

YOLOv7 based Automation for Parotid Tumor Detection

Author: Bo-Sen, Wang

Outline:

1. Introduction
 - 1.1 Motivation&Objective
2. Tumor Detection
 - 2.1 資料來源以及符號定義
 - 2.2 資料前處理
 - 2.3 YOLOv7 腫瘤偵測
 - 2.4 YOLOv7 訓練數據分析
 - 2.5 Evaluation Metrics
3. Metrics
 - 3.1 Based on slices space
 - 3.2 Based on bounding box space
 - 3.3 Based on tumor space
 - 3.4 Data distribution
 - 3.5 視覺化呈現結果
4. 後續實驗
 - 4.1 單張影像輸入 v.s. 上下文資訊整合輸入

1. Introduction

1.1 Motivation&Objective:

為什麼需要人臉低解析度還原至高解析度?

■ 應用

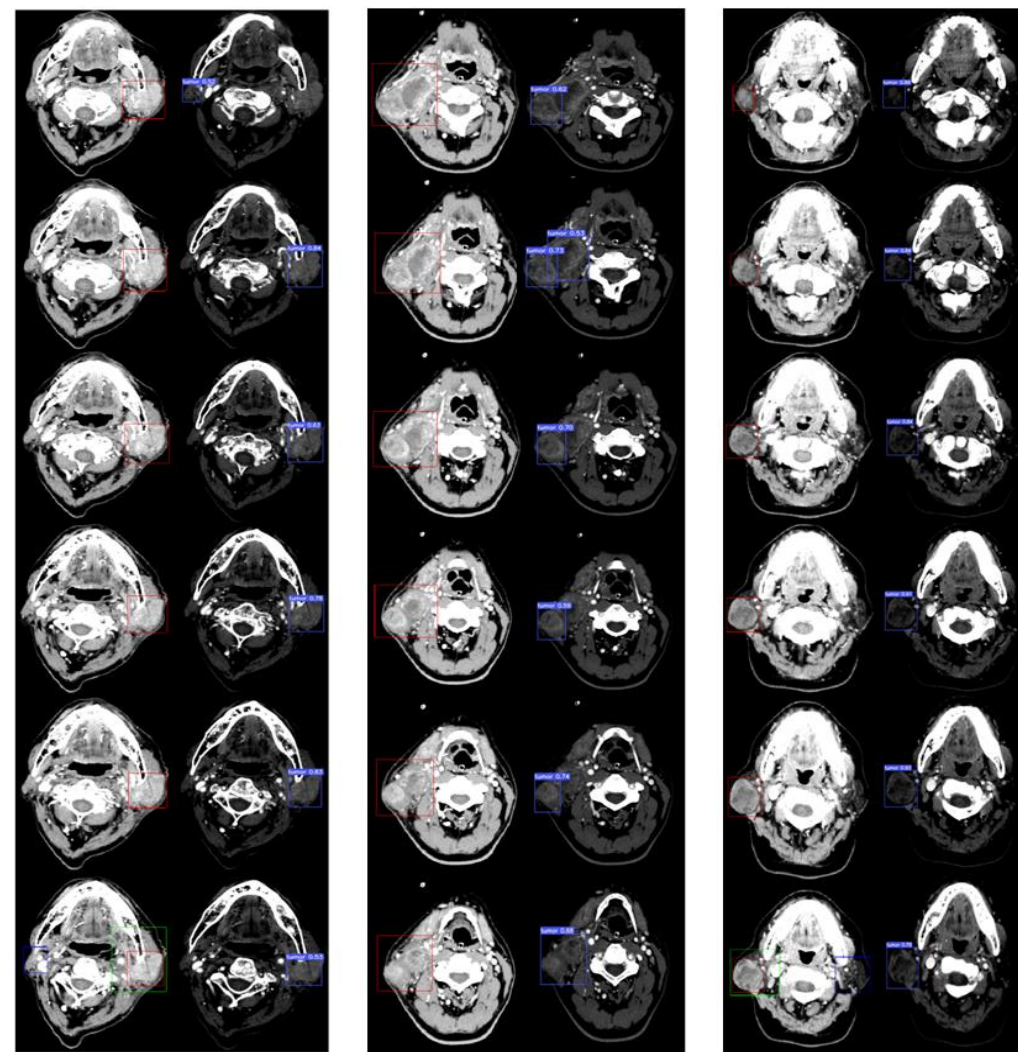
- **減少人工標記所耗的時間:**人工標記腫瘤會耗費醫生許多時間，也可能導致要辨別的影像需要標記使得人工標記錯誤的狀況發生。
- **早期治療以及治療計畫:**AI的興起讓腫瘤影像可以經由人工智慧快速判讀，幫助病人及早治療，以及及早制定治療計畫。

■ 困難

- **資料不足:**人工標記需要耗費許多時間會導致訓練資料難以蒐集。
- **資料多樣性不足:**儘管蒐集很多資料但如果出現較為特異的資料使得病人病情判斷錯誤，這項的代價是巨大的。

■ 研究

- **多方面評估準則:**我們將提供3種不同層面的評估準則以衡量YOLOv7模型的效能，同時這樣的方式可以根據使用者不同的情境用不同的方式應用我們模型，而不局限於一種方式。



腫瘤偵測視覺化示意圖

2. Tumor Detection

2.1 資料來源以及符號定義(1/2)

- 整體資料:

學長整理的excel中挑選編號1~46(9排除)、91~335號病例當作資料集，定義為：

$$DB^{overall} \equiv \{I_{i,j,type} | i = 1, \dots, 300; j = 1, \dots, n; type = P, N\}$$

我們會執行K-fold Validation，因此不同的K會有不同的訓練以及測試資料

定義如下：

- 腮腺腫瘤訓練資料

- $DB^{train} \equiv \{I_{i,j,r} | i = 1 \sim (k-1) * 60 + 1 \text{ and } k * 60 \sim 300; j = 1, \dots, n; t = P, N\}$

- 腮腺腫瘤測試資料

- $DB^{val} \equiv \{I_{i,j,r} | i = (k-1) * 60 + 1, \dots, k * 60; j = 1, \dots, n; t = P, N\}$

- 資料屬性以及符號說明

- i 代表病人編號;
- j 代表第i個病人的第j張slices;
- t代表該張slices有沒有腫瘤;
- P代表有腫瘤
- N代表沒有腫瘤

1				46(CYCN)	22(CYCN)	Ncase(109年)
2	年份	性別	年齡	排序	編號	Type
23	108	女	31	21	Case_021	PMA
24	108	女	25	22	Case_022	PMA
25	107	男	39	23	Case_023	PMA
26	108	男	50	24	Case_024	PMA
27	108	男	31	25	Case_025	PMA
28	108	男	27	26	Case_026	PMA
29	107	女	64	27	Case_027	PMA
30	107	男	62	28	Case_028	Warthin
31	107	男	46	29	Case_029	Warthin
32	108	男	46	30	Case_030	Warthin
33	108	女	49	31	Case_031	Warthin
34	108	男	60	32	Case_032	Warthin
35	108	男	65	33	Case_033	Warthin
36	106	男	66	34	Case_034	Warthin
37	106	男	54	35	Case_035	Warthin
38	108	男	65	36	Case_036	Warthin
39	108	男	71	37	Case_037	Warthin
40	108	男	61	38	Case_038	Warthin
41	107	男	59	39	Case_039	Warthin
42	108	男	64	40	Case_040	Warthin
43	108	男	57	41	Case_041	Warthin
44	108	男	53	42	Case_042	Warthin
45	108	男	63	43	Case_043	Warthin
46	108	男	78	44	Case_044	Warthin
47	108	男	65	45	Case_045	Warthin
48	108	男	61	46	Case_046	Warthin

107	男	61	322	Case_334	Warthin
107	男	46	323	Case_335	Warthin
107	男	41	324	Case_336	Warthin
107	男	53	325	Case_337	Warthin
107	男	58	326	Case_338	Warthin
108	男	39	327	Case_339	Warthin
108	男	62	328	Case_340	Warthin
108	男	47	329	Case_341	Warthin
108	男	68	330	Case_342	Warthin
108	男	76	331	Case_343	Warthin
108	男	47	332	Case_344	Warthin
108	男	60	333	Case_345	Warthin
108	男	55	334	Case_346	Warthin
108	男	84	335	Case_347	Warthin

2.1 資料來源以及符號定義(2/2)

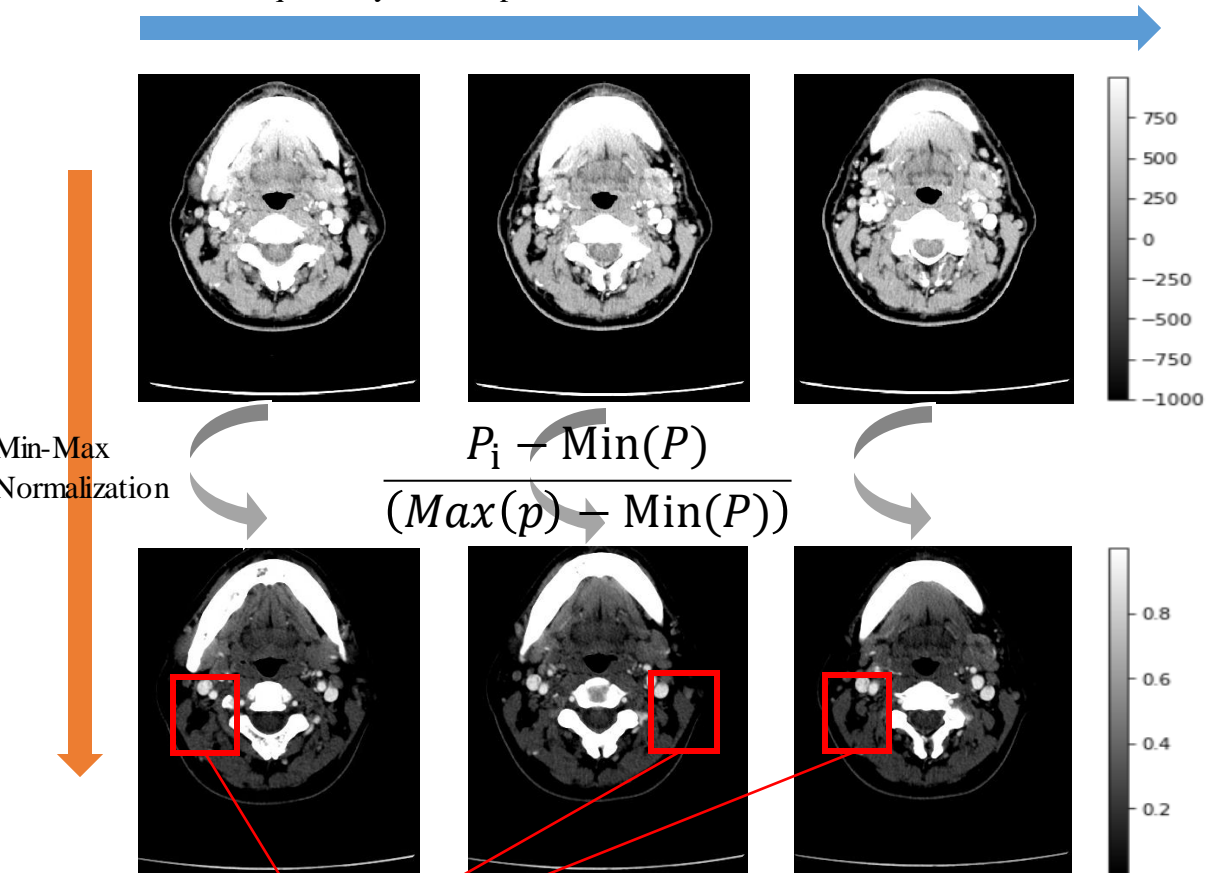
資料數量：

	訓練資料	驗證資料
Cross validation 1	Case: 61~300; 2801 tumor slices	Case: 1~60; 644 tumor slices; 5887 nontumor slices
Cross validation 2	Case: 1~60 and Case 121~300 ; 2769 tumor slices	Case: 61~120; 676 tumor slices; 6684 nontumor slices
Cross validation 3	Case: 1~120 and Case 181~300; 2895 tumor slices	Case: 121~180; 550 tumor slices; 6364 nontumor slices
Cross validation 4	Case: 1~180 and Case 241~300; 2764 tumor slices	Case: 181~240; 681 tumor slices; 6456 nontumor slices
Cross validation 5	Case: 1~240; 2551 tumor slices	Case: 241~300; 894 tumor slices; 6506 nontumor slices

2.2 資料前處理

一般情況下會使用min max normalization來進行前處理

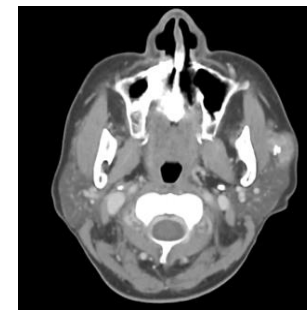
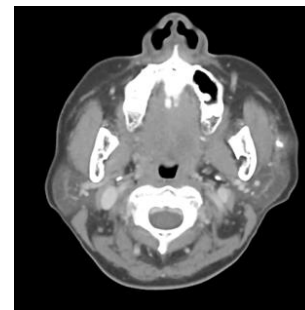
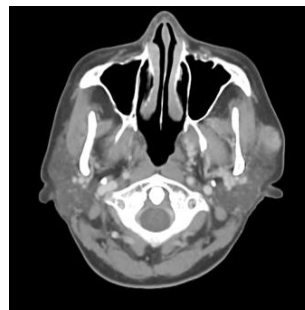
The CT images of the patient's head were acquired sequentially from top to bottom.



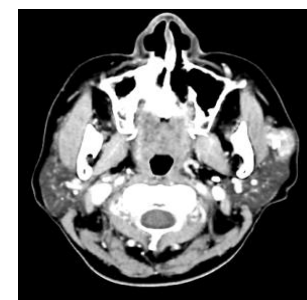
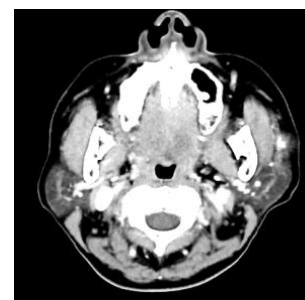
腫瘤位置的地方都會有點黯淡 會有鑑別性不夠的問題

