

## 二、研究計畫內容

### (一) 摘要

微笑美學被認為是患者選擇牙科治療的重要原因，因此數位微笑設計 (DSD) 技術的出現，將幫助矯正醫師更方便診斷分析患者的微笑美學，並向患者視覺化呈現。數位微笑設計技術可以幫助矯正醫師在治療過程中，對患者的牙齒進行模擬，並根據患者的需求和預期效果調整。此外，數位微笑設計還可以提供更加直觀和易于理解的建議，讓患者更清晰地了解矯正過程的全貌。

但是如今的微笑設計軟體仍需要矯正醫師手動標記參考線、牙齒位置等資訊，無法做到軟體自動分析患者微笑美觀構成，因此尚不能做到即時向患者視覺化呈現。因此本研究旨在將實例分割技術應用於微笑設計 (DSD)，希望透過實例分割模型自動分析微笑照片，分割並標記牙齒、嘴唇、牙齦，最後將分割結果用於計算微笑美觀參數。因為在 DSD 的應用場景中，主要分析患者正顎與側顎的影像，因此本研究將針對正顎與側顎訓練分割模型。我們會使用 AI 生成的人臉與向患者搜集的照片作為訓練集，並使用 YOLACT++ 與 BlendMask 模型進行訓練，最後比較評估兩種模型的辨識結果。本研究計畫假設能在符合 DSD 臨床使用的情況下，得到更好的分割精度 (AP)。該分析模型將用於開發一套易於使用的微笑特徵辨識系統，用以簡化 DSD 軟體操作流程，幫助矯正醫師及時向患者呈現矯正或是微笑訓練的成果。

## (二) 研究動機與研究問題

現今患者越來越意識到有吸引力的微笑對臉部美觀的重要性，因此矯正學除了重視咬合關係外，也開始重視咬合與周圍軟組織的協調性，微笑的吸引力是專業人士評斷審美偏好的重要標準<sup>[1]</sup>。此外，改善面部和牙齒的吸引力亦是患者尋求牙科治療的主因之一。在日常社交互動，微笑對於面部美學有顯著影響<sup>[2]</sup>，因此微笑在矯正中是十分重要的。

許多因素都會影響微笑美學，例如，微笑線、牙唇關係、微笑弧度、協調的面部比例以及對稱性<sup>[3][4][5]</sup>。然而微笑美學是主觀的，受到個人經歷與社會社會環境影響。例如由於牙科經歷不同，矯正牙醫、一般牙醫和非專業人士之間對微笑的偏好存在明顯差異<sup>[2][6]</sup>。甚至在不同地區或性別的矯正醫師間仍從在顯著差異<sup>[2][7]</sup>。因此在矯正治療與微笑訓練上，矯正醫師需要考慮患者對微笑的美觀偏好，診斷評估並與患者溝通微笑美觀的各種構成。

若能有效率的分析微笑構成並呈現，將能輔助矯正醫師診斷，並有效降低醫師與患者的溝通成本，而這些正是現今數位微笑設計 (DSD,digital smile design) 技術的目標<sup>[8][9][10]</sup>。目前的 DSD 技術主要是紀錄患者前牙列的 2D 圖像或是面部掃描的 3D 數據上<sup>[8][9][11]</sup>。但是在影像分析上，仍需要醫師先手動繪製參考線、標記輪廓與形狀，例如面部中線、門牙切線等，軟體方能取得影像或模型的美學參數資訊<sup>[8]</sup>。

人工智能 (AI) 已經大幅改變我們的生活。尤其是機器學習 (ML) 技術，得利於其優異的圖像處理和決策支援能力，其在各領域皆有廣泛應用。在醫學影像上，卷積神經網路 (CNN) 透過學習標記過的醫學影像，已經在診斷視網膜病變<sup>[12]</sup>、間質性肺疾病 (Interstitial Lung Disease)<sup>[13]</sup> 有良好成效。在口腔矯正領域，卷積神經網路 (CNN) 影像處理也在 2D 頭部 X 光影像的標註<sup>[14]</sup>、檢測矯正患者牙齦炎的早期徵兆<sup>[15]</sup> 方面取得有效成果。

