資料蒐集與處理

資料蒐集與處理

爬蟲

Web Crawl

因爬蟲後圖片品質不一，
最後選擇人工搜圖

人工找圖

IG+FB

搜圖同時也過濾掉不適合進行訓練，背景也比較乾淨的圖片

資料標記

Labelimg

採人工標記/
code(自動標記)

Deployment

2D會比較好 for Yolo

Box: Pose: 5 Landmarks:

Landmarks: Mask: 3D mesh(Ours): 3D mesh:

More Informative

我們可以發現出yolo 好對邊緣特正擷取有較好的結果

實驗偵測發現


joeman 0.75
beef 0.70


mc jeng 0.85
超廢 敗財
不只廢還有很貴的東西

模型比較問題

模型使用

Nano

YOLOv5n

4 MB_{FP16}
6.3 ms_{V100}
28.4 mAP_{COCO}

Small

YOLOv5s

14 MB_{FP16}
6.4 ms_{V100}
37.2 mAP_{COCO}

Medium

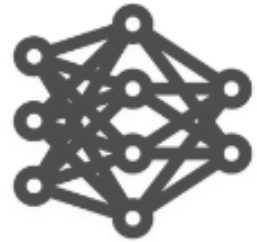
YOLOv5m

41 MB_{FP16}
8.2 ms_{V100}
45.2 mAP_{COCO}

Large

YOLOv5l

89 MB_{FP16}
10.1 ms_{V100}