我們使用window width window level normalization

Window level = 20
Window width = 500



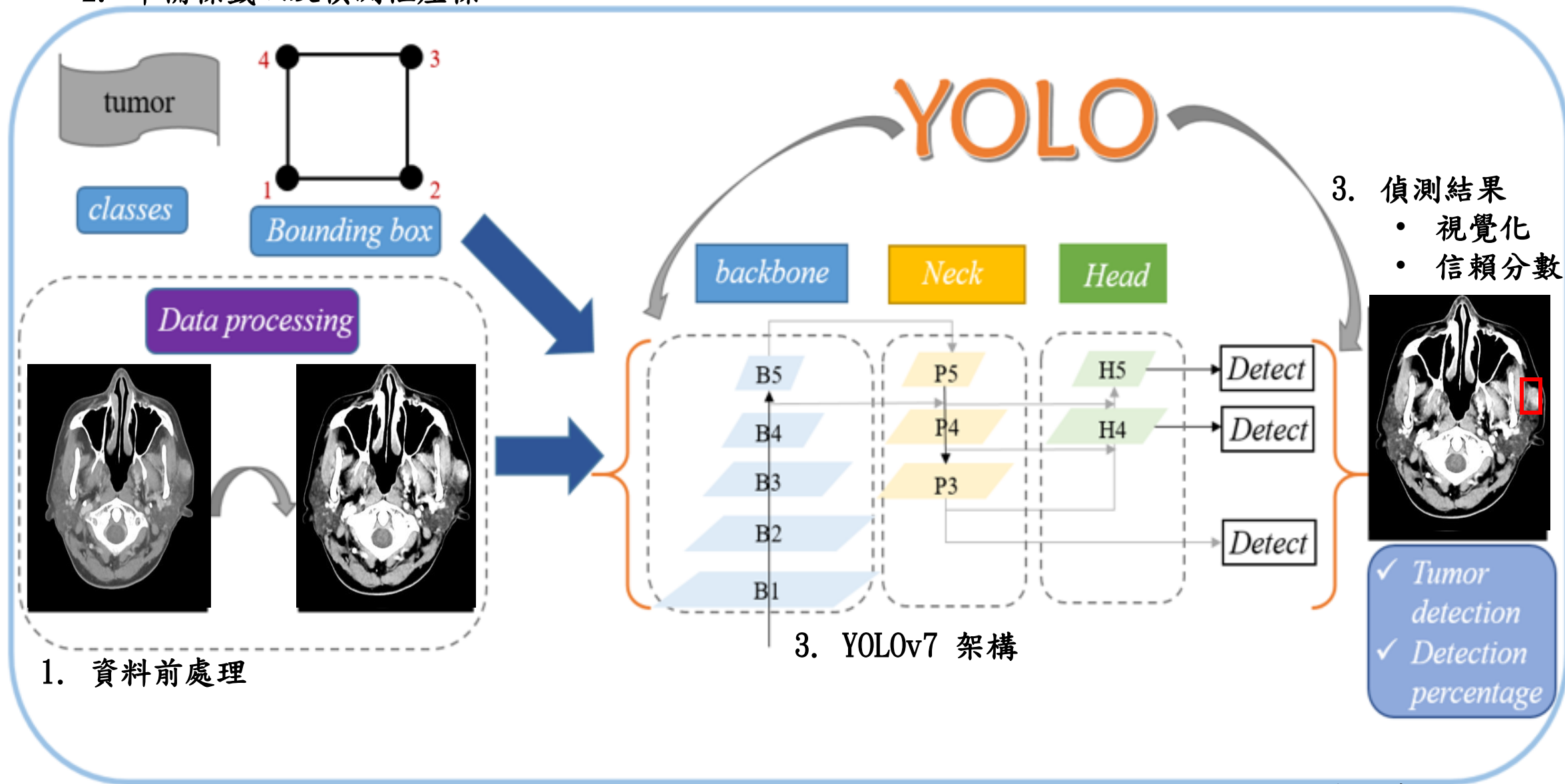
Window level = 20
Window width = 180



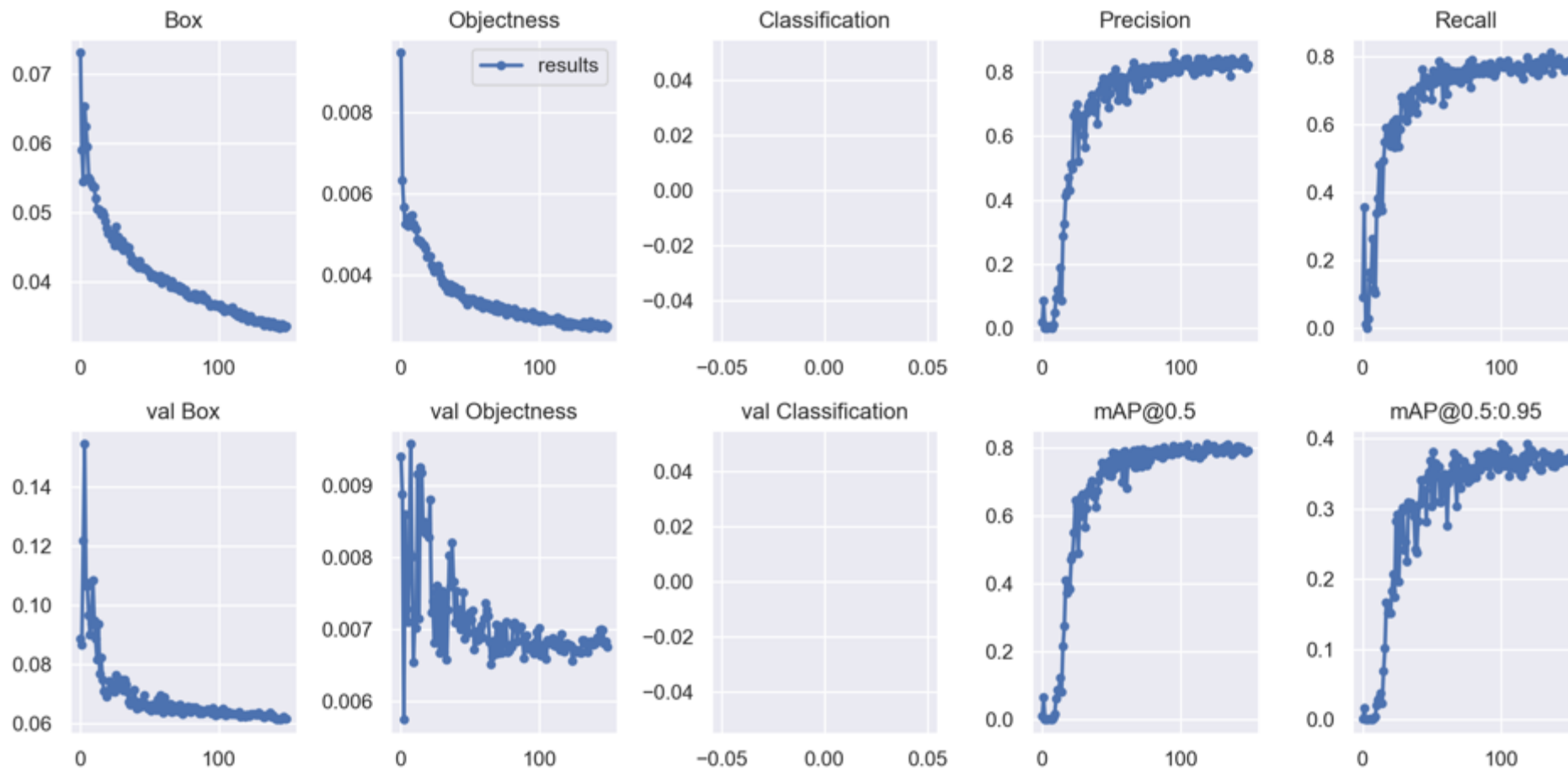
根據醫生的經驗法則我們選擇level=20; width=180
像素值超過 $20 + (180/2) = 110$ ，則設定為1
像素值小於 $20 - (180/2) = -70$ ，則設定為0，
再對整體做normalization.

2.3 YOLOv7 腫瘤偵測-流程

2. 準備標籤以及偵測框座標



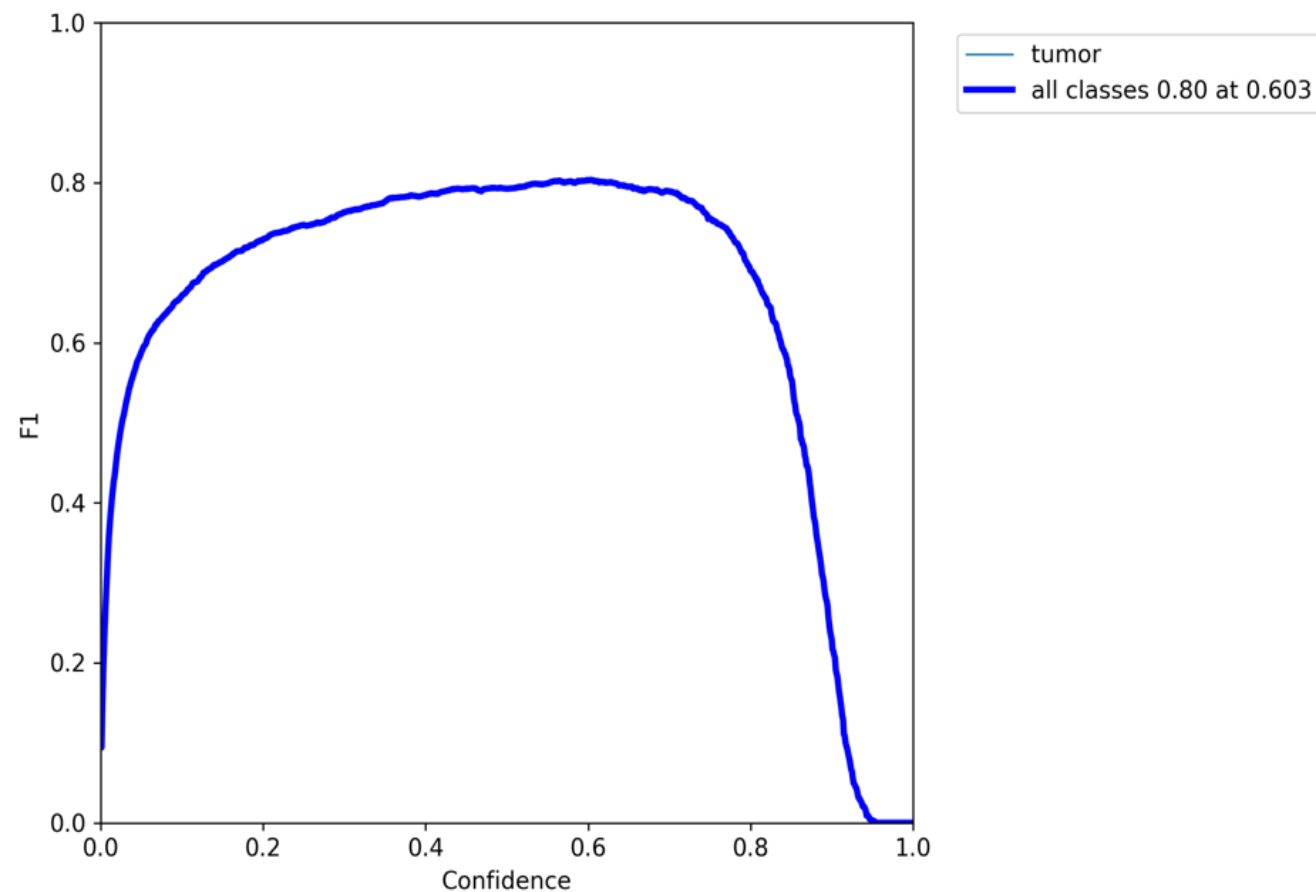
2.4 YOLOv7訓練數據分析(1/4)



- **box:** 回歸損失，針對預測的bounding box以及正確答案的bounding box的差異
- **objectness:** 在 YOLO 的設計中，每個預測框（bounding box）都有一個與之相對應的 objectness 分數，這個分數反映了模型對於該框內是否包含目標物體的信心。
- **classification:** 種類的分類損失，因為我們做的是單分類，所以沒有這個損失。

2.4 YOLOv7訓練數據分析(2/4)

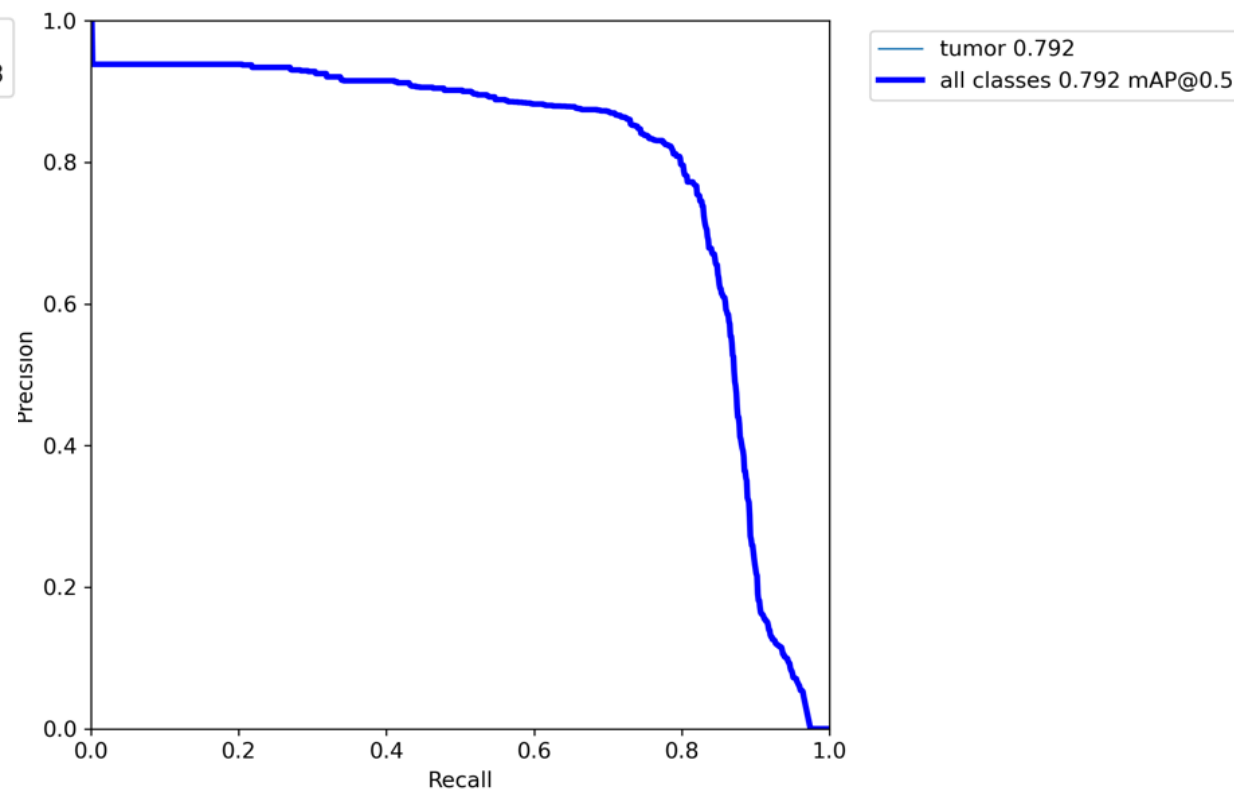
F1 score



不同的 confidence 會有不同的 precision 以及 recall 會有對應的 f1 score，以此類推可以繪製 f1 curve.

F1分數，他被定義為 Precision 以及 Recall 的調和平均數，最大為1，最小為0，以以下圖為例，可以得知在 confidence score = 0.603，則模型在 validation datasets 上有較好的成績。

PR curve



PR 曲線中P代表的precision，R代表的是Recall，他們兩者之間是一個矛盾的性能指標，及一個值越高，另一個值越低;

提高Precision($TP/(TP+FP)$) >> 則FP越小

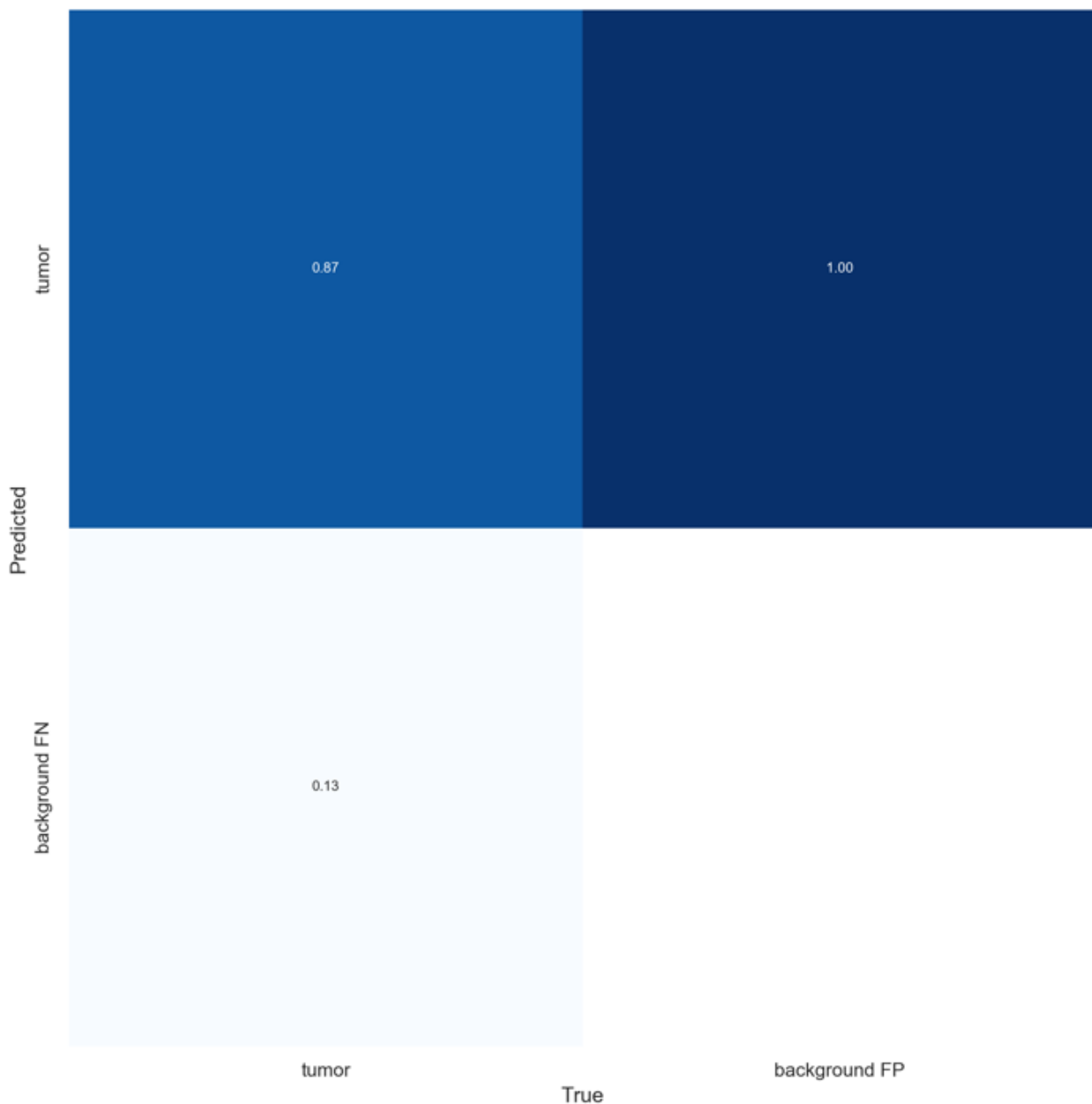
提高Recall($TP/(TP+FN)$) >> 則FN越小

For example:

全部被預測成正樣本>> precision 很低，recall 為100%

全部被預測成負樣本>> recall 很低，precision 為100%

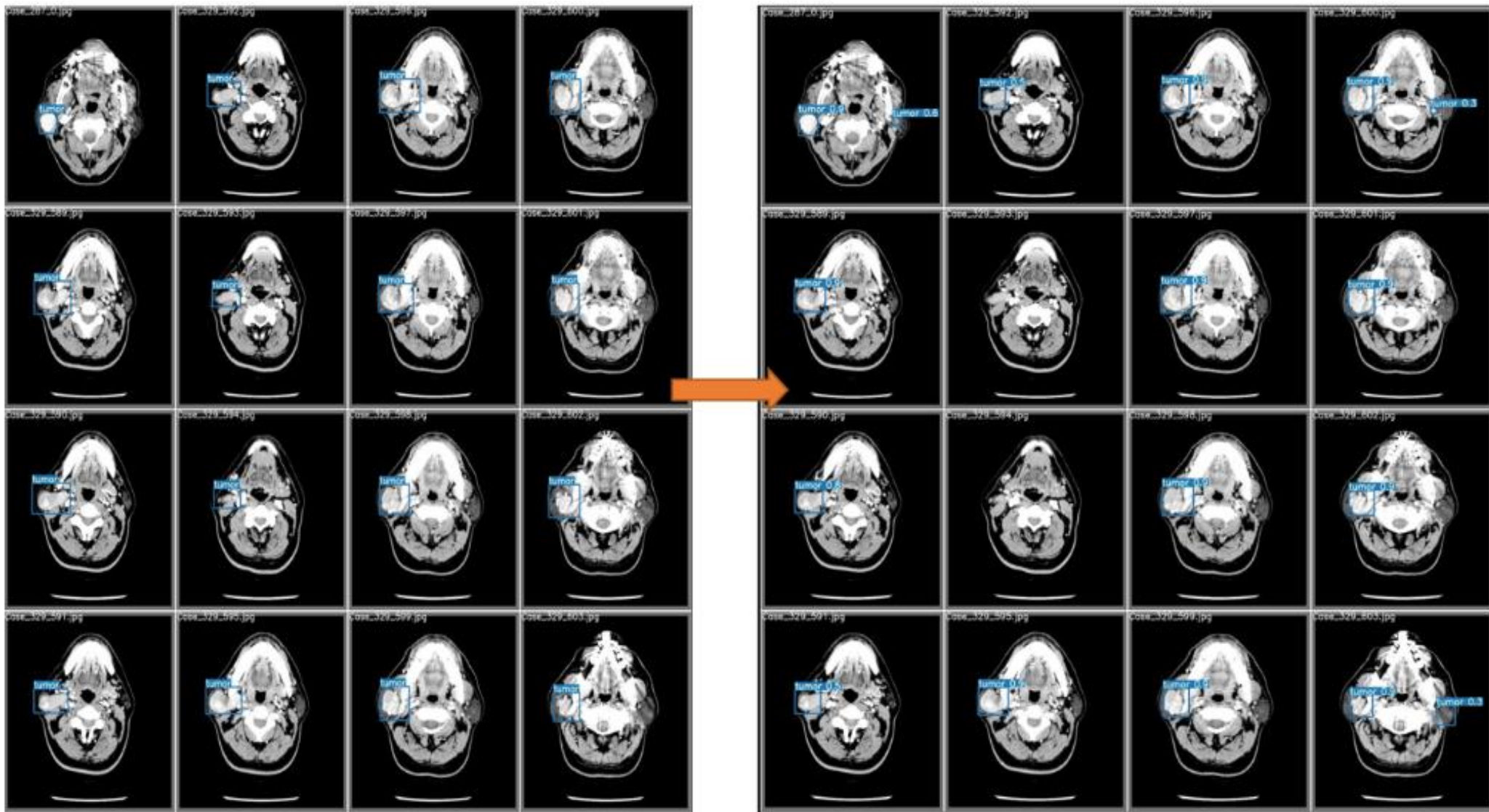
2.4 YOLOv7訓練數據分析(3/4)



- **左上角**代表真實標籤是腫瘤，預測出來也為腫瘤的機率。
- **左下角**代表真實標籤是腫瘤，預測出來為背景（代表背景）的機率。
- **Background FP** 代表被檢測為背景，但實際非背景。
- **Background FN** 代表被檢測為非背景，但實際為腫瘤。
- 根據background FP, FN的定義，**右下角**代表同時被檢測為背景以及非背景，這種可能性為0。
- 只有一個類別的情況下，**右上角**的分母以及分子數量一致，**背景被檢測為有物體的數量=背景被檢測為有腫瘤的數量**。

2.4 YOLOv7訓練數據分析(4/4)

視覺化validation datasets的 label 以及prediction.