但是對於 DSD 技術中的二維圖像標註，目前仍須矯正醫師手動進行，並未應用卷積神經網路 (CNN) 影像辨識。雖然目前已有研究訓練 YOLACT++ 模型分割微笑的二維圖像，但在牙齒的分割上成效不佳<sup>[16]</sup>。該研究是以各種不同角度的微笑圖片進行訓練，而在 DSD 的使用場景中，分析對象是以固定角度、固定方法拍攝的微笑照片（正顎、側顎）<sup>[8][9][17]</sup>。另外，因為 YOLACT++ 是較重視辨識速度的模型，因此其準確度比起辨識速度較慢的 BlendMask 模型差<sup>[18][19]</sup>。所以本研究希望分別利用 YOLACT++ 與 BlendMask 模型，針對不同角度（正顎、側顎）分別重新訓練微笑圖像的實例分割模型，並評估比較其辨識效果，希望能提升各辨識目標精度 (AP) 的分割模型。最後開發一套易於使用自動微笑特徵辨識系統，將辨識結果用以輔助醫師與患者溝通微笑構成與評估微笑訓練成效。

## (三) 文獻回顧與探討

### 1 微笑的美觀構成

關於研究中微笑的美觀構成，目前已經有許多論文對微笑的構造提出數種美觀參數以量化，並透過問卷調查得知不同族群對這些參數的偏好<sup>[2][6][7]</sup>，例如在 Pasukdee 等人的研究中探討了矯正醫師、一般牙醫師、矯正患者、一般大眾分別對數個美觀參數的偏好，其中包含<sup>[2]</sup>：

**Table I.** Definitions and range of values of smile esthetic variables

Smile variable	Range of values	Definition
Arc ratio <sup>a,b,c</sup>	-1, -0.5, 0, 0.5, and 1	The perpendicular distance from the incisal edge of maxillary central incisors to a line connecting the cusp tips of the maxillary canine divided by the distance between a tangent line of the upper border of the lower lip and the maxillary intercanine line
Most posterior maxillary teeth visible <sup>a,b,c</sup>	Canine (3), first premolar (4), second premolar (5), first molar (6), and second molar (7)	The most posterior maxillary teeth exposed on both sides while smiling
Maxillary midline <sup>c</sup>	0, 1, 2, 3, and 4 mm	Maxillary dental midline (measured between central incisors) compared with the facial midline (a line representing the midline of the face as defined by the center of the philtrum and nasal tip)
Maxillary teeth exposure <sup>a,b,c</sup>	50%, 75%, 100%, 125%, and 150%	Height of visible maxillary central incisor on smiling divided by the actual height of maxillary central incisor
Mandibular teeth exposure <sup>a,b,c</sup>	0%, 25%, 50%, 75%, and 100%	Height of visible mandibular central incisor while smiling divided by the actual height of mandibular central incisor (measured from the maxillary incisal edge to gingival margin of mandibular incisors)
Buccal corridor <sup>a</sup>	0%, 5%, 10%, 15%, and 20%	The horizontal distance of the total dark space on both sides of the mouth divided by the total smile width
Maxillary gingival display <sup>a</sup>	0, 1, 2, 3, and 4 mm	Height of visible gingiva between the lower border of the upper lip and gingival zenith of the maxillary central incisors
Cant <sup>a</sup>	0°, 3°, 6°, 9°, and 12°	Amount of rotation in the maxillary teeth from the horizontal plane through the middle of the maxillary central incisors
Upper lip thickness <sup>b</sup>	4, 5, 6, 7, and 8 mm	The vertical distance from the most superior point of the cupid's bow to the most inferior portion of the tubercle of the upper lip
Mouth angle <sup>b</sup>	20°, 25°, 30°, 35°, and 40°	The angle formed by the most inferior border of the upper lip and most superior border of the lower lip on smiling
Nasolabial angle <sup>b</sup>	70°, 80°, 90°, 100°, and 110°	The angle between columella of the nose and anterior surface of the upper lip

<sup>a</sup>Variables in frontal view; <sup>b</sup>Variables in three-quarter view; <sup>c</sup>Variables in lateral view.