48.8 mAP_{COCO}



XLarge

YOLOv5x

166 MB_{FP16}
12.1 ms_{V100}
50.7 mAP_{COCO}

Key Numbers

1500

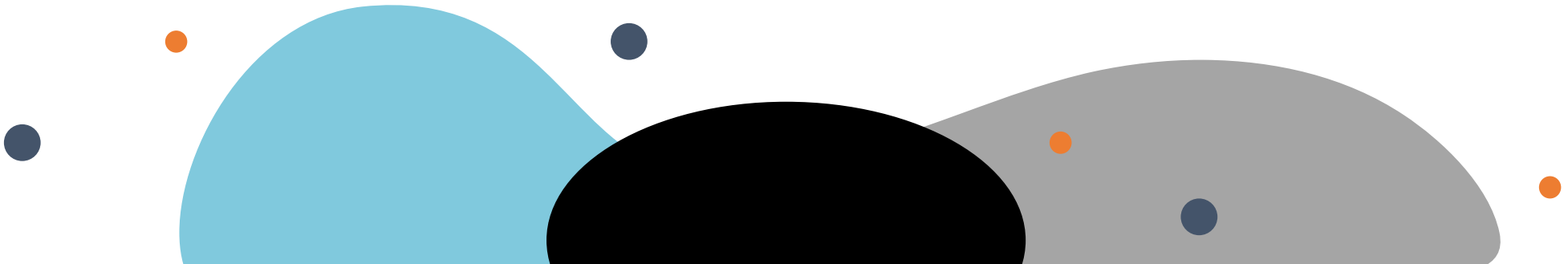
photos

50

Youtubers

83.2

mAP



Key Numbers

80%

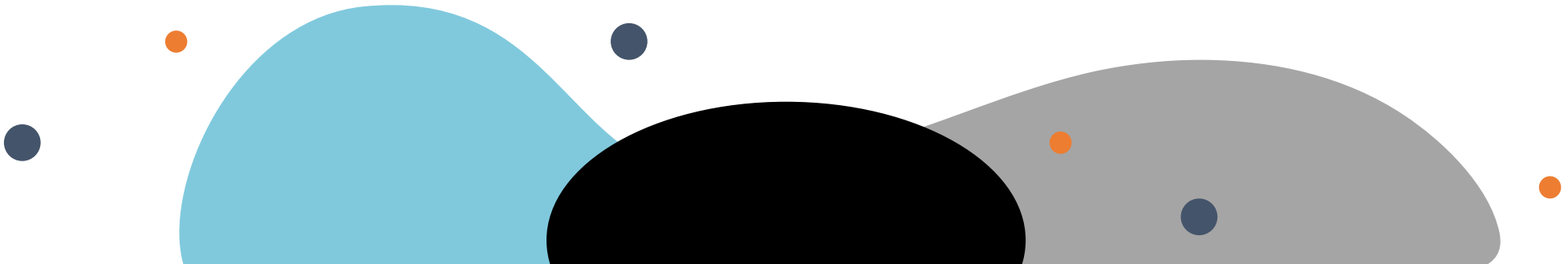
Training Set

13%

Validation Set

7%

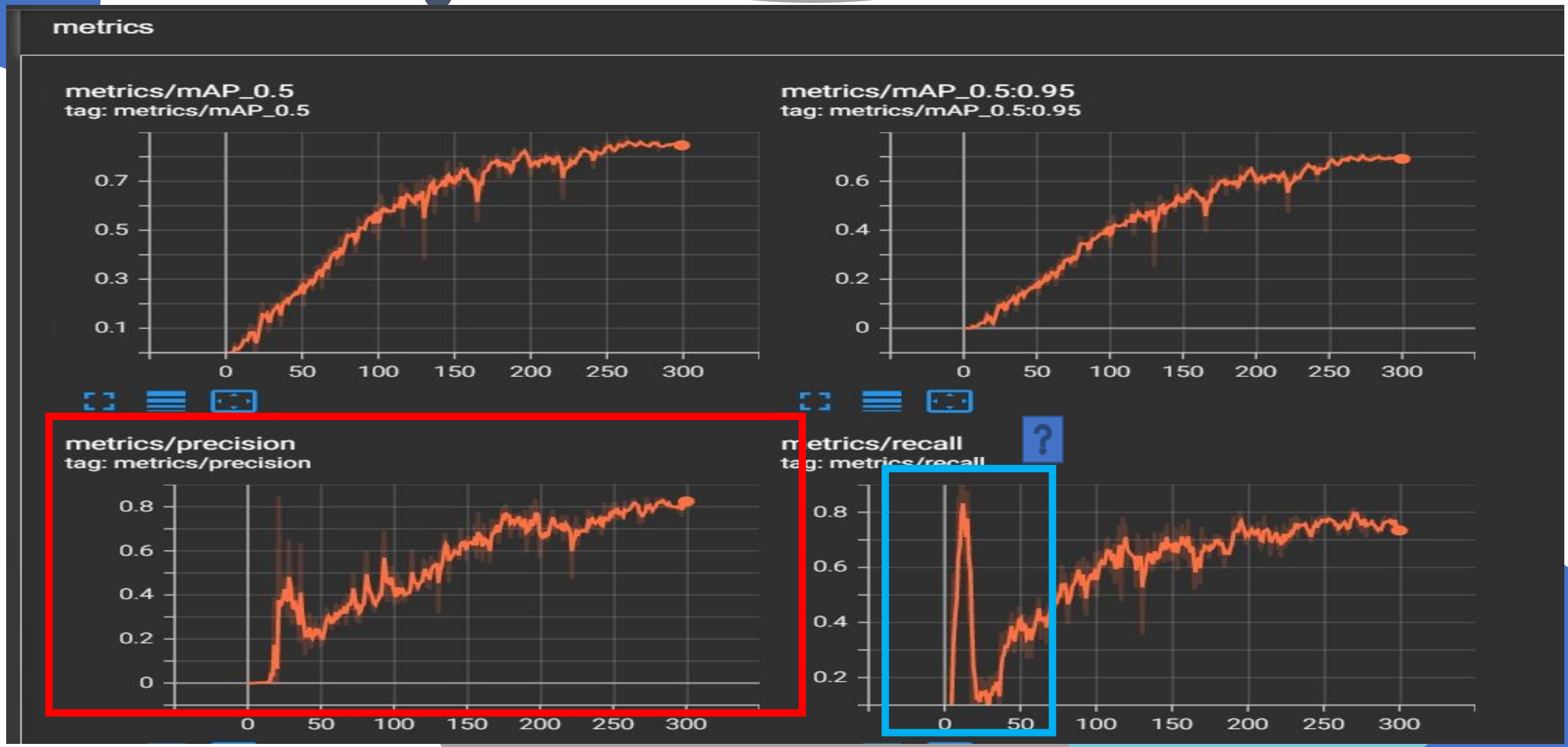
Testing Set



RESULT

```
custom_YOLOv5s summary: 232 layers, 7384065 parameters, 0 gradients, 17.2 GFLOPs
                 Class     Images     Labels          P          R     mAP@.5 mAP@.5:.95: 100%
                   all        200        205      0.718      0.767      0.832      0.686
```