2.5 Evaluation Metrics (1/2)

準確率為所有樣本中，**預測正確的比例**，該指標顯示模型的綜合能力。

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (1)$$

敏感度則代表在所有陽性樣本中，**正確預測為陽性的樣本比例**，該指標在醫療領域中十分重要，它代表著漏檢測的風險這在醫療領域上是具有較大的代價，在統計學上我們又稱之為型一誤差(Type I error)。

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

精確度則衡量陽性預測的準確性，通過測量正確預測的陽性樣本佔所有陽性預測的比例來實現。**該指標顯示模型減少偽陽性的能力**。

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

2.5 Evaluation Metrics (2/2)

特異度是指實際為陰性的樣本中，判斷為陰性的比例（例如真正未生病的人中，被醫院判斷為未生病者的比例）。該指標顯示模型減少偽陽性的能力。

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (4)$$

陰性預測值是指在所有陰性樣本中，被正確預測為陰性樣本的比例。

$$\text{NPV} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Negative}} \quad (5)$$

偵測相關的演算法中經常會分別提到**精確率**（precision）和**召回率**（recall），F-score能**同時考慮這兩個數值**，平衡地反映這個演算法的精確度。

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (6)$$

IOU表示了產生的候選框（candidate bounding box）與原始標記框（ground truth bound）的交疊率或說重疊度，也就是它們的交集與並集的比值。

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (7)$$

3. Metrics

3.1 Based on slices space-Definition

- 基於切片空間上評估 YOLOv7 模型效能：

Definition:

■ TP:

- Slices 有腫瘤，有框腫瘤。  問題：不論腫瘤位置，只要有偵測到就會被計算 True Positive.

■ TN:

- Slices 無腫瘤，無框腫瘤。

■ FP:

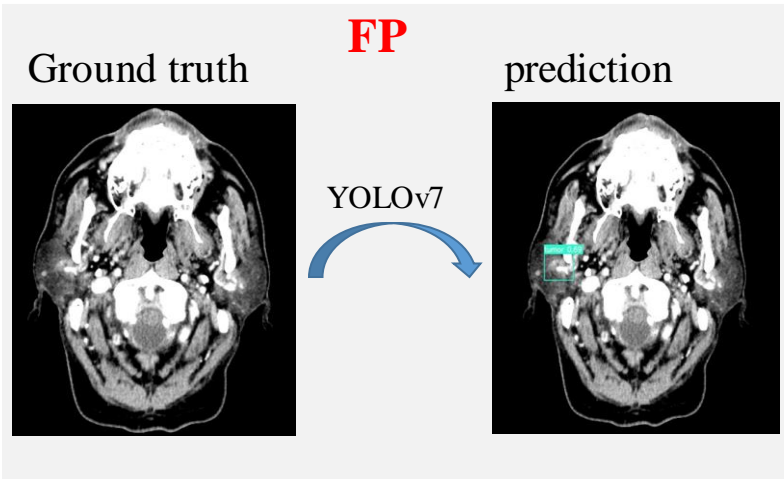
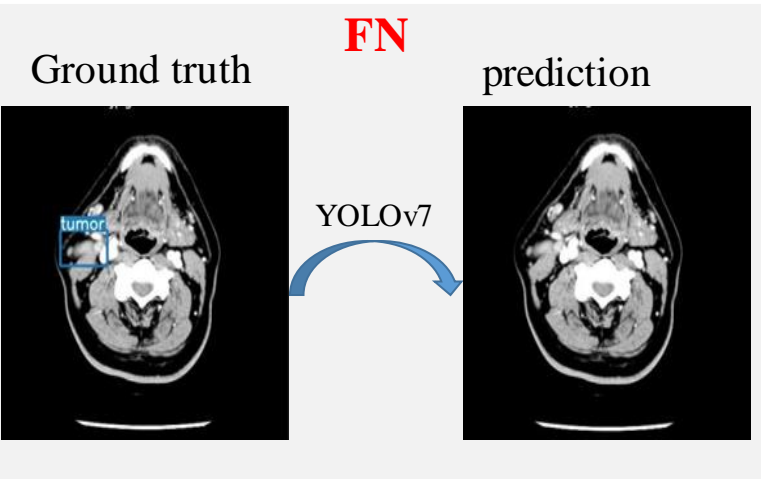
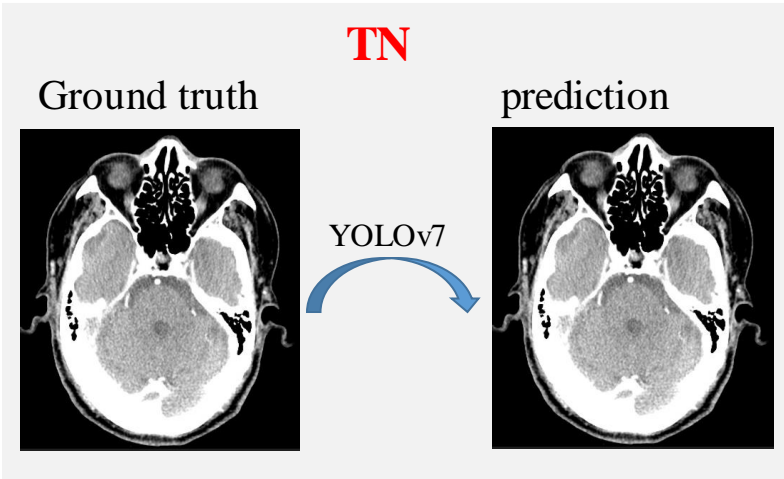
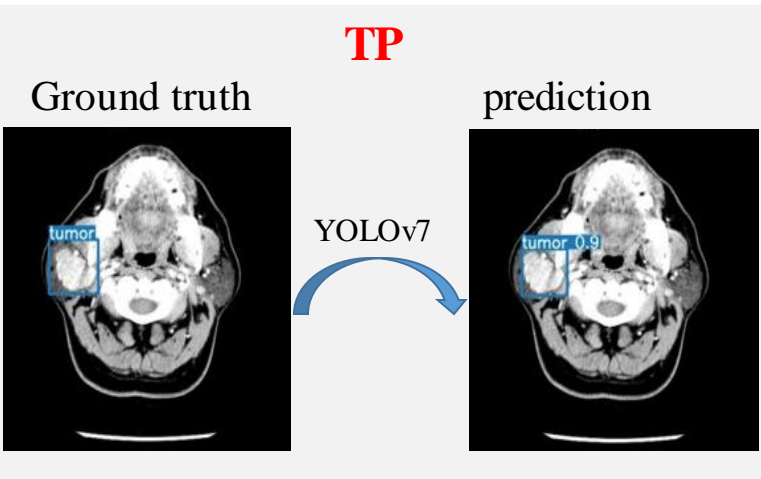
- Slices 無腫瘤，有框腫瘤。

■ FN:

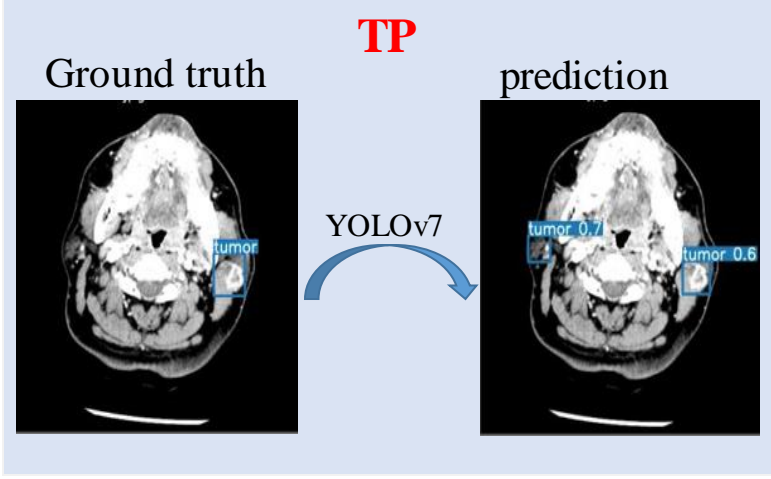
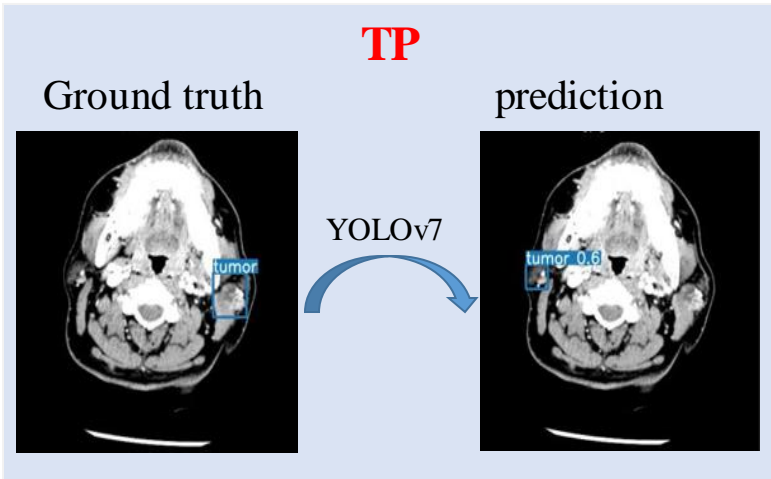
- Slices 有腫瘤，無框出腫瘤。

3.1 Based on slices space-視覺化示意圖

◆ (a) In various scenarios based on slices below.



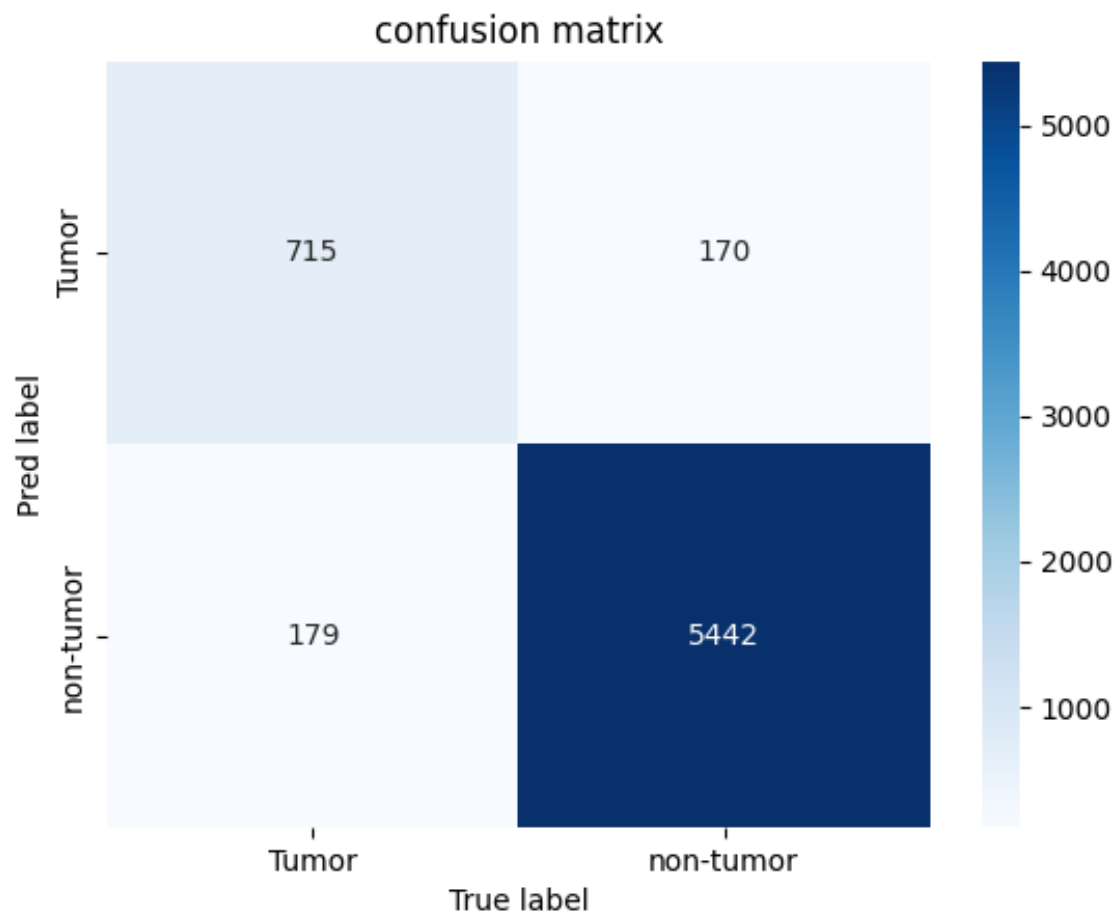
◆ (b) In peculiar scenarios.



3.1 Based on slices space-量化

資料來源:

$$DB^{val} \equiv \{I_{i,j,r} | i = 240 \sim 300; j = 1, \dots, n; t = P, N\}$$



■ TP:

- Slices 有腫瘤，有框腫瘤，共715 slices.

■ TN:

- Slices 無腫瘤，無框腫瘤，共5442 slices.

■ FP:

- Slices 無腫瘤，有框腫瘤，共170 slices.

■ FN:

- Slices 有腫瘤，無框出腫瘤，共179 slices.

Accuracy : 0.946

Recall : 0.799

Precision : 0.807

Specificity : 0.969

NPV : 0.968

3.2 Based on boundingbox space

- 基於偵測框空間上評估 YOLOv7 模型效能：

Definition:

■ TP:

- Slices 有腫瘤，偵測框框出腫瘤 $iou > threshold$ 。

■ TN:

- Slices 無腫瘤，無框出腫瘤。

■ FP:

- Slices 無腫瘤，偵測框框出腫瘤。
- Slices 有腫瘤，框出的偵測框 $IoU < threshold$ 。(物理意義為預測為腫瘤但實際沒有腫瘤。)
For example: 一張圖像有一個腫瘤，框出兩個偵測框，兩個偵測框的 IoU 都小於 $threshold$ ，則計算兩次 FP。

■ FN:


- Slices 有腫瘤，無框出腫瘤。
- Slices 有腫瘤，框出的偵測框 $IoU < threshold$ 。(物理意義為實際有腫瘤但沒有預測出來。)


紅色為重複計算，後續會針對有無重複計算做不同的評估

3.2 Based on boundingbox space-視覺化示意圖

- Illustrative Figure

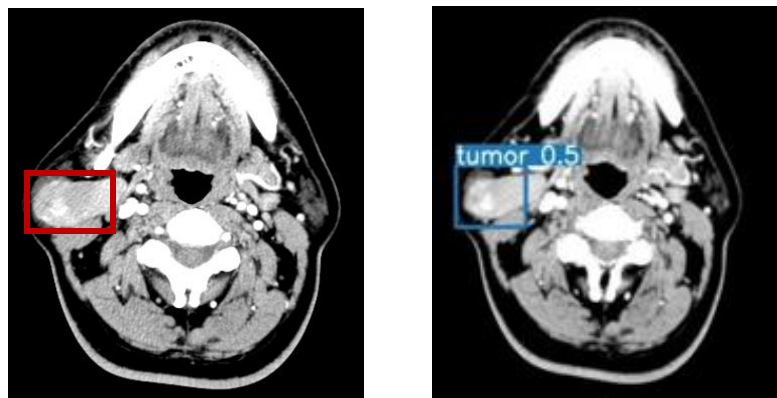


 Ground truth bounding box

 Prediction bounding box

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of union}} = \frac{A \cap B}{A \cup B}$$