Fig 1: 微笑美觀參數定義 [2]

這些美觀參數多可從患者的正面、側面的微笑影相中獲得，為了自動辨識這些參數，本研究希望能透過深度學習模型，自動從微笑照片中的分割以下特徵：鼻子、上唇、下唇、各顆牙齒、牙齦與 Buccal corridor

## 2 DSD 軟體

根據 Coachman 等人的研究，將會拍攝患者正面、側面照片用作 DSD 分析<sup>[9]</sup>。而在 Jaffri 等人的研究中，他們操作了現今常見的 12 種 DSD 軟體，認為這些軟體確實能幫助面部美學視覺化。他們不僅能幫助患者了解治療結果，也能幫助醫師診斷與治療<sup>[8]</sup>。但是這些軟體仍需要臨床醫師標註參考線、牙齒位置等，仍不能自動分析微笑美觀構成。

Fig 2: 正面照<sup>[9]</sup>Fig 3: 側面照<sup>[9]</sup>

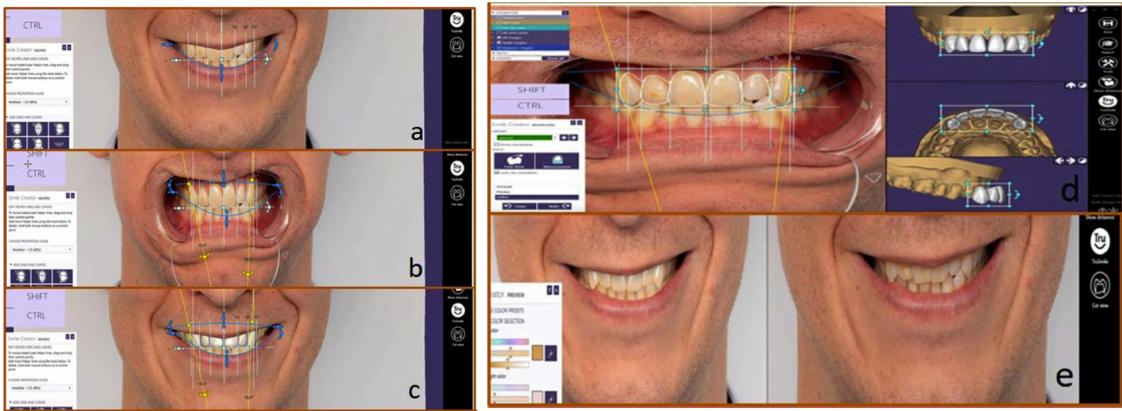


Fig 4: DSD 軟體操作介面<sup>[8]</sup>

### 3 實例分割

在實例分割研究上經典的 Mask R-CNN 模型是基於檢測框的雙階段算法，在訓練與辨識皆需要等待檢測框<sup>[20]</sup>。而在單階段算法上，YOLOACT 模型移除了 localization step 來加速，實現實時實例分割 (Real-time Instance Segmentation)，但精度較差<sup>[21]</sup>。而 YOLOACT++ 是在 YOLOACT 算法的基礎上改善精度，根據 Bolya 等人的研究結果，YOLOACT++ 可在 COCO 資料集取得 33.5fps 下獲得 34.1mAP 的成績。而跟據 Chen 等的研究成果，BlendMask 模型相較於 YOLOACT 雖然速度較慢，但提升了精度，在 ResNet-101 骨架下能在 COCO 資料集取得 37.8 的 AP 值<sup>[18]</sup>。