RESULT


metrics
metrics/mAP_0.5
tag: metrics/mAP_0.5
0.7
0.5
0.3
0.1
0
50
100
150
200
250
300
metrics/mAP_0.5:0.95
tag: metrics/mAP_0.5:0.95
0.6
0.4
0.2
0
0
50
100
150
200
250
300
metrics/precision
tag: metrics/precision
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
etrics/recall
g: metrics/recall
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300

實驗

1. 我們只有時間了解準確率問題
2. 但是Recall rate 很高的地方也是我們可以深入研究的
3. 交叉質**(Intersection over union)**

   交疊比例並無顯著差異，所以這邊不探討

.

真正例：$TP = TruePositive$

真反例：$TN = TrueNegative$

假正例：$FP = FalsePositive$

假反例：$FN = FasleNegative$

则，查准率和查全率计算公式：

查准率：$Precision = \frac{TP}{TP+FP}$

查全率：$Recall = \frac{TP}{TP+FN}$

模型使用

Nano

YOLOv5n

4 MB_{FP16}
6.3 ms_{V100}
28.4 mAP_{COCO}

Small

YOLOv5s

14 MB_{FP16}
6.4 ms_{V100}
37.2 mAP_{COCO}

Medium

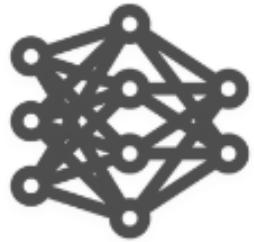
YOLOv5m

41 MB_{FP16}
8.2 ms_{V100}
45.2 mAP_{COCO}

Large

YOLOv5l

89 MB_{FP16}
10.1 ms_{V100}
48.8 mAP_{COCO}



XLarge

YOLOv5x

166 MB_{FP16}
12.1 ms_{V100}
50.7 mAP_{COCO}

Key Numbers

1500

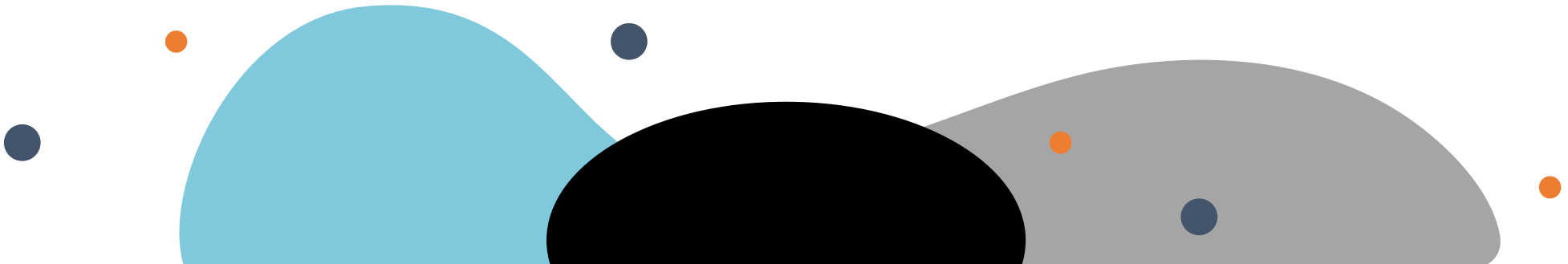
photos

25

Youtubers

87.4

mAP



Key Numbers

80%

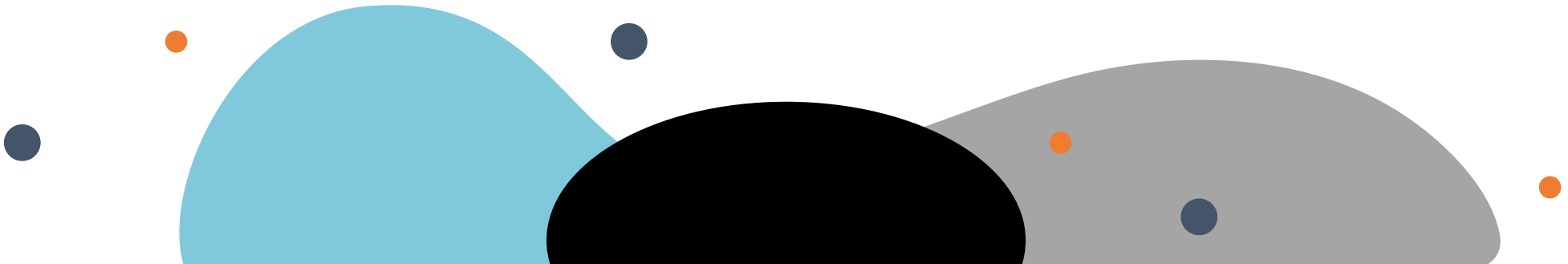
Training Set

13%

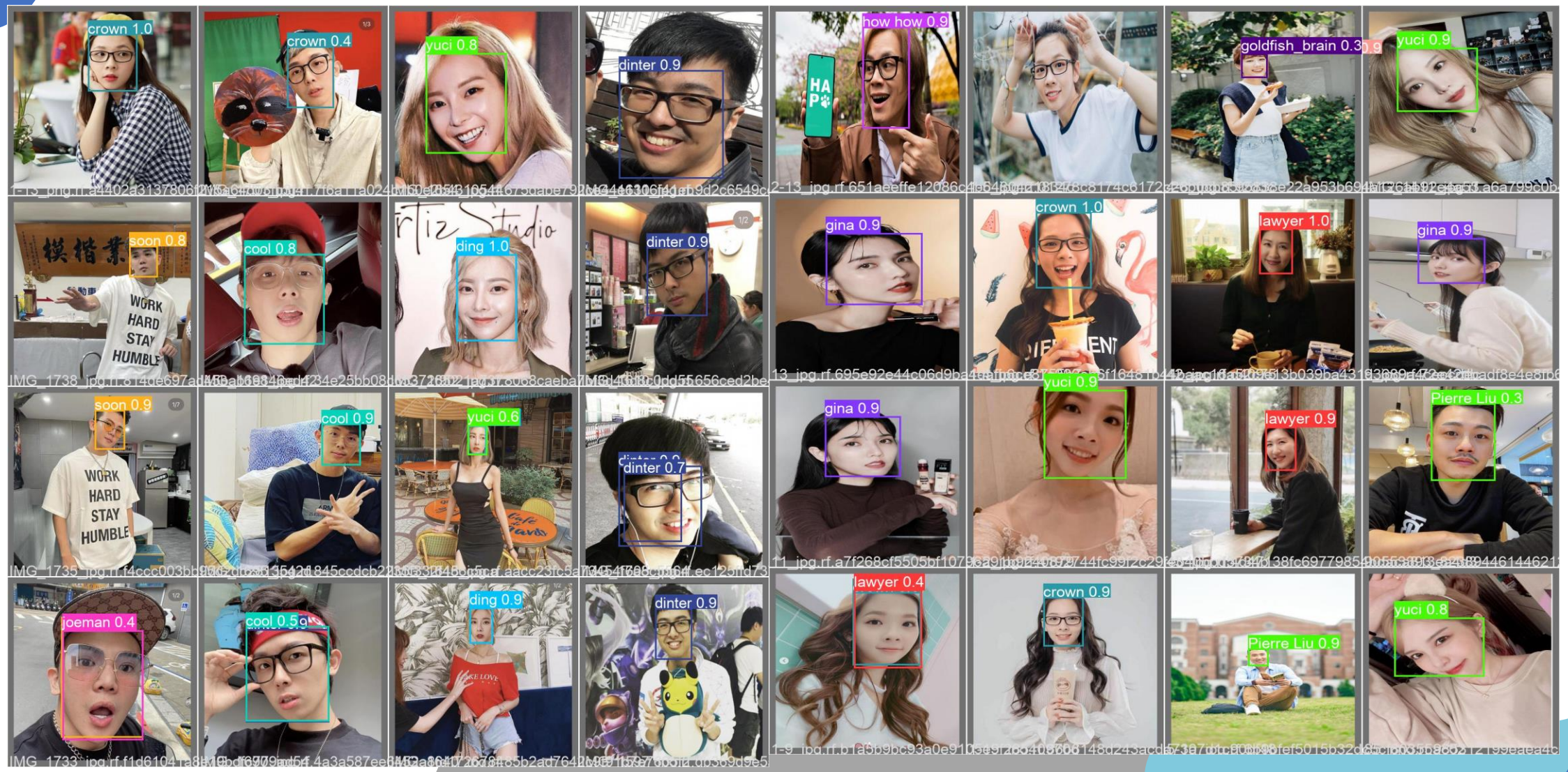
Validation Set

7%

Testing Set



RESULT


crown 1.0
crown 0.4
yuci 0.8
dinter 0.9
how how 0.9
goldfish_brain 0.3
yuci 0.9