Case1



$$IoU = 0.91$$

$IoU > \text{threshold}$

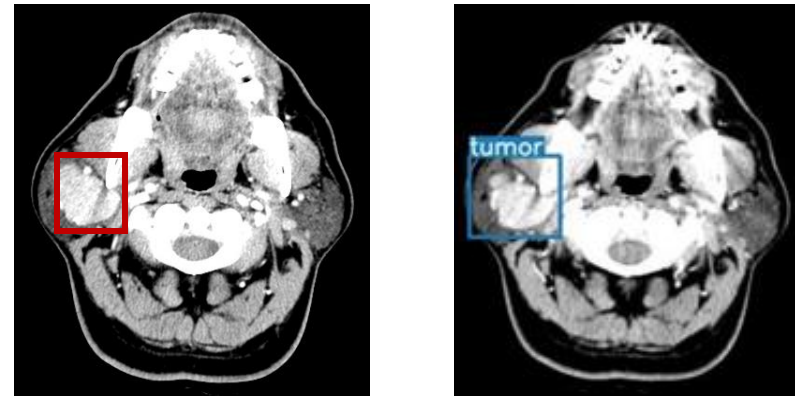
TP

$IoU < \text{threshold}$

FP & FN

or
FN

Case2



$$IoU = 0.735$$

$IoU > \text{threshold}$

TP

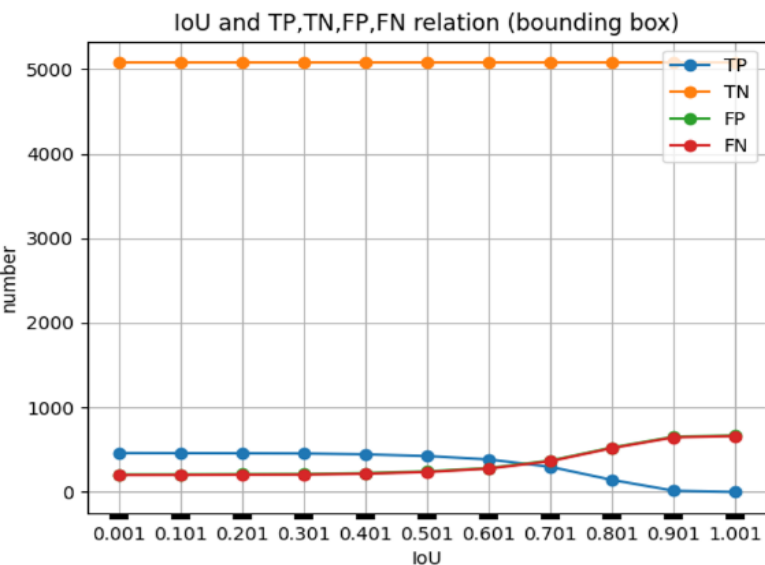
$IoU < \text{threshold}$

FP & FN

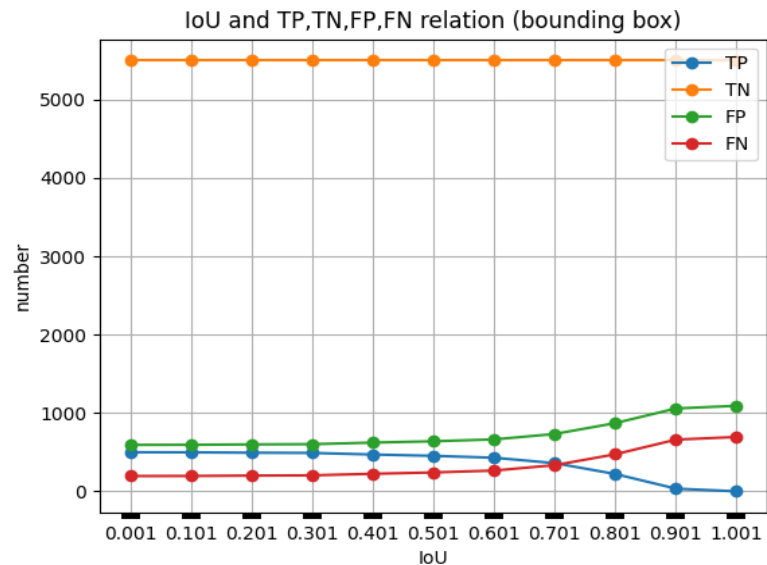
or
FN

3.2 Based on boundingbox space- IoU 以及各項指標關聯性(有重複計算FP)

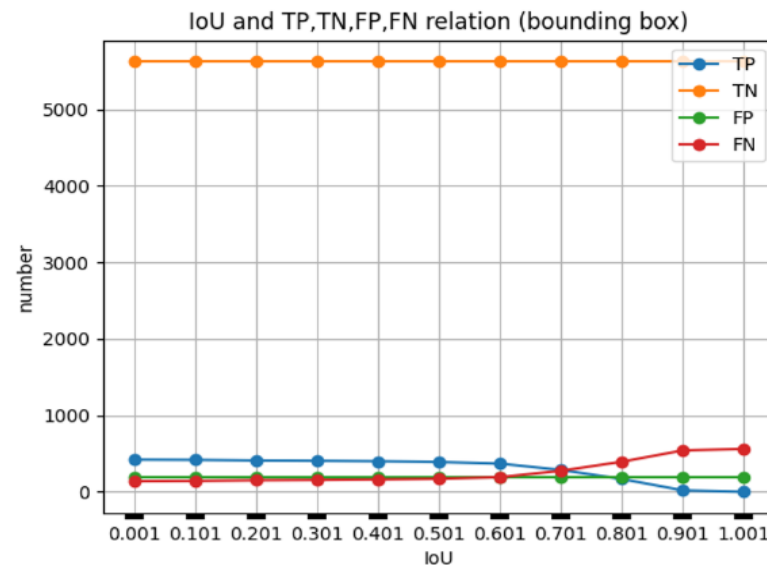
Cross validation 1



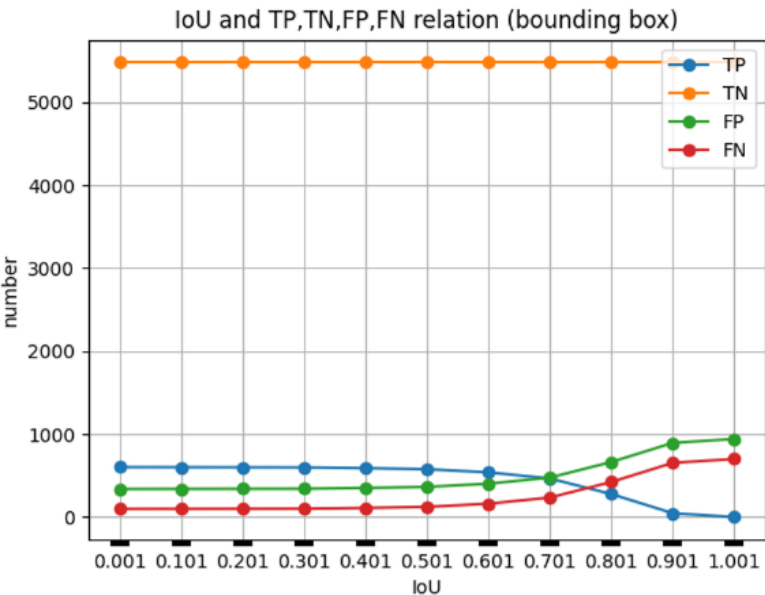
Cross validation 2



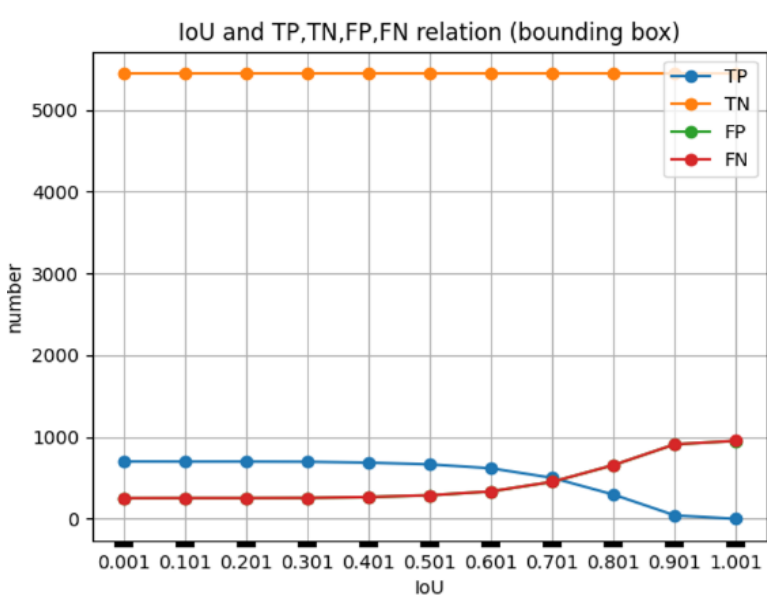
Cross validation 3



Cross validation 4



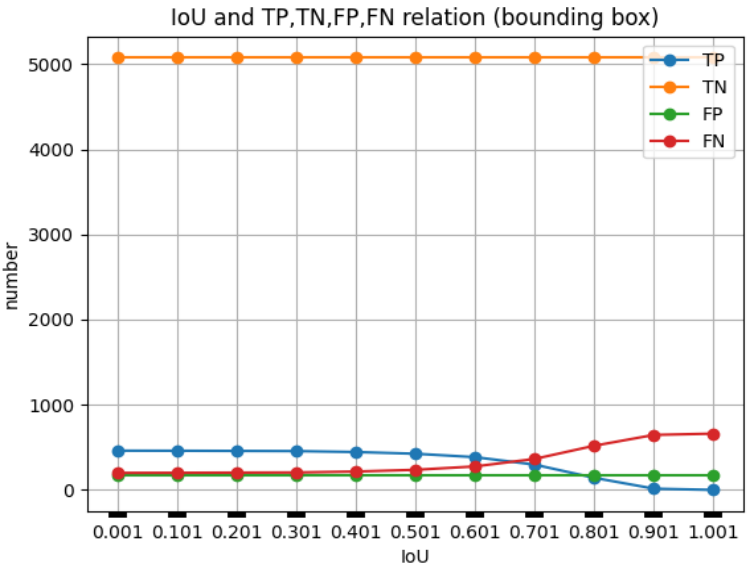
Cross validation 5



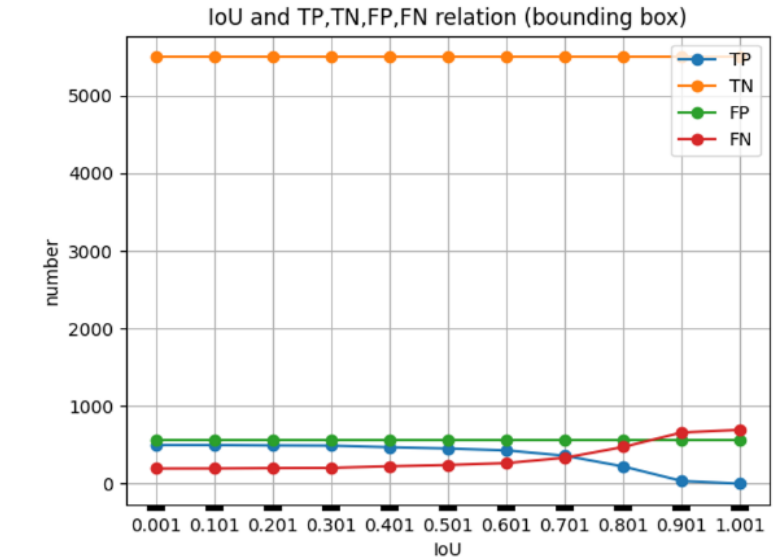
X 軸 代表IoU的變化量
Y 軸 代表指標的數據

3.2 Based on boundingbox space- IoU 以及各項指標關聯性(無重複計算FP)

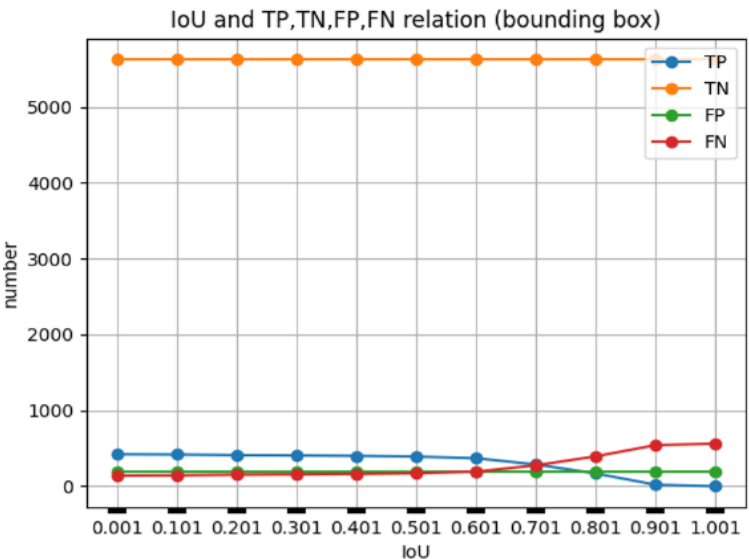
Cross validation 1



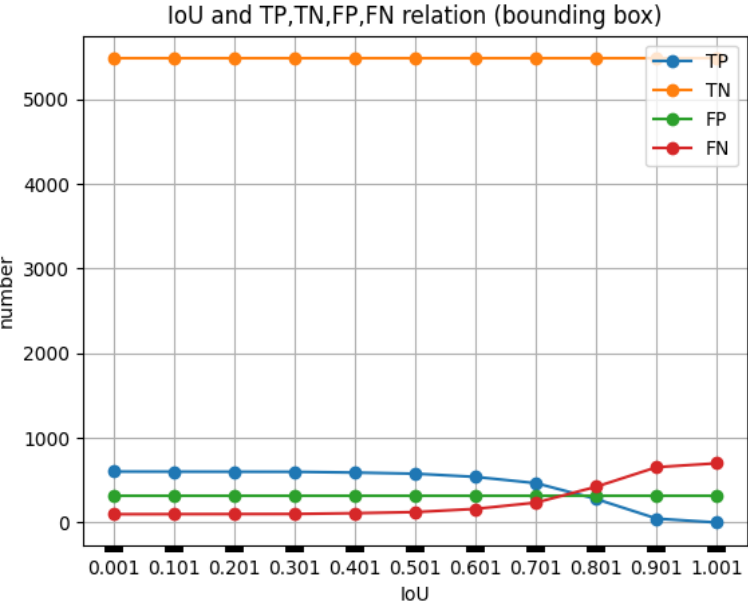
Cross validation 2



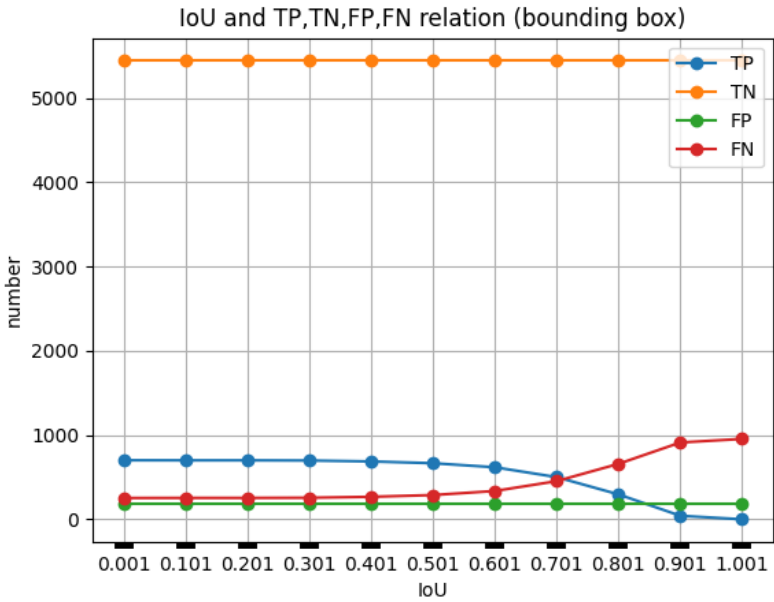
Cross validation 3



Cross validation 4



Cross validation 5



X 軸 代表IoU的變化量
Y 軸 代表指標的數據

3.2 Based on boundingbox space-量化數據(有重複計算FP)

- 前提:
- 1. IoU threshold = 0.001
 - 2. Cross validation 取平均

Cross validation based on bounding box					
評估準則	Cross number = 1	Cross number = 2	Cross number = 3	Cross number = 4	Cross number = 5
(TP/TN/FP/FN)	(461/5074/207/200)	(499/5499/593/195)	(421/5625/211/140)	(602/5480/339/98)	(701/5442/252/253)
Accuracy	0.931	0.883	0.945	0.932	0.924
Recall(Sensitivity;TPR)	0.697	0.719	0.75	0.86	0.734
Precision	0.69	0.456	0.666	0.639	0.735
Specificity	0.96	0.902	0.963	0.941	0.955
NPV	0.962	0.965	0.975	0.982	0.955

3.2 Based on boundingbox space-量化數據(無重複計算FP)

前提:

- 1. IoU threshold = 0.001
- 2. Cross validation 取平均

Cross validation based on bounding box					
評估準則	Cross number = 1	Cross number = 2	Cross number = 3	Cross number = 4	Cross number = 5
(TP/TN/FP/FN)	(461/5074/176/200)	(499/5499/568/195)	(421/5625/193/140)	(602/5480/314/98)	(701/5442/178/253)
Accuracy	0.936	0.887	0.947	0.936	0.934
Recall(Sensitivity;TPR)	0.697	0.719	0.75	0.86	0.734
Precision	0.723	0.467	0.685	0.657	0.797
Specificity	0.966	0.906	0.966	0.945	0.968
NPV	0.962	0.965	0.975	0.982	0.955

3.3 Based on tumor space

- 基於腫瘤空間上評估 YOLOv7模型效能:

- **TP:**

- 病人中有腫瘤部分，連續n張有預測到腫瘤且IoU>threshold.

- **TN:**

- 病人中沒有腫瘤部分，連續n張沒有預測到腫瘤.

- **FP:**

- 病人中沒有腫瘤的部分，連續n張預測到腫瘤.

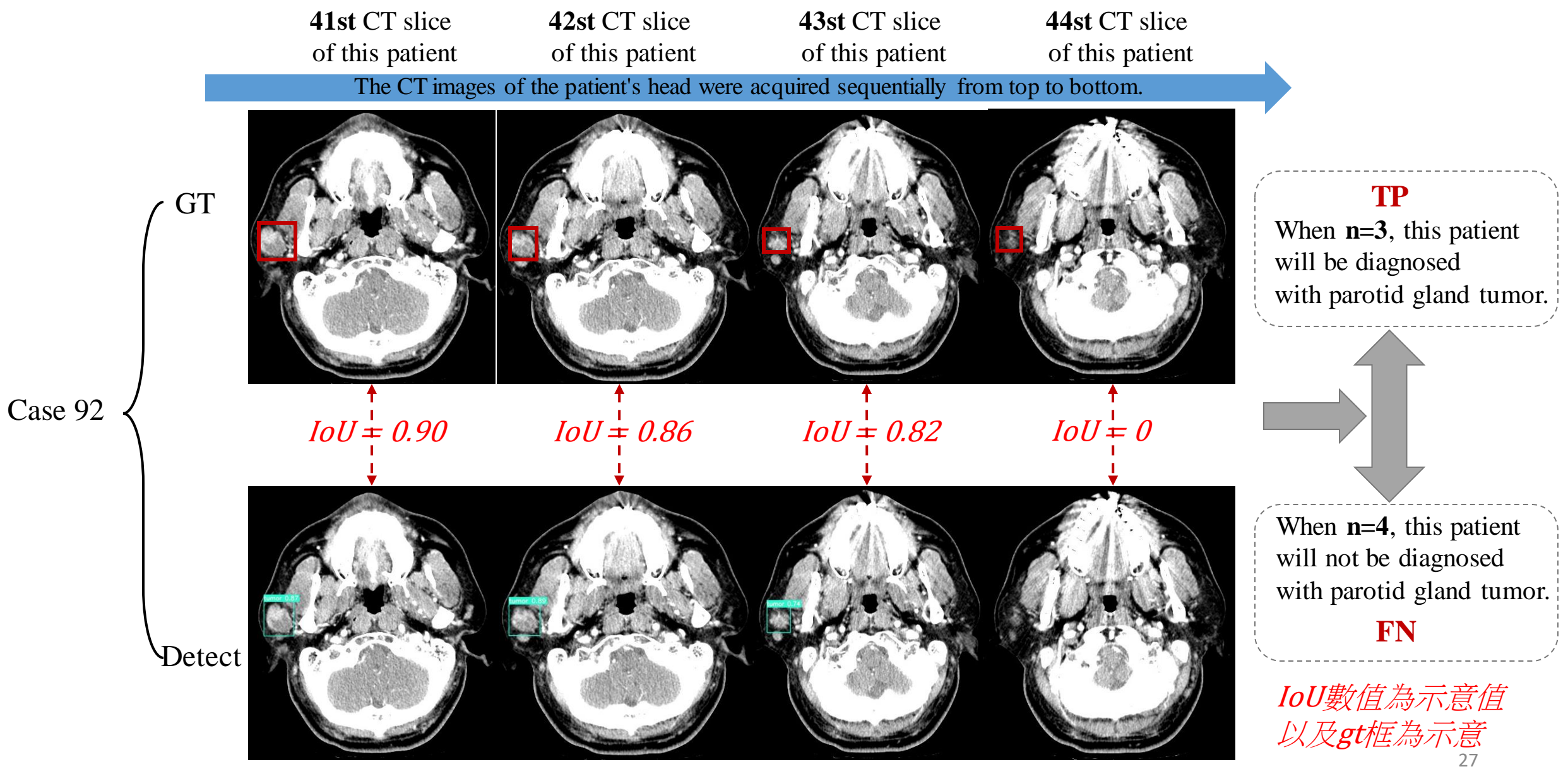
- **FN:**

- 病人中有腫瘤部分，沒有連續n張有預測到腫瘤.