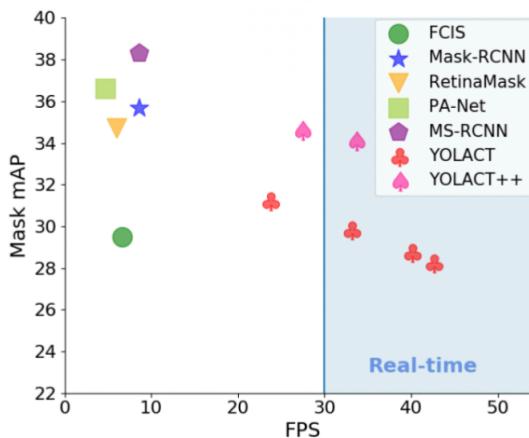


Fig 5: 各模型在 COCO 資料及的測試結果<sup>[19]</sup>

Method	Backbone	Epochs	Aug.	Time (ms)	AP	AP <sub>50</sub>	AP <sub>75</sub>	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>
Mask R-CNN [13]	R-50	12		97.0	34.6	56.5	36.6	15.4	36.3	49.7
Mask R-CNN*		72	✓	97+	36.8	<b>59.2</b>	39.3	17.1	38.7	52.1
TensorMask [7]		72	✓	400+	35.5	57.3	37.4	16.6	37.0	49.1
BlendMask		12		78.5	34.3	55.4	36.6	14.9	36.4	48.9
BlendMask		36	✓	78.5	37.0	58.9	39.7	17.3	39.4	52.5
BlendMask*		36	✓	<b>74.0</b>	<b>37.8</b>	58.8	<b>40.3</b>	<b>18.8</b>	<b>40.9</b>	<b>53.6</b>
Mask R-CNN	R-101	12		118.1	36.2	58.6	38.4	16.4	38.4	52.1
Mask R-CNN*		36	✓	118+	38.3	61.2	40.8	18.2	40.6	54.1
TensorMask		72	✓	400+	37.3	59.5	39.5	17.5	39.3	51.6
SOLO [26]		72	✓	-	37.8	59.5	40.4	16.4	40.6	54.2
+deform convs [26]		72	✓	-	40.4	62.7	43.3	17.6	43.3	58.9
BlendMask		36	✓	101.8	38.4	60.7	41.3	18.2	41.5	53.3
BlendMask*		36	✓	<b>94.1</b>	39.6	61.6	42.6	22.4	42.2	51.4
+deform convs (interval = 3)		60	✓	105.0	<b>41.3</b>	<b>63.1</b>	<b>44.6</b>	<b>22.7</b>	<b>44.1</b>	<b>54.5</b>

Fig 6: BlendMask 在 COCO 資料集的測試結果<sup>[18]</sup>

在 Lee & Kim(2022) 的研究<sup>[16]</sup> 中，使用 NVIDIA 公司發布的 Flickr-Faces-HQ (FFHQ) 資料集的 5500 張圖片進行訓練，其中包含各種角度的成人臉孔。並對眼睛、眉、鼻子、上唇、下唇、耳屏、牙齦、頰走廊、各顆牙齒標註。在實例分割模型選擇上，他們使用 YOLACT++ 架構，基於 ResNet-101 骨架和 FPN。最後對每個特徵的辨識結果計算平均精度 (AP) 以定量評價模型品質。其中平均精度 (AP) 是透過精確率-召回率曲線得到，綜合考慮精度、召回率、IoU 得到的指標。最後在 IoU 域值 0.5 下得到在眼睛、鼻子、嘴唇都獲得高於 0.9 的 AP 值，在 t13-t23 的牙齒範圍取得 80% 以上的準確率。以下為其成果演示：