soon 0.8
cool 0.8
ding 1.0
dinter 0.9
gina 0.9
crown 1.0
lawyer 1.0
gina 0.9
soon 0.9
cool 0.9
yuci 0.6
dinter 0.7
gina 0.9
yuci 0.9
lawyer 0.9
Pierre Liu 0.3
joeman 0.4
cool 0.5
ding 0.9
dinter 0.9
lawyer 0.4
crown 0.9
Pierre Liu 0.9
yuci 0.8

最後選擇modelS的原因

1500

photos

可辨識廣

Youtubers多才比較符合我們所需要的情況

無顯著差異

mAP只有差距4%

結論

結論

1. 2D is better for yolo :

   因為人臉是屬於3D 所以在辨識細節會比較困難(Transfer 會有難度)

2. 資料收集困難:

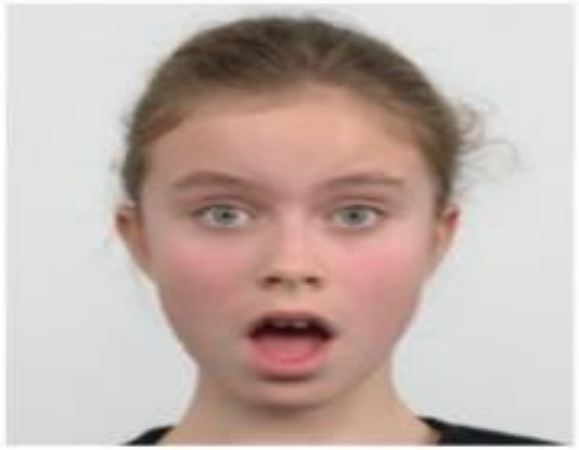
   希望在未來可以解決這個問題

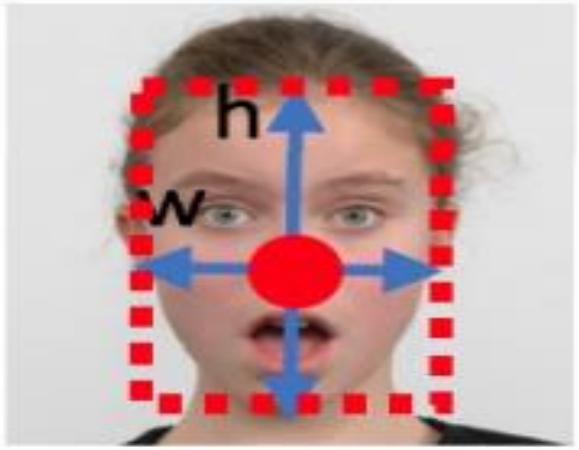
3. 源頭決定一切

未來展往

- State of the art :
  - 所以我們嘗試使用yolo7模型，但發現模型過於巨大，GPU 容易不足
- More Function:
  - 優質的圖篇增加辨識準確率
  - 加入不同元素的臉(有戴帽子或是沒戴帽子的資料)
- 自動化:
  - 自動化標記(RetinaFace)