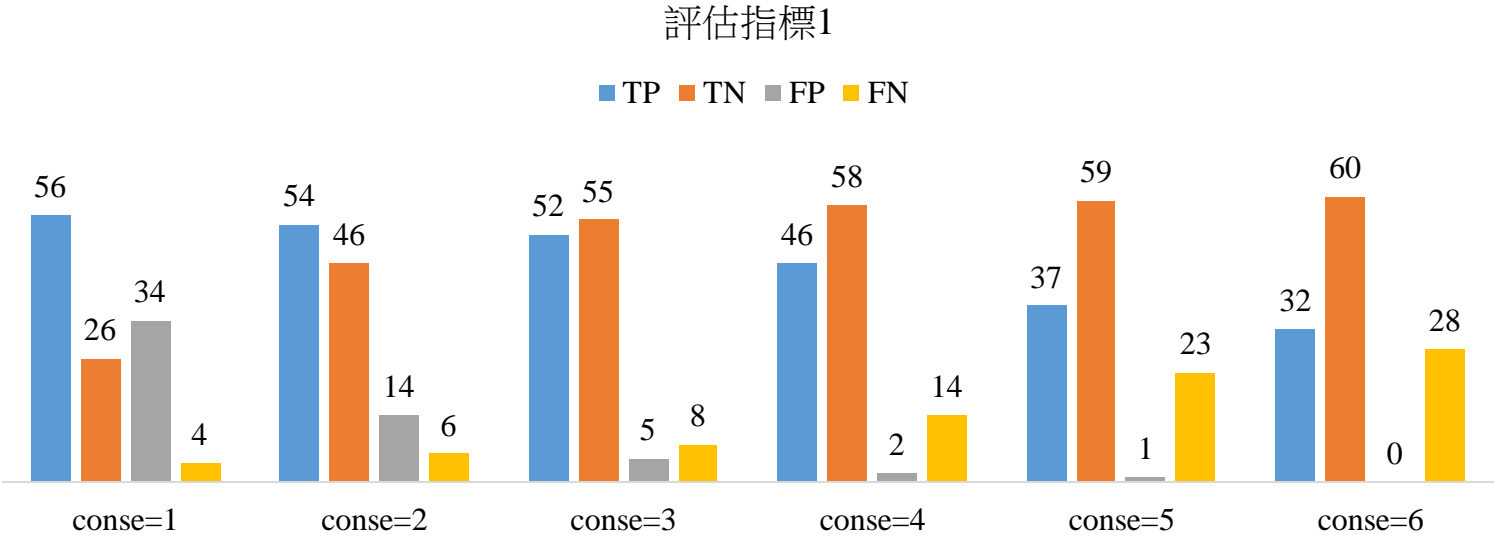
資料後處理:

會將靠近頭部以及頸部的slices 先行篩選掉，因為此部分絕對不會出現腫瘤。

3.3 Based on tumor space-視覺化示意圖



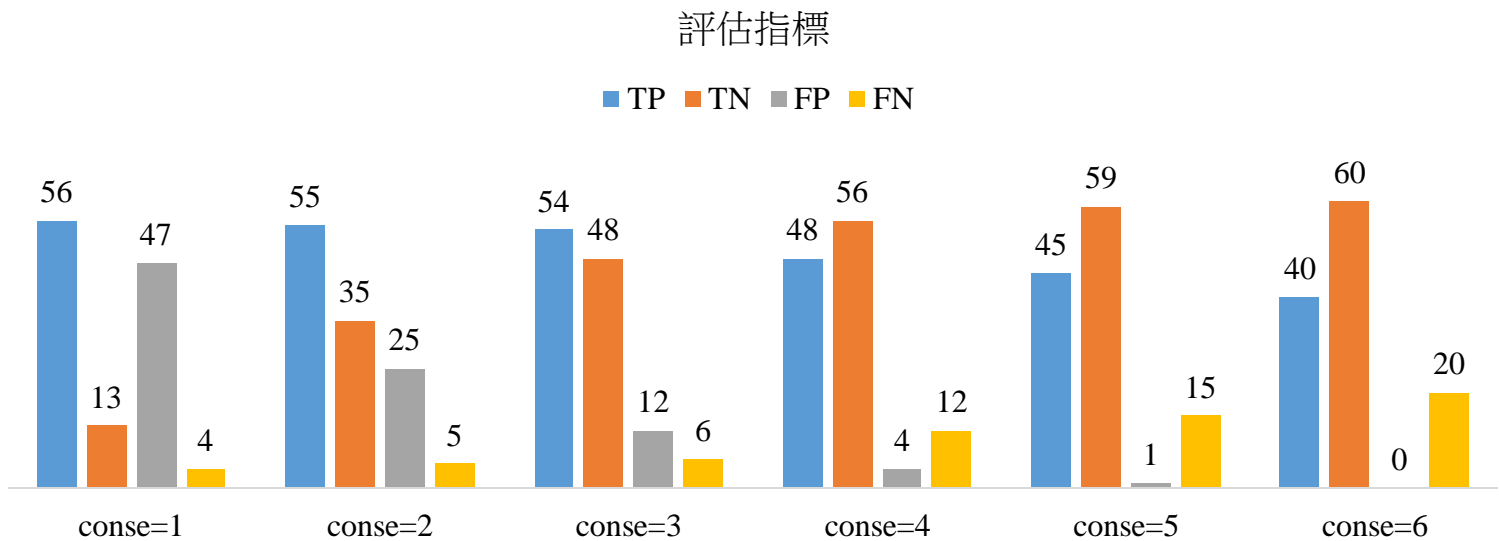
3.3 Based on tumor space-量化數據 (cross validation = 1)



(b) With bounding box optimizers(Cross validation 1)

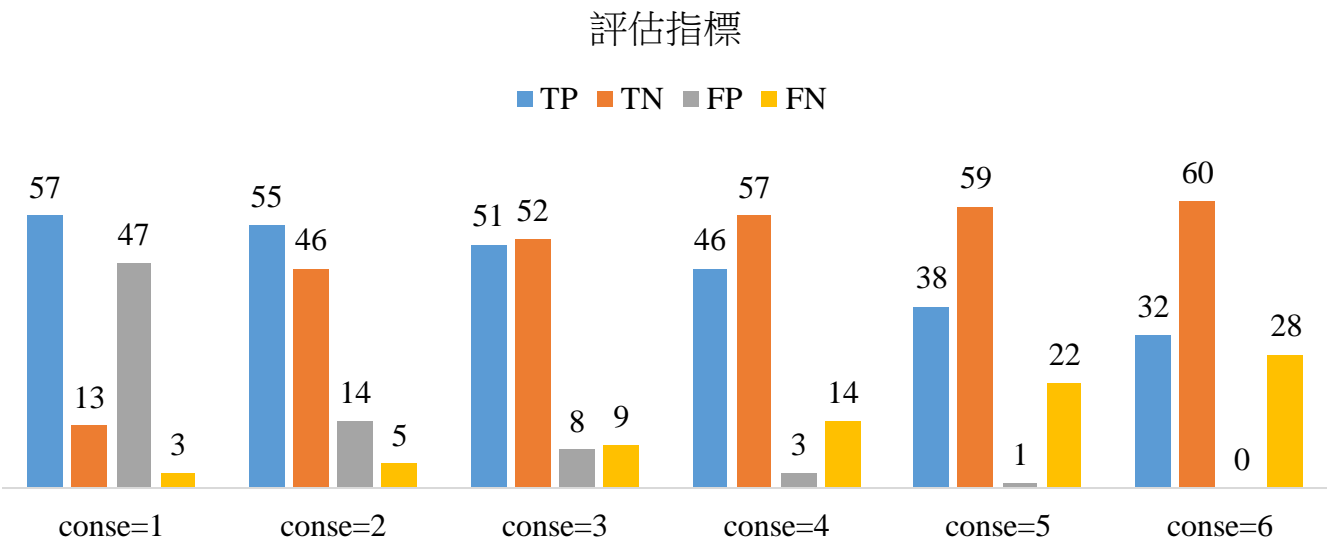
評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
Accuracy	0.683	0.833	0.891	0.866	0.8	0.766
Recall(Sensitivity;TPR)	0.933	0.9	0.866	0.766	0.616	0.533
Precision	0.622	0.794	0.912	0.958	0.973	1
Specificity	0.433	0.766	0.916	0.966	0.983	1
NPV	0.866	0.884	0.873	0.802	0.719	0.681

3.3 Based on tumor space-量化數據(cross validation = 2)



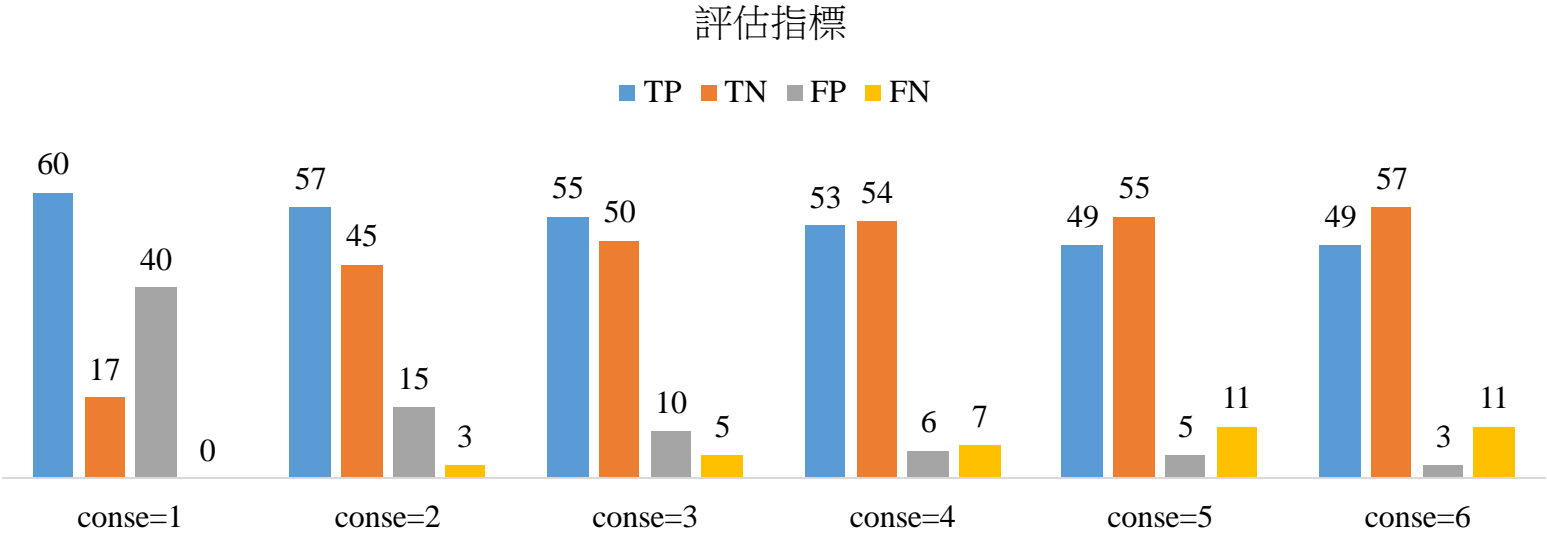
(b) With bounding box optimizers(Cross validation 2)						
評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
Accuracy	0.575	0.75	0.85	0.866	0.866	0.833
Recall(Sensitivity;TPR)	0.933	0.916	0.9	0.8	0.75	0.666
Precision	0.543	0.687	0.818	0.923	0.978	1
Specificity	0.216	0.583	0.8	0.933	0.983	1
NPV	0.764	0.875	0.888	0.823	0.797	0.75

3.3 Based on tumor space-量化數據(cross validation = 3)



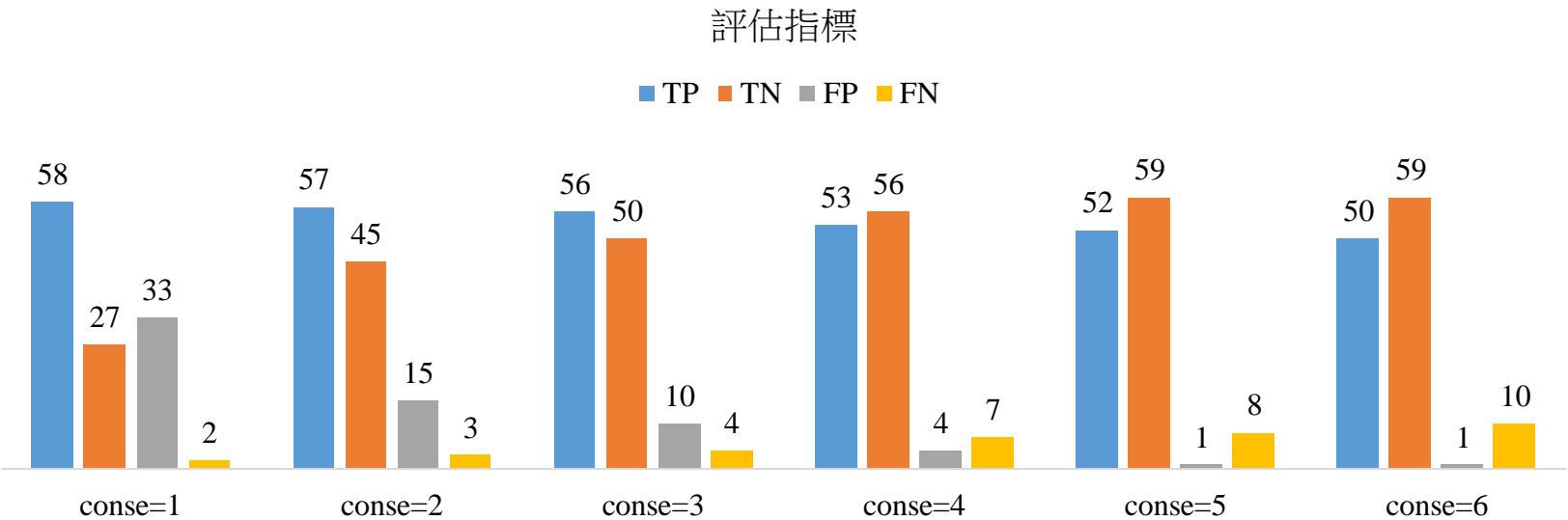
(b) With bounding box optimizers(Cross validation 3)						
評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
Accuracy	0.583	0.841	0.858	0.858	0.808	0.766
Recall(Sensitivity;TPR)	0.95	0.916	0.85	0.766	0.633	0.533
Precision	0.548	0.797	0.864	0.938	0.974	1
Specificity	0.216	0.766	0.866	0.95	0.983	1
NPV	0.812	0.901	0.852/	0.802	0.728	0.68

3.3 Based on tumor space-量化數據(cross validation = 4)



(b) With bounding box optimizers(Cross validation 4)						
評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
Accuracy	0.641	0.85	0.875	0.891	0.866	0.883
Recall(Sensitivity; TPR)	1	0.95	0.916	0.883	0.816	0.816
Precision	0.582	0.791	0.846	0.898	0.907	0.942
Specificity	0.283	0.75	0.833	0.9	0.916	0.95
NPV	1	0.937	0.909	0.885	0.833	0.838

3.3 Based on tumor space-量化數據(cross validation = 5)



(b) With bounding box optimizers(Cross validation 5)

評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
Accuracy	0.708	0.85	0.883	0.908	0.925	0.908
Recall(Sensitivity; TPR)	0.966	0.95	0.933	0.883	0.866	0.833
Precision	0.637	0.791	0.848	0.928	0.981	0.98
Specificity	0.45	0.75	0.833	0.933	0.983	0.983
NPV	0.931	0.937	0.925	0.888	0.880	0.855

3.3 Based on tumor space-量化數據(Mean)

Cross validation Metrics						
評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
(TP/TN/FP/FN)	(57.4/19.2/40.8/2.6)	(55.6/43.4/16.6/4.4)	(53.6/51/9/6.4)	(49.2/56.2/3.8/10.8)	(44.2/58.2/1.8/15.8)	(40.6/59.2/0.8/19.4)
Accuracy	0.638	0.825	0.876	0.878	0.853	0.831
Recall(Sensitivity; TPR)	0.956	0.926	0.893	0.82	0.736	0.676
Precision	0.584	0.77	0.856	0.928	0.96	0.98
Specificity	0.32	0.723	0.85	0.936	0.97	0.986
NPV	0.88	0.907	0.888	0.838	0.786	0.753

3.4 Data distribution(1/3)

Data distribution

- Size of Tumor(mm): 病人的腫瘤大小
- Number of Case: 病人數量
- 圖片語意: 統計病人數量與腫瘤大小的關聯性，以方便統計Based on tumor space 設計的連續張數。

