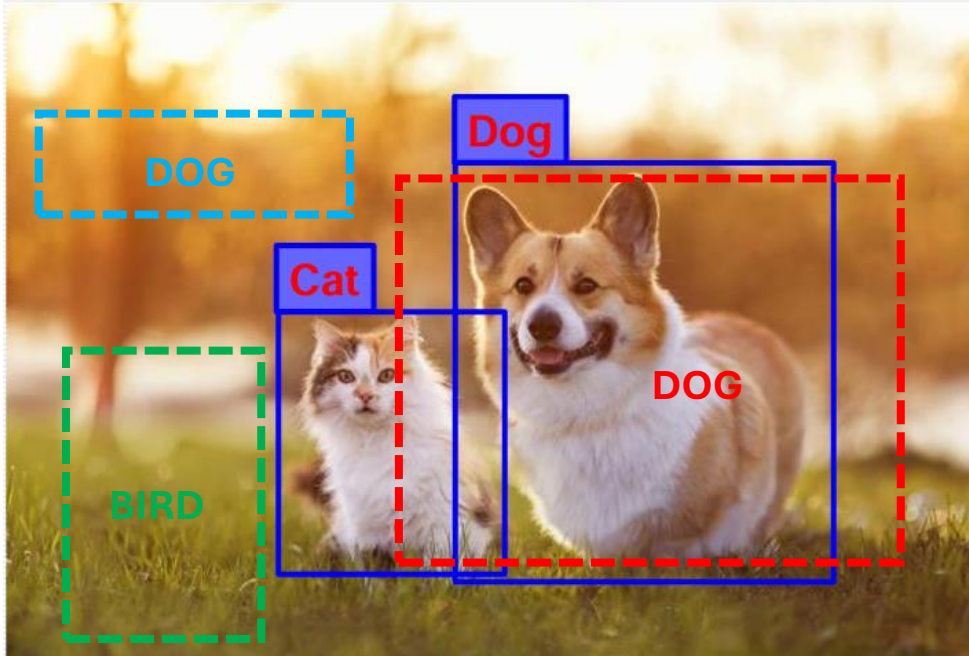
在聽完學長介紹後，非常好奇 HAC algorithm 等 clustering 演算法是如何實現，加上前兩個優點，使此 project 相較其他兩者要引人入勝。

Part II. Questions and Answers

- Q1: Explain the meanings of true positive(TP), False positive(FP), true negative(TN), false negative(FN).

A1:

	True: 模型判斷正確	False: 模型判斷錯誤
Positive: 物件有出現	TP: 預測物在圖中有出現且有正確預測(紅框)。	FP: 預測物不在預測框內(藍框)或模型判斷無出現。
Negative: 物件沒出現	TN: 預測物沒有出現且模型正確判斷無物件。	FN: 預測物沒有出現模型卻錯誤預測有出現(綠框)。



- Q2: What is the correlation between precision and recall? Positive or negative? Why?

A2:

觀查 precision 和 recall 的公式，會發現 precision 著重在 prediction 的品質，給予 FP 處罰；而 recall 主要判斷模型正確提取物件的能力，處罰 FN。

當我們提高模型判斷閾值(threshold)，模型傾向判斷無物件，也就是判定 negative 機率變大，因此，FN 數量會增加，FP 則會減少；反之，若減少閾值，則 FP 會增加、FN 減少。

由此可知，precision 和 recall、positive 和 negative 兩者皆為負向關係。

- Q3: What makes a two-stage model slow? List at least two reasons.

A3:

許多 object detection 任務的 model 皆為 two-stage，第一步為 region proposal，第二步為 region classification and refinement，主要目的及好處如下：

Reason 1: 經過第一步的區域分割，通常第二步驟能在有限的算量下達到較佳的準確度，因為模型只需集中在較小區域的特徵，而不需要處理整張圖片。

Reason 2: 透過分開兩步驟，能增加模型 generalize 能力。先對不同大小的圖片進行處理，再經由模型來判斷，如此來增加模型彈性。

- Q4: If there is a 384x384 image and inferenced by a YOLO model. It has 4 output feature maps, which are stride 4, stride 8, stride 16, and stride 32. It also has 2 different types of anchors. How many bboxes before post-processing?

A4:

首先需計算 4 張 feature map 的 grid:

Stride 4: Image size / stride = $384 / 4 = 96 \times 96$

Stride 8: Image size / stride = $384 / 8 = 48 \times 48$

Stride 16: Image size / stride = $384 / 16 = 24 \times 24$

Stride 32: Image size / stride = $384 / 32 = 12 \times 12$

再來將 anchor 放入每個 grid 中，也就是 grid*anchor:

Stride 4: $96 * 96 * 2 \text{ anchors} = 18432 \text{ bboxes}$

Stride 8: $48 * 48 * 2 \text{ anchors} = 4608 \text{ bboxes}$

Stride 16: $24 * 24 * 2 \text{ anchors} = 1152 \text{ bboxes}$

Stride 32: $12 * 12 * 2 \text{ anchors} = 288 \text{ bboxes}$

Number of bbox = $18432 + 4608 + 1152 + 288 = 24480 \text{ bboxes}$

- Q5: There are two images name img1 and img2. img1 has a target (x1, y1, w1, h1, cls[0, 0, 1]), and img2 has a target (x2, y2, w2, h2, cls[0, 1, 0]). How to mixup them to a new image img3? And what “are” the targets of img3?

A5:

MixUp 會將兩張圖片透明度設為一半並疊合，target 混合方式參考以下公式，若 $\lambda=0.5$ ，可以得到

```
target_img3 =  
(  
    0.5*x1 + 0.5*x2,  
    0.5*y1 + 0.5*y2,  
    0.5*w1 + 0.5*w2,  
    0.5*h1 + 0.5*h2,  
    cls[0, 0.5, 0.5]  
)
```

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j, & \text{where } x_i, x_j \text{ are raw input vectors} \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j, & \text{where } y_i, y_j \text{ are one-hot label encodings}\end{aligned}$$

- Q6: Considering the worst case of NMS, what is its time complexity?

A6:

NMS 最差的時間複雜度是 $O(N^2)$ ，其中 N 為 bbox 數量。Non-maximum suppression 進行步驟如下：

1. 刪除信心程度(confidence)小於閾值的 bbox。
2. 按信心程度對 bbox 進行排序。
3. 從信心程度最高者和其他 bbox 計算 IoU
4. 若兩者 IoU 高於 IoU 閾值，刪除信心程度較低者
5. 換至下一個 bbox，回到步驟 3.

當所有 bbox 的 IoU 並未高於 IoU 閾值，則計算會迭代所有 bbox，因此最差的時間複雜度是 $O(N^2)$ 。

- Q7: A two-stage model usually has better ability to classification than an one-stage. Why?

A7:

原因如 A3 中所提，two-stage model 對於 object detection 任務擁有較高的彈性，能處理不同大小的圖片、物件，且經過第一階段處理後的分割圖片，讓模型分類能著重在小區域特徵上，由此增加模型準確度。