RetinaFace

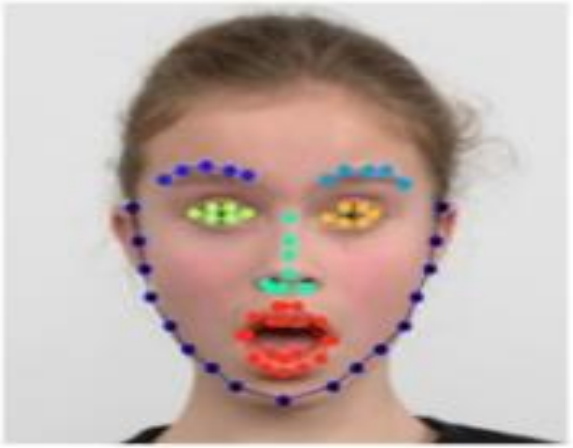




Box:
4 scalars

Pose:
7 scalars

5 Landmarks:
10 scalars



68 Landmarks:
136 scalars

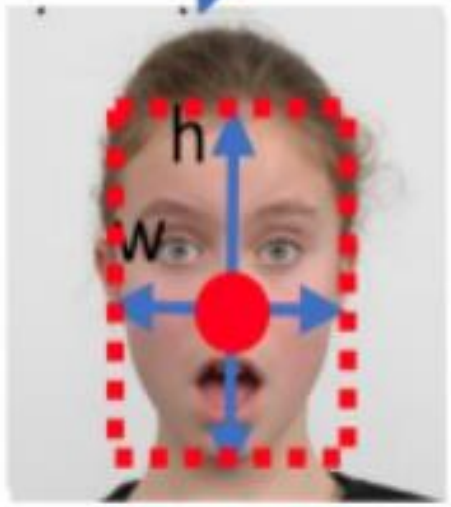
Mask:
H x W matrix

3D mesh(Ours):
3 x 1k vertices

3D mesh:
3 x 53k vertices

More Informative

(1) More semantic points, more accurate box prediction



1k 3D points enhance
pose-invariant 5 points

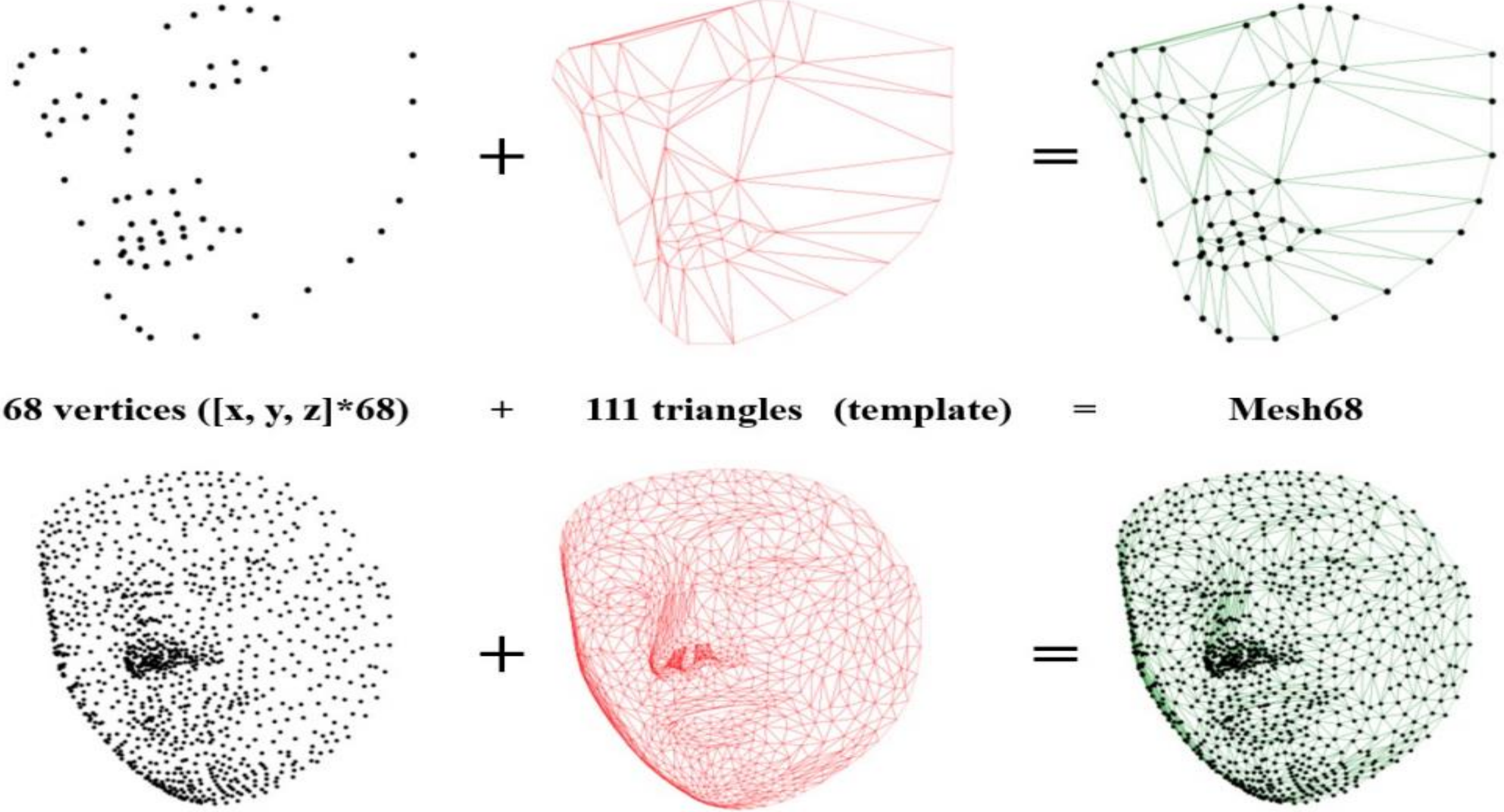