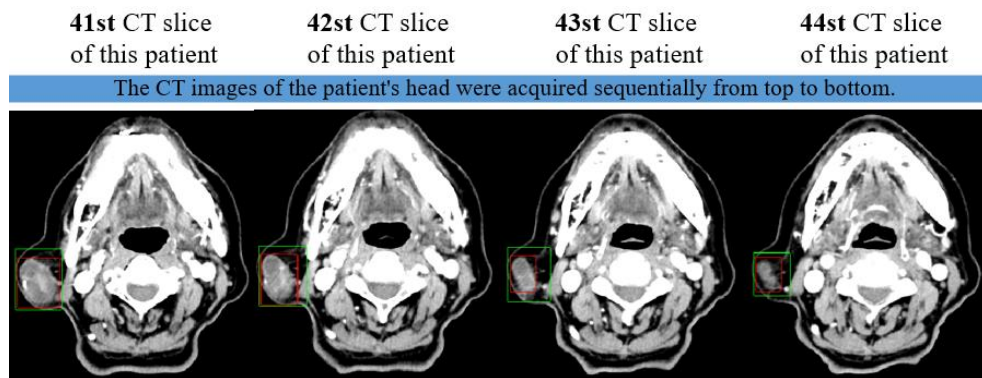
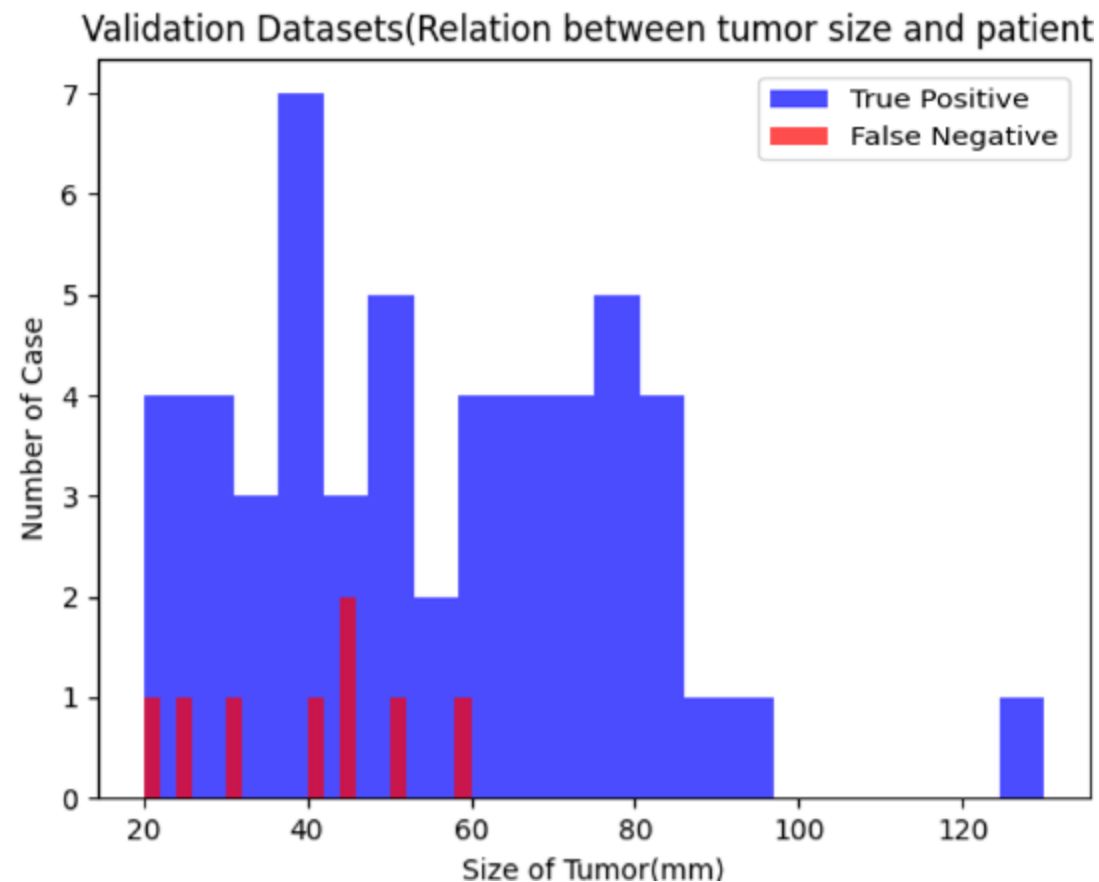


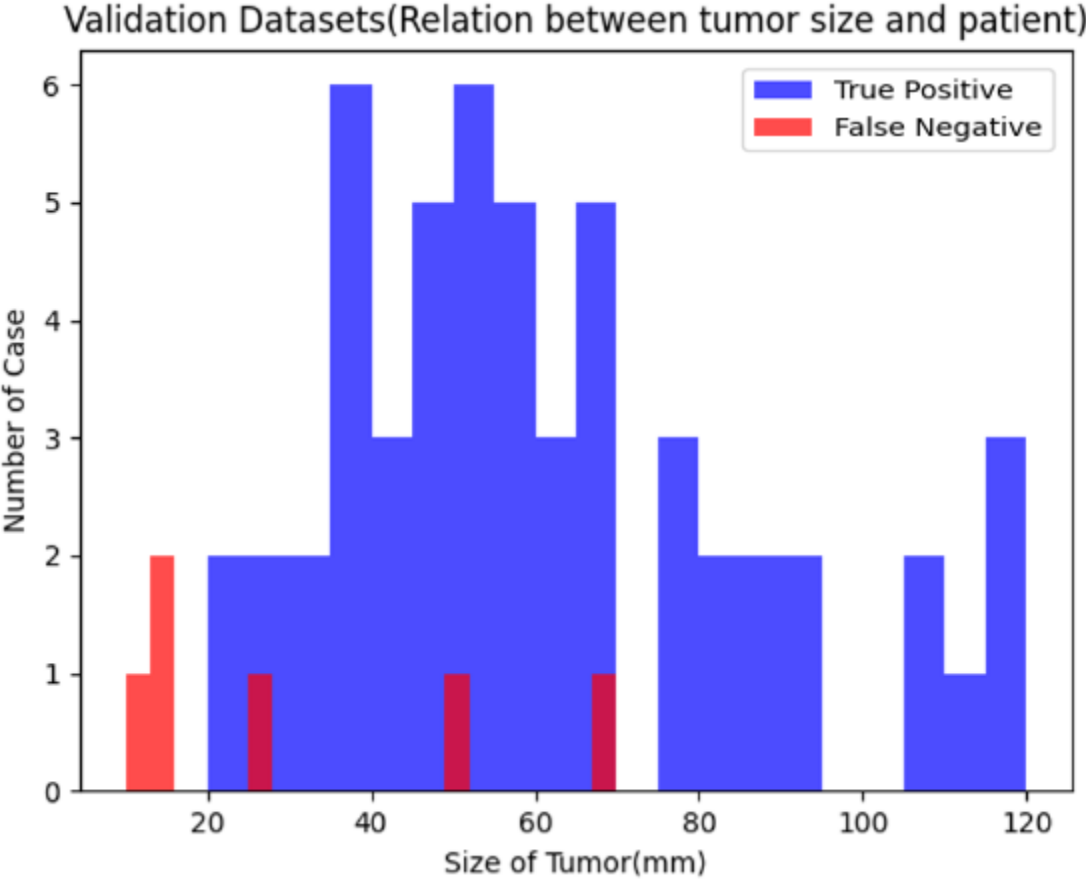
Fig. 連續四張代表此病人的腫瘤20mm

病人編號1-60的分布

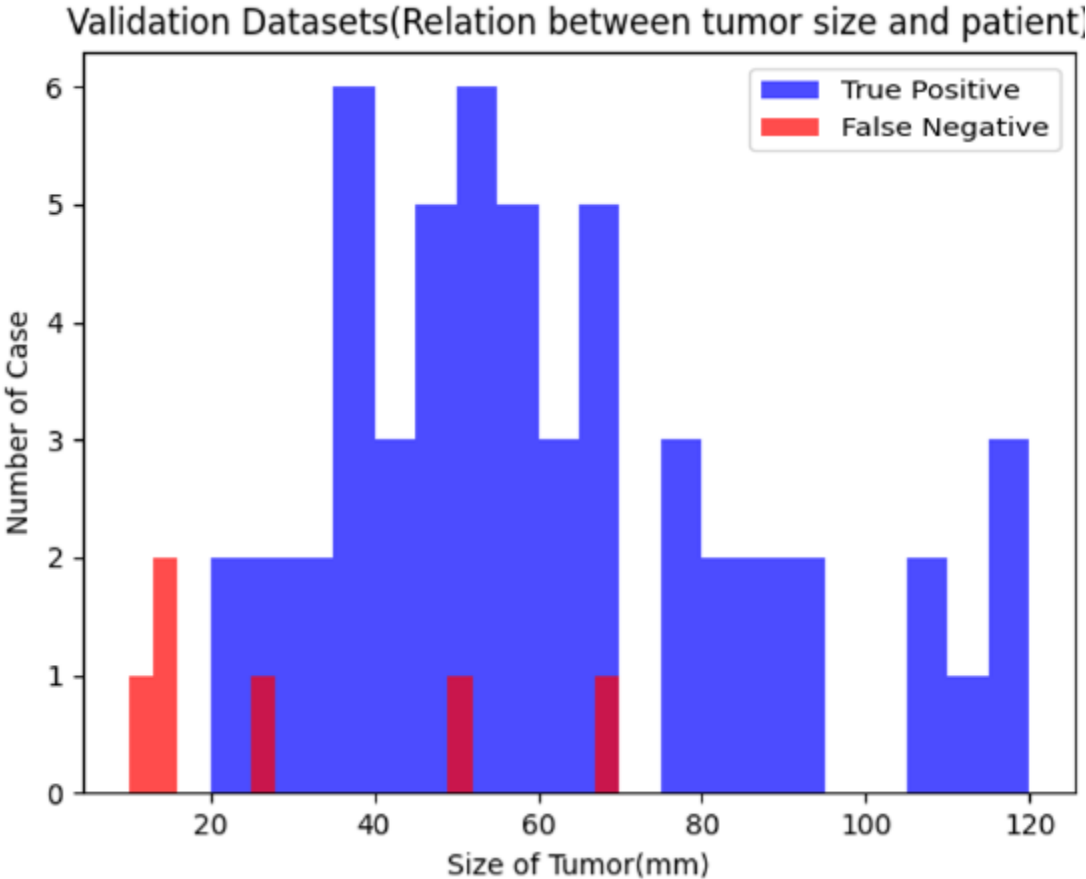


3.4 Data distribution(2/3)

病人編號61-120的分布

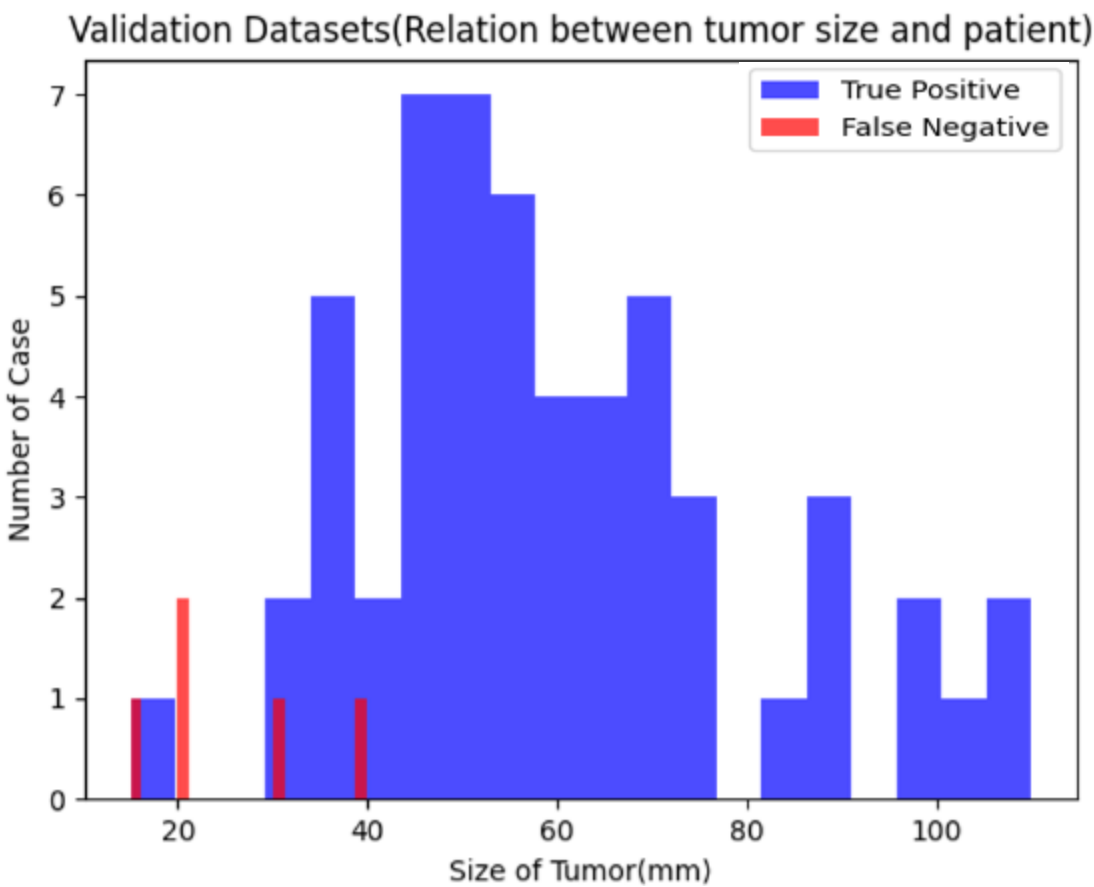


病人編號121-180的分布

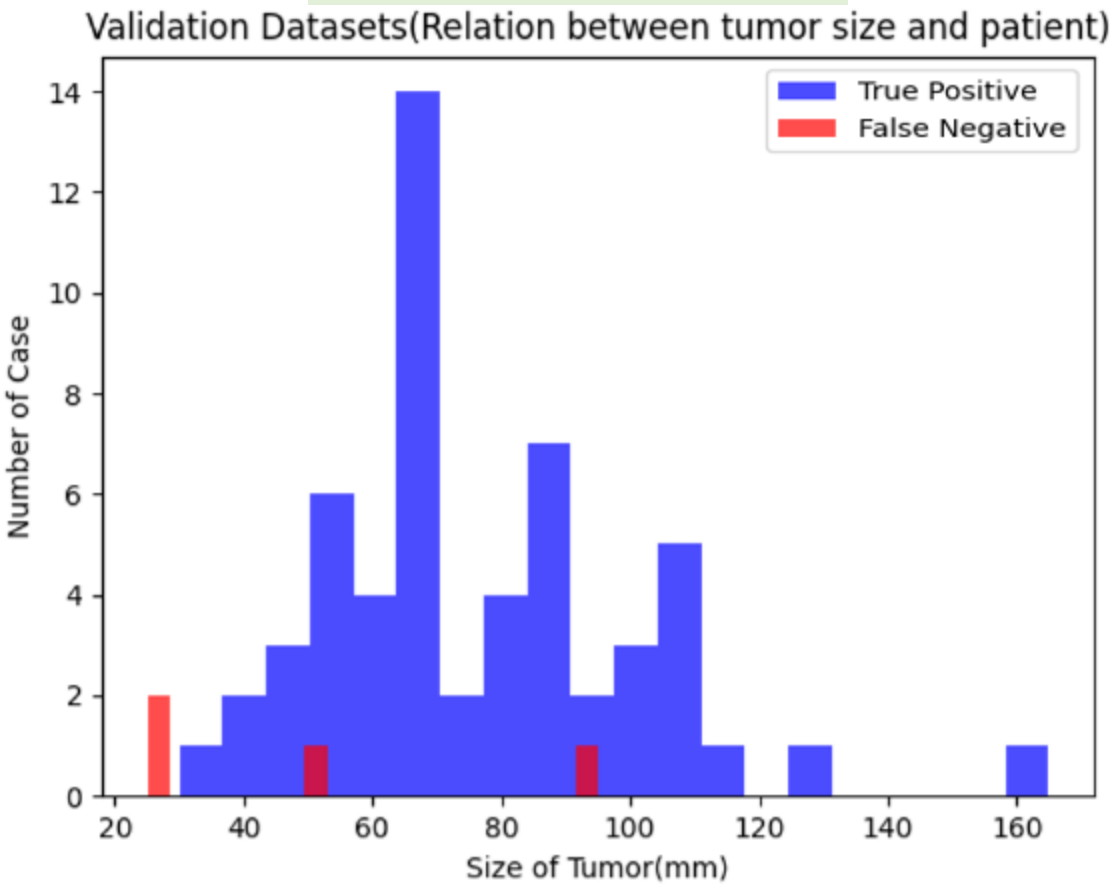


3.4 Data distribution(3/3)

病人編號181-240的分布



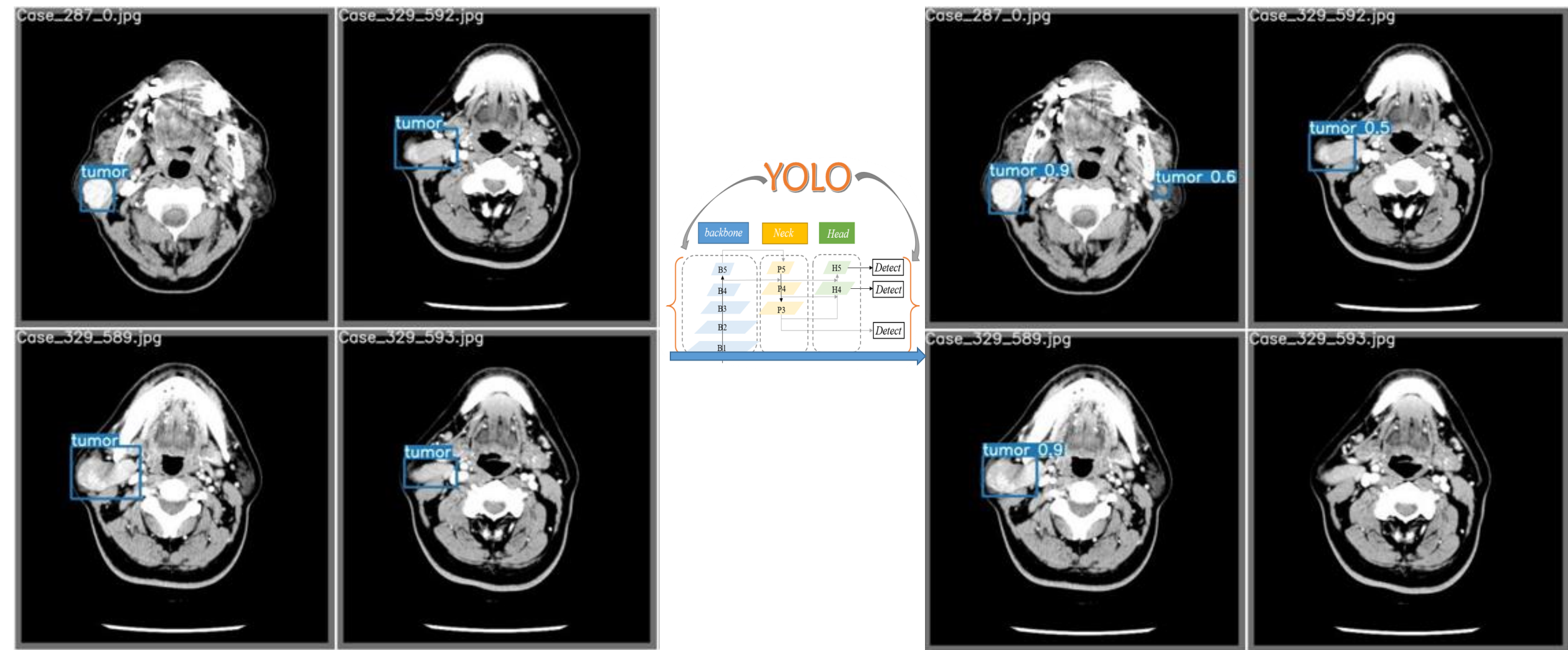
病人編號241-300的分布



3.5 視覺化呈現結果(1/4)