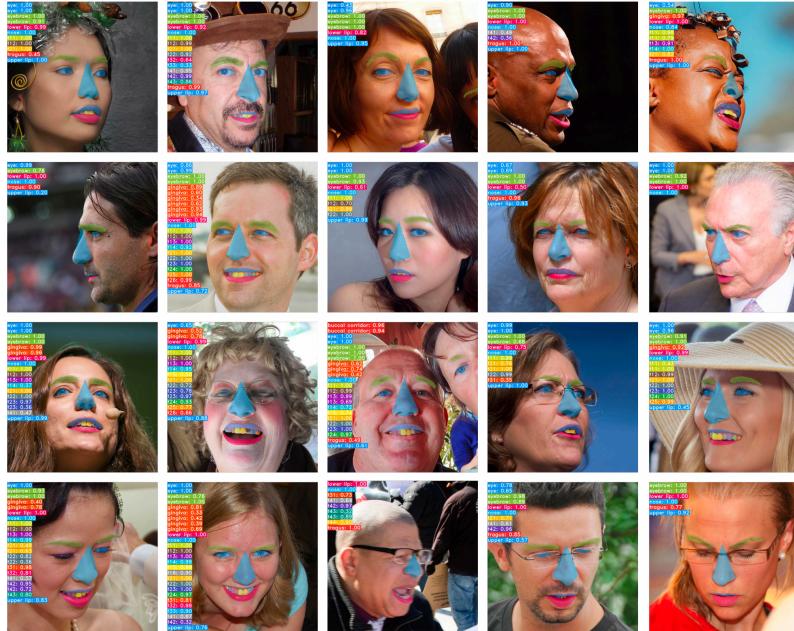


Fig 7: YOLACT++ 分割結果<sup>[16]</sup>

Type	AP <sub>all</sub> (AP <sup>50</sup> <sub>all</sub> )	AP <sub>eye</sub> (AP <sup>50</sup> <sub>eye</sub> )	AP <sub>nose</sub> (AP <sup>50</sup> <sub>nose</sub> )	AP <sub>lips</sub> (AP <sup>50</sup> <sub>lips</sub> )	AP <sub>teeth</sub> (AP <sup>50</sup> <sub>teeth</sub> )
Box	0.341 (0.635)	0.621 (0.946)	0.879 (0.990)	0.645 (0.942)	0.303 (0.604)
Mask	0.229 (0.472)	0.570 (0.945)	0.855 (0.990)	0.541 (0.921)	0.175 (0.411)

Fig 8: Lee & Kim(2022) 訓練的模型 AP 值<sup>[16]</sup>

這說明了基於 YOLACT++ 的實例分割模型確實能在微笑美觀構成之圖像分割上取得良好效果。但是因為 Lee & Kim 的訓練資料是各個角度的人臉照片，牙齒的資訊較少，所以在牙齒的分個效果較差（如 Fig 8，在IoU 閾值 0.5 下，辨識框與分割的 AP 值分別 0.604, 0.411）。但是在 DSD 的使用場景中，2D 圖像的微笑分析對象是固定角度拍攝的，而且比起辨識的速度，更關心辨識的精準度<sup>[8][9]</sup>。因此本研究希望只使用固定角度的微笑影像進行訓練，並且比較 BlendMask 模型與 YOLACT++ 模型的精度 (AP)，本研究計畫假設能在符合 DSD 臨床使用的情況下，得到好的辨識效果。

## (四) 研究方法及步驟

### 1 數據集準備

微笑照片的來源分為兩部分，首先是向患者搜集，再來是使用 AI 生成的人臉照片（來自 <https://generated.photos/> 或是 <https://this-person-does-not-exist.com/>）。

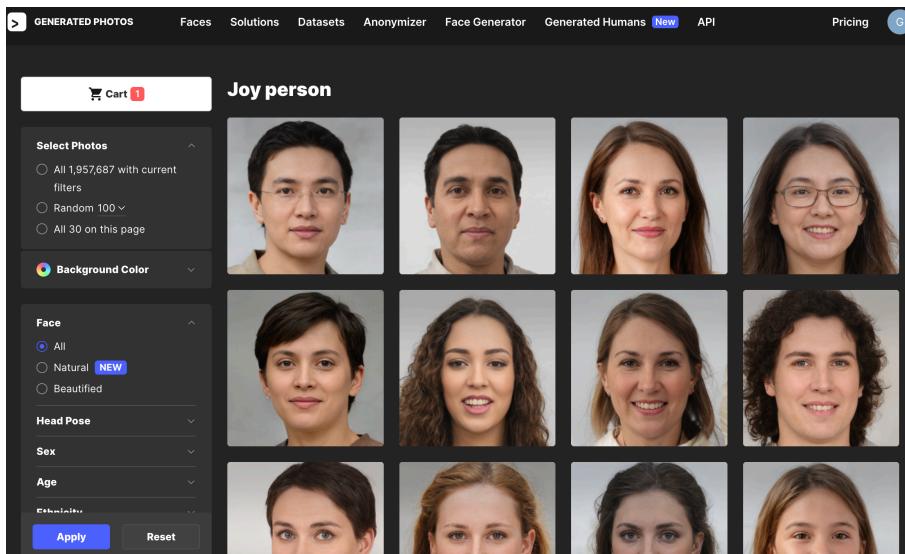


Fig 9: generated.photos 生成的人臉

## 2 數據標註

將使用 Labelme 以多邊形標註待辨識的特徵。其中包含：上嘴唇、下嘴唇、各顆牙齒、牙齦和 Buccal corridor。