Cheap 5 points enhance
robust 1k points

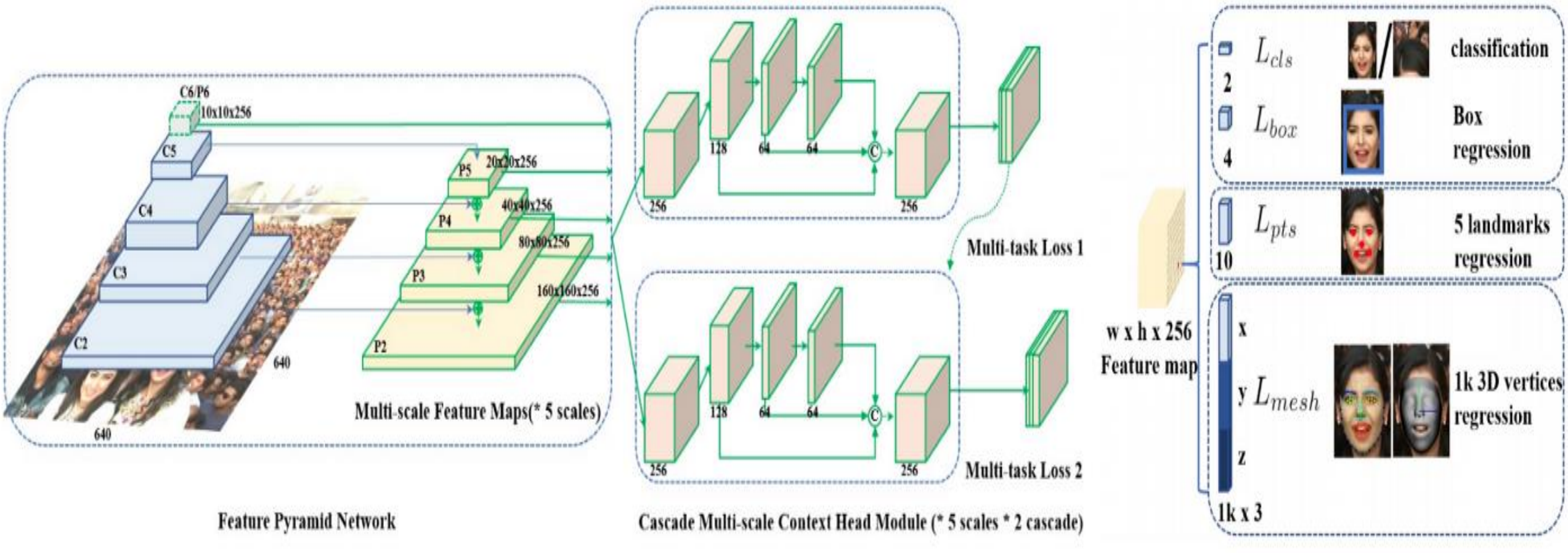
Face Detection
(one center point)

2D Face Alignment
(five points)

3D Face Reconstruction
(1k points)

(2) More challenging training scenario, more robust point prediction


68 vertices ([x, y, z]*68) + 111 triangles (template) = Mesh68
1035 vertices ([x, y, z]*1035) + 1999 triangles (template) = Mesh1k



(a) Network Structure

(b) Multi-task Loss