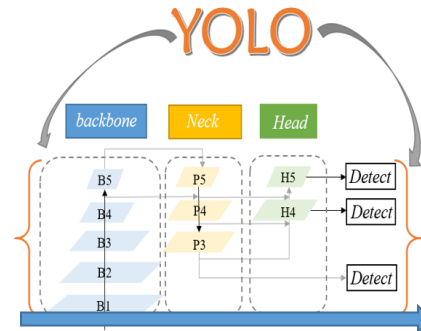
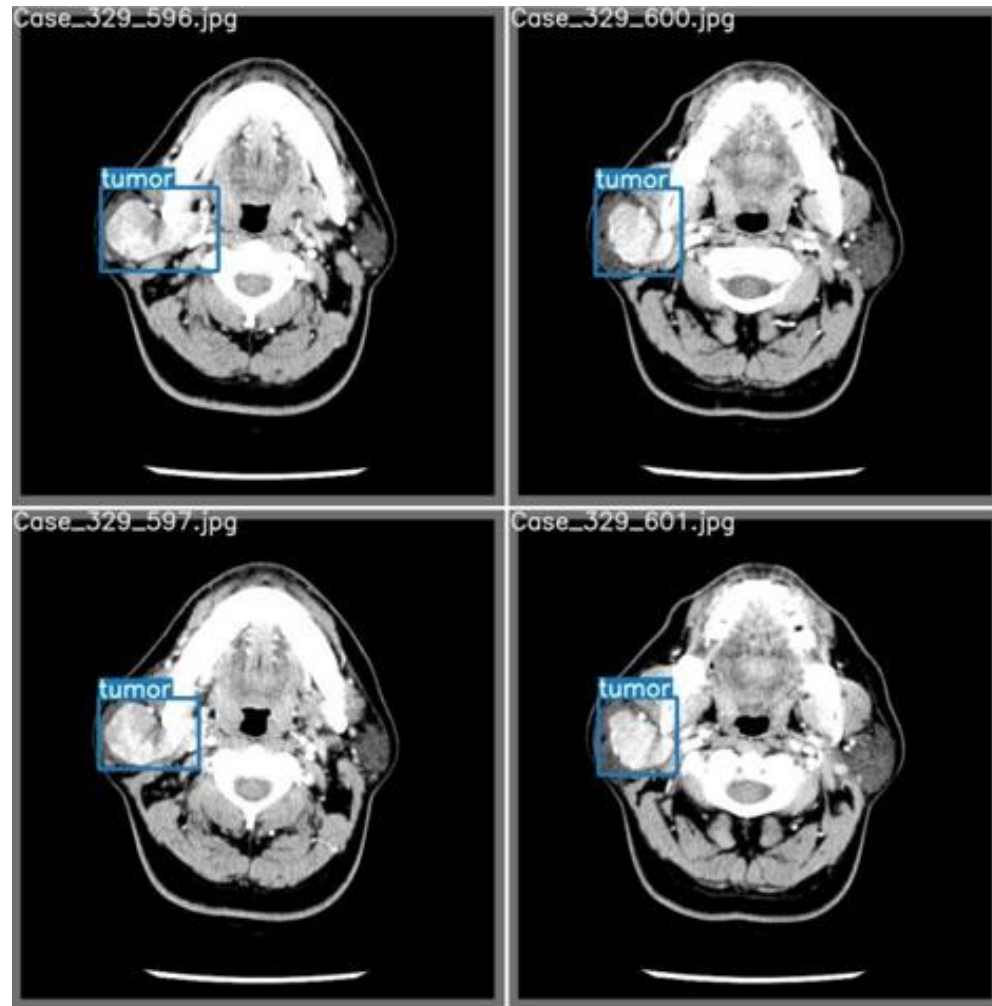
Ground Truth

Prediction

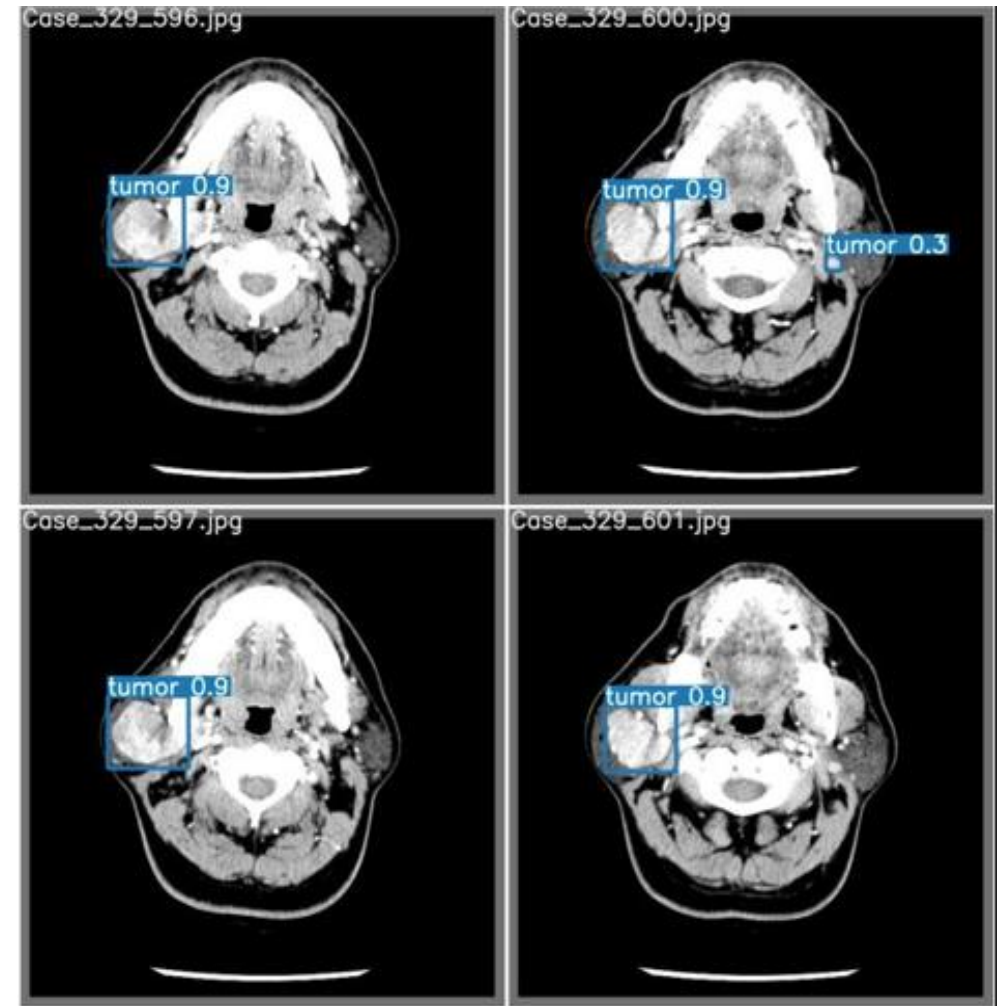


3.5 視覺化呈現結果(2/4)

Ground Truth

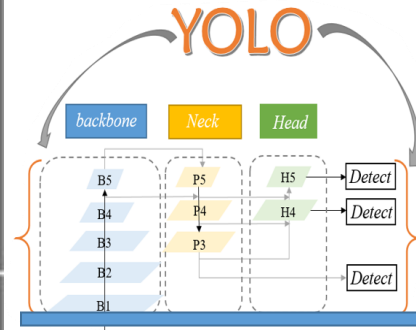
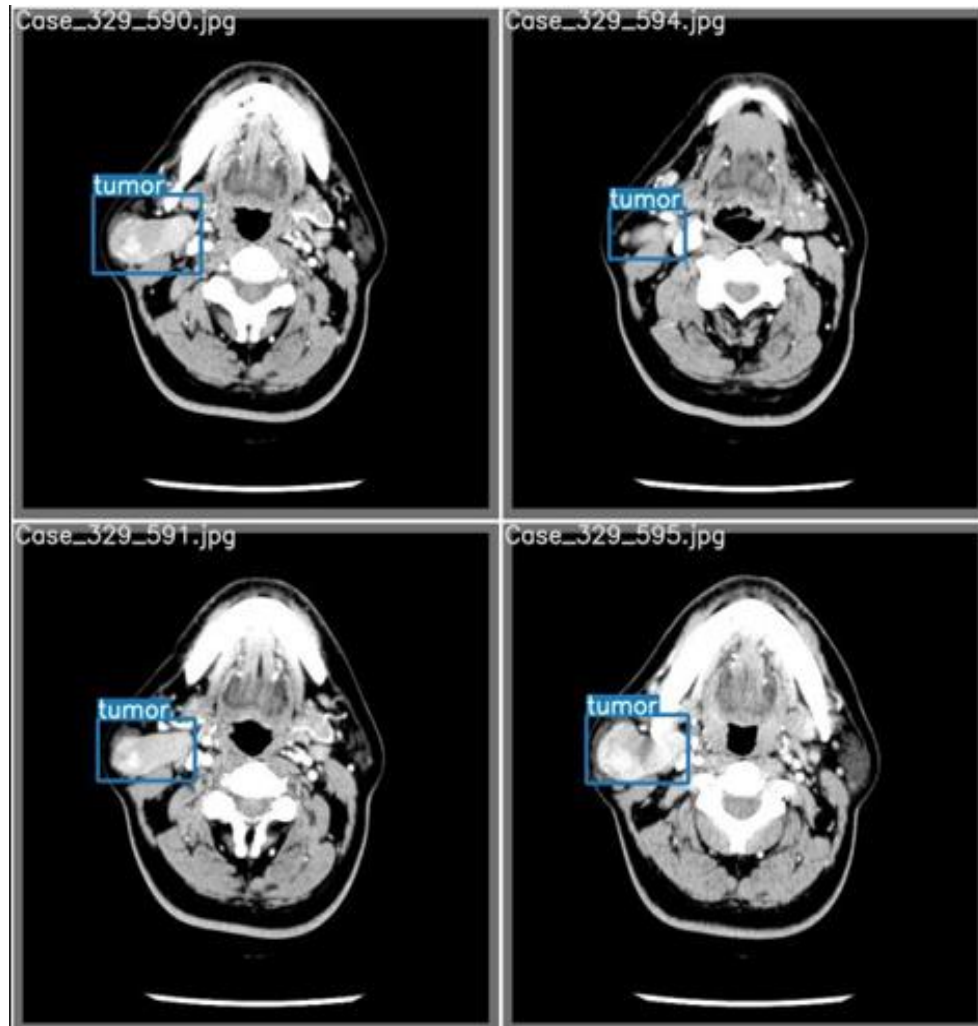


Prediction

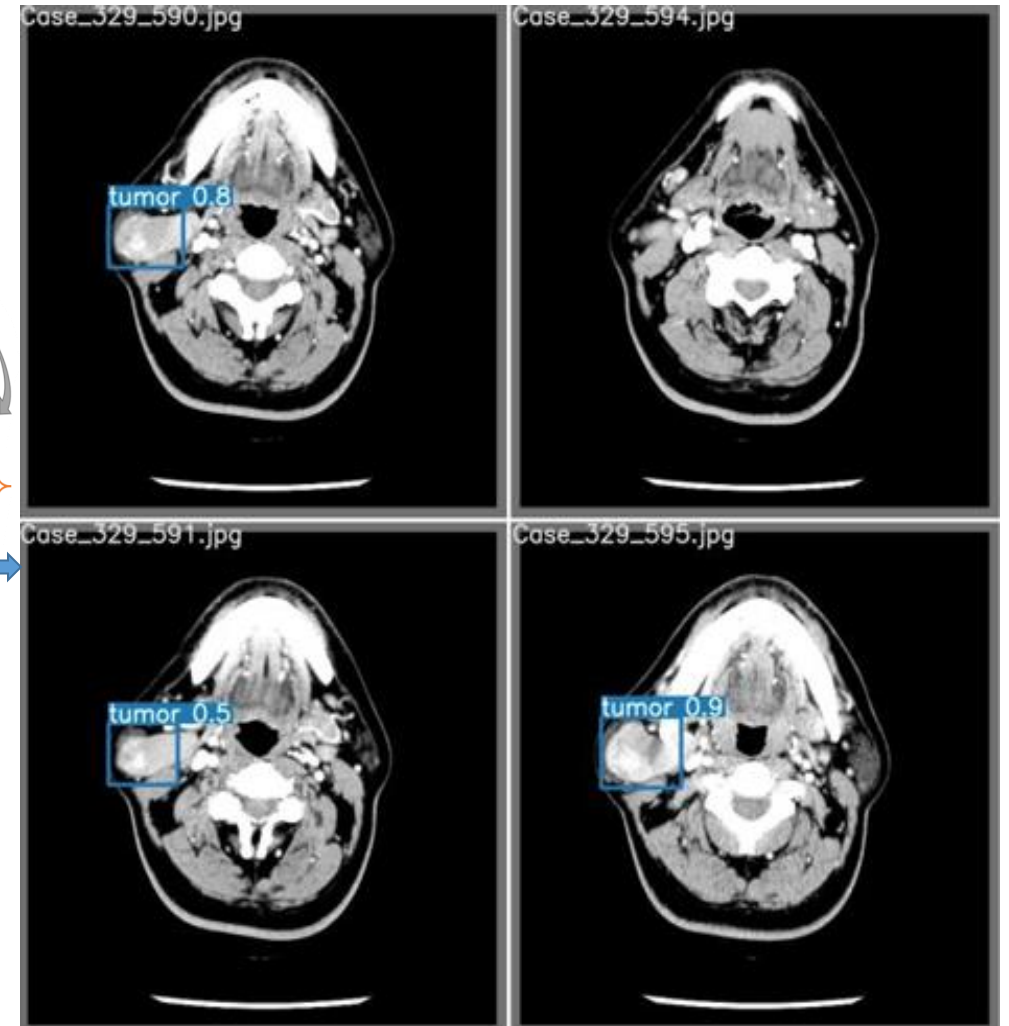


3.5 視覺化呈現結果(3/4)

Ground Truth

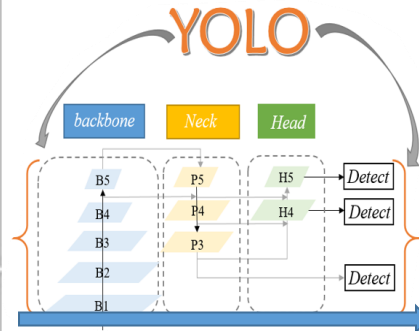
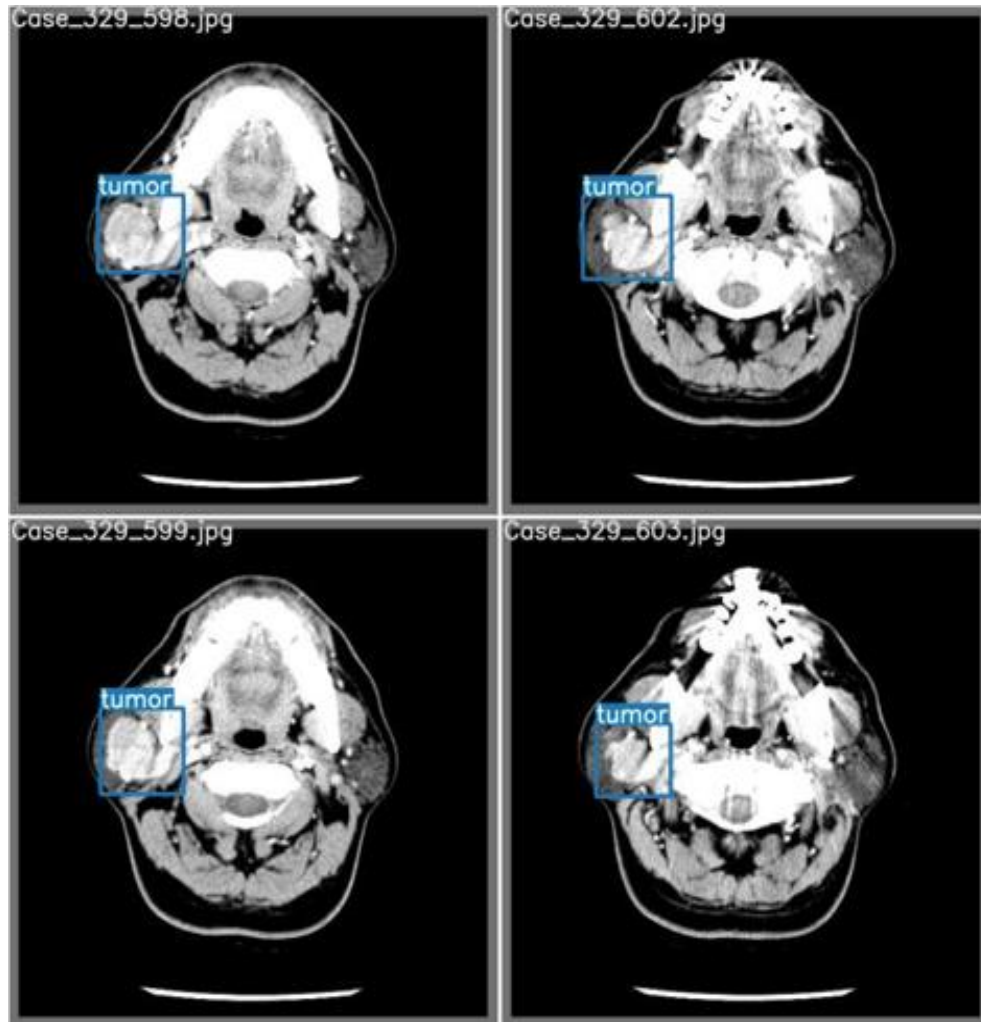


Prediction

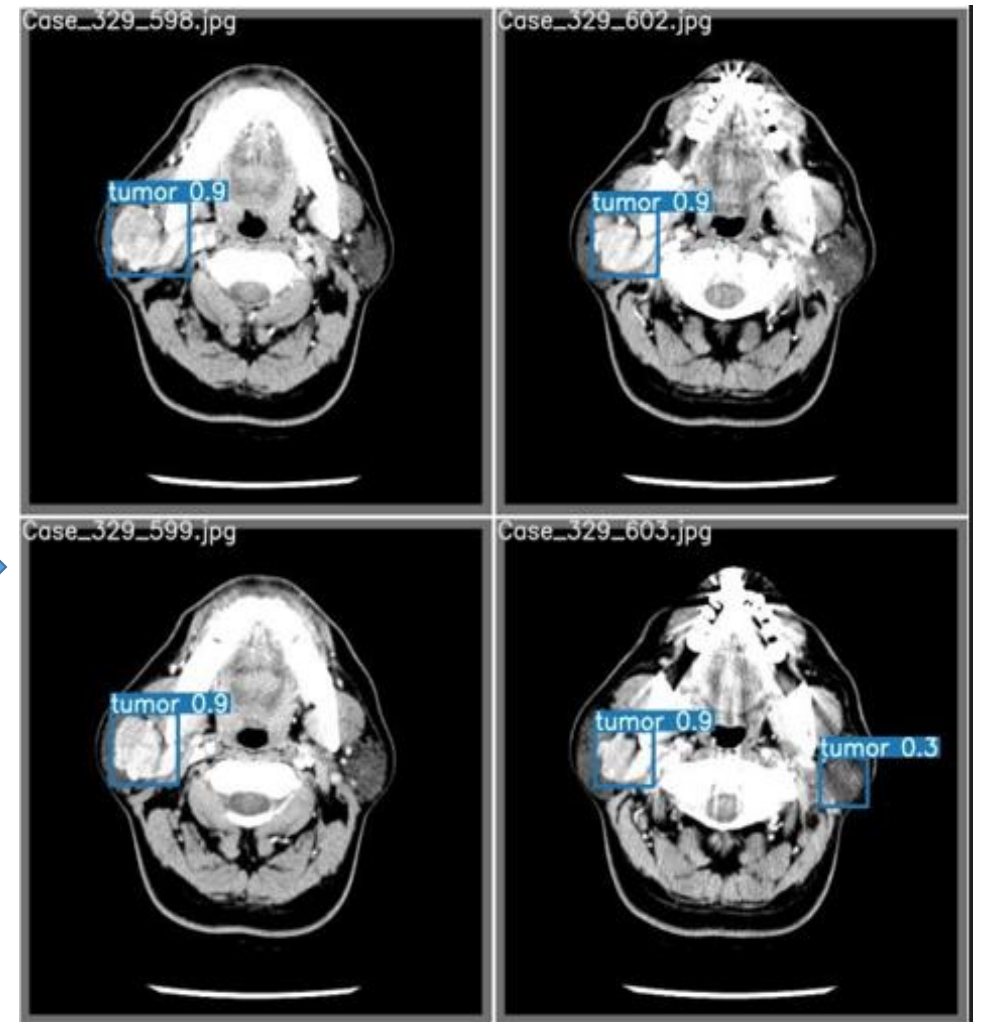


3.5 視覺化呈現結果(4/4)

Ground Truth



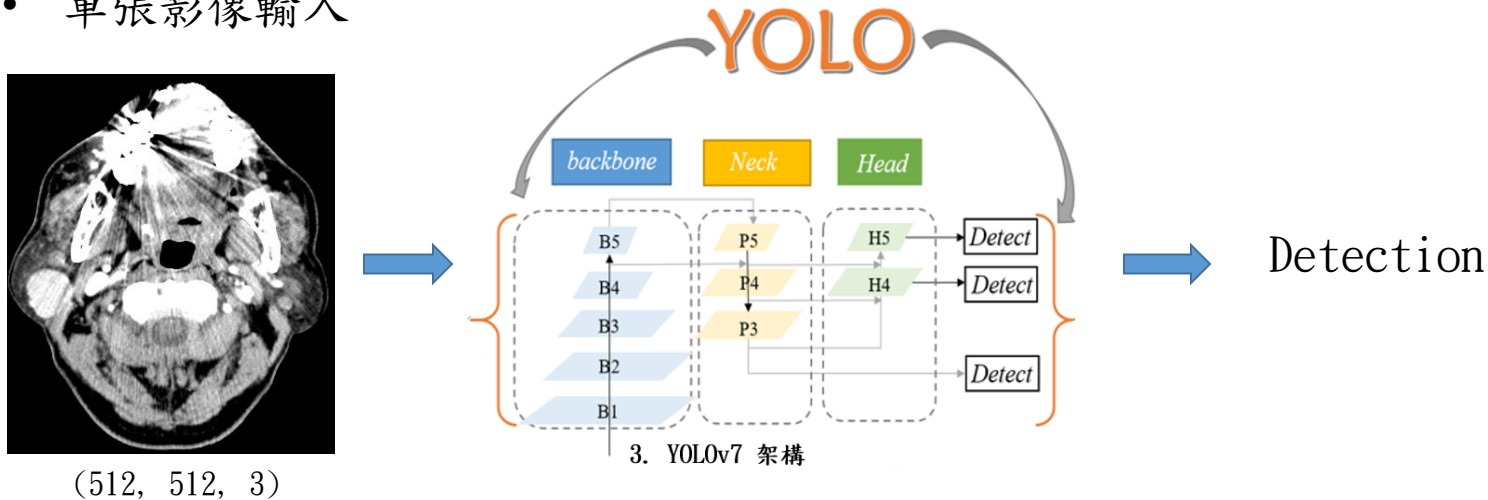
Prediction



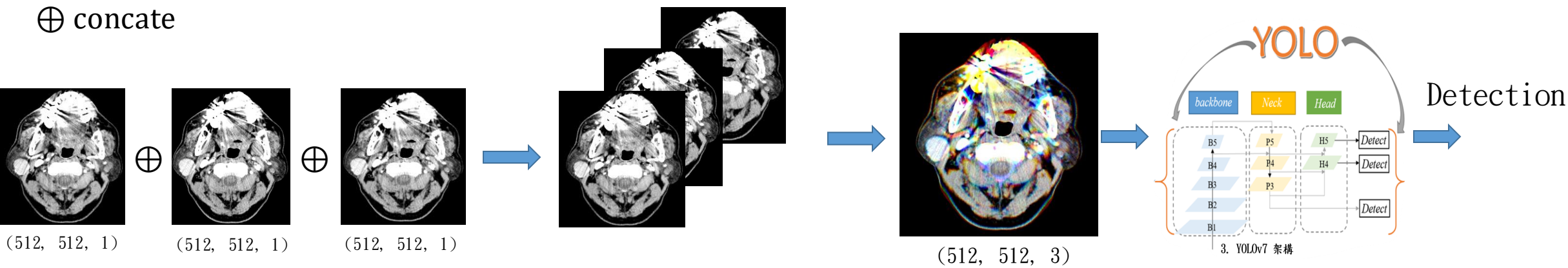
4. 後續實驗

4.1 單張影像輸入 v.s. 上下文資訊整合輸入－示意圖

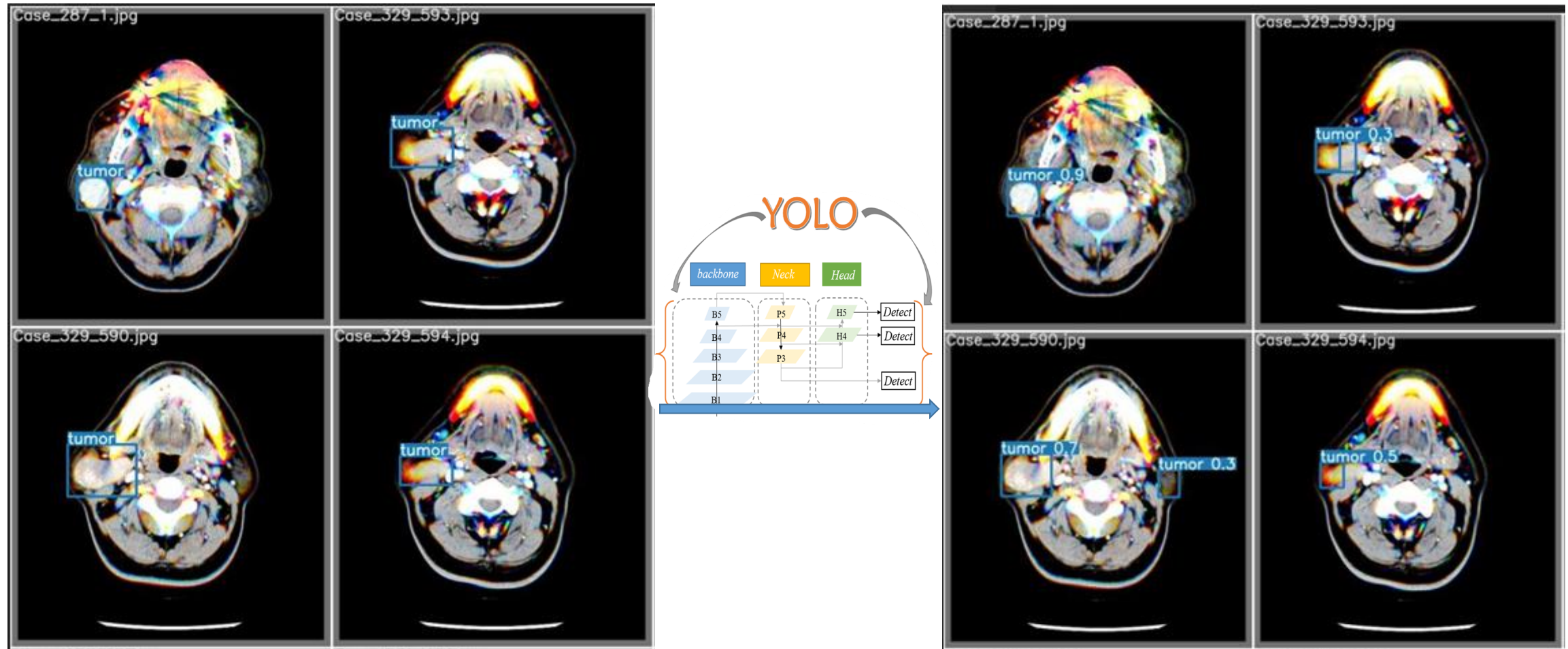
- 單張影像輸入



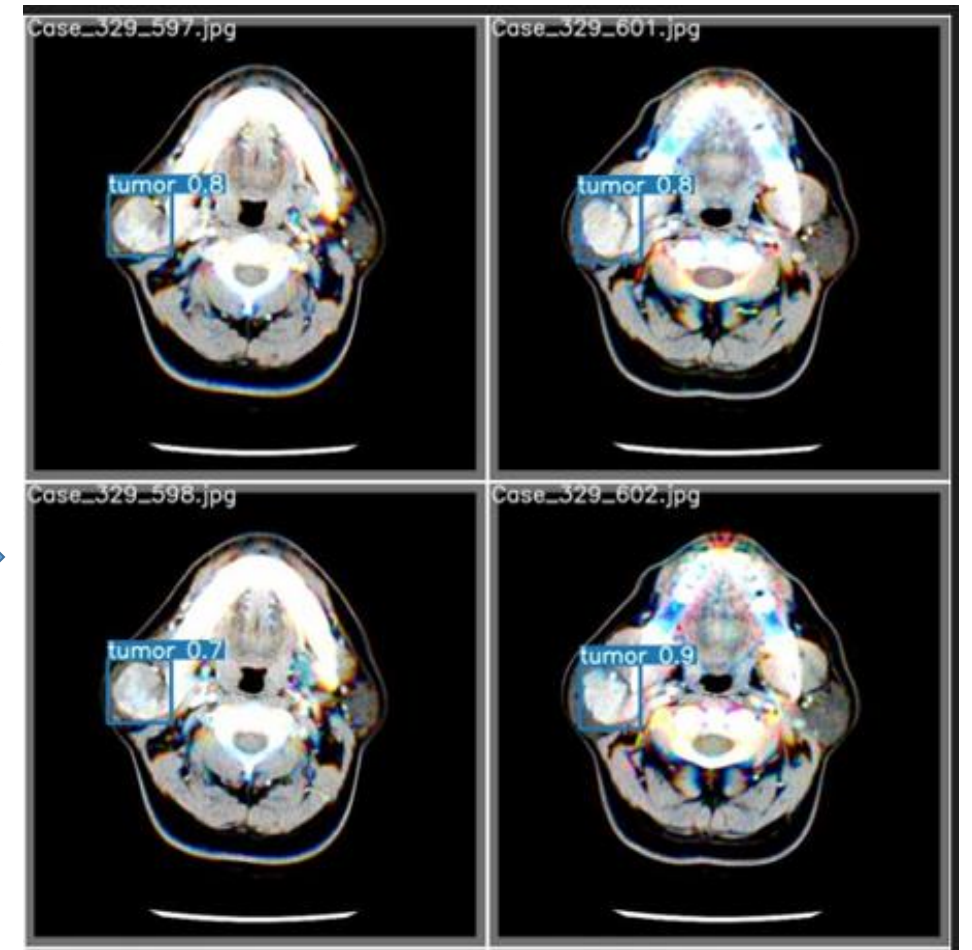
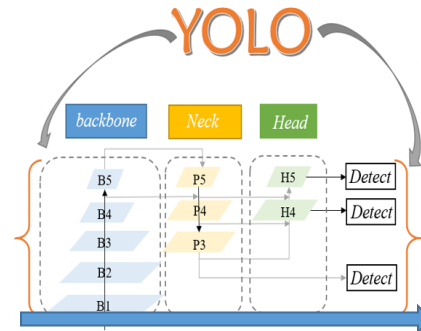
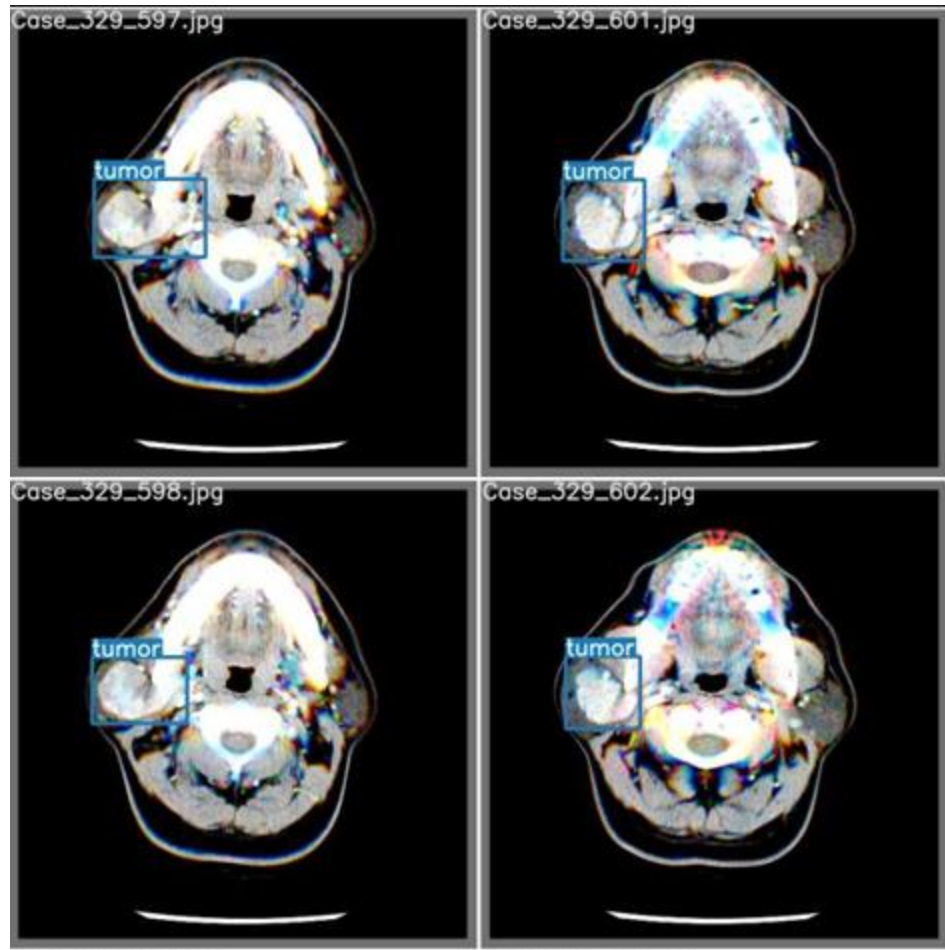
- 上下文資訊整合輸入



4.1 單張影像輸入 v.s. 上下文資訊整合輸入－視覺化



4.1 單張影像輸入 v.s. 上下文資訊整合輸入－視覺化



4.1 單張影像輸入 v.s. 上下文資訊整合輸入－量化

藍色代表輸入是連續的三張
其他代表輸入為單張影像

Cross validation Metrics						
評估準則	consecutive=1	consecutive=2	consecutive=3	consecutive=4	consecutive=5	consecutive=6
(TP/TN/FP/FN)	(57.4/19.2/40.8/2.6) (54.4/20.2/39.8/5.6)	(55.6/43.4/16.6/4.4) (51.8/40.4/19.6/8.2)	(53.6/51/9/6.4) (48.2/48.8/11.2/11.8)	(49.2/56.2/3.8/10.8) (44.8/53.4/6.6/15.2)	(44.2/58.2/1.8/15.8) (40.2/56.2/3.8/19.8)	(40.6/59.2/0.8/19.4) (36.8/57.6/2.4/23.2)
Accuracy	0.638/ 0.621	0.825/ 0.768	0.876/ 0.808	0.878/ 0.818	0.853/ 0.803	0.831/ 0.786
Recall(Sensitivity; TPR)	0.956/ 0.906	0.926/ 0.863	0.893/ 0.803	0.82/ 0.746	0.736/ 0.67	0.676/ 0.613
Precision	0.584/ 0.577	0.77/ 0.725	0.856/ 0.811	0.928/ 0.871	0.96/ 0.913	0.98/ 0.938
Specificity	0.32/ 0.336	0.723/ 0.673	0.85/ 0.813	0.936/ 0.89	0.97/ 0.936	0.986/ 0.96
NPV	0.88/ 0.782	0.907/ 0.831	0.888/ 0.805	0.838/ 0.778	0.786/ 0.739	0.753/ 0.712

結論：
懷疑可能是overfitting的問題，給予更多的資訊學習的方向可能不正確，導致給予上下文資訊的結果反而沒有比較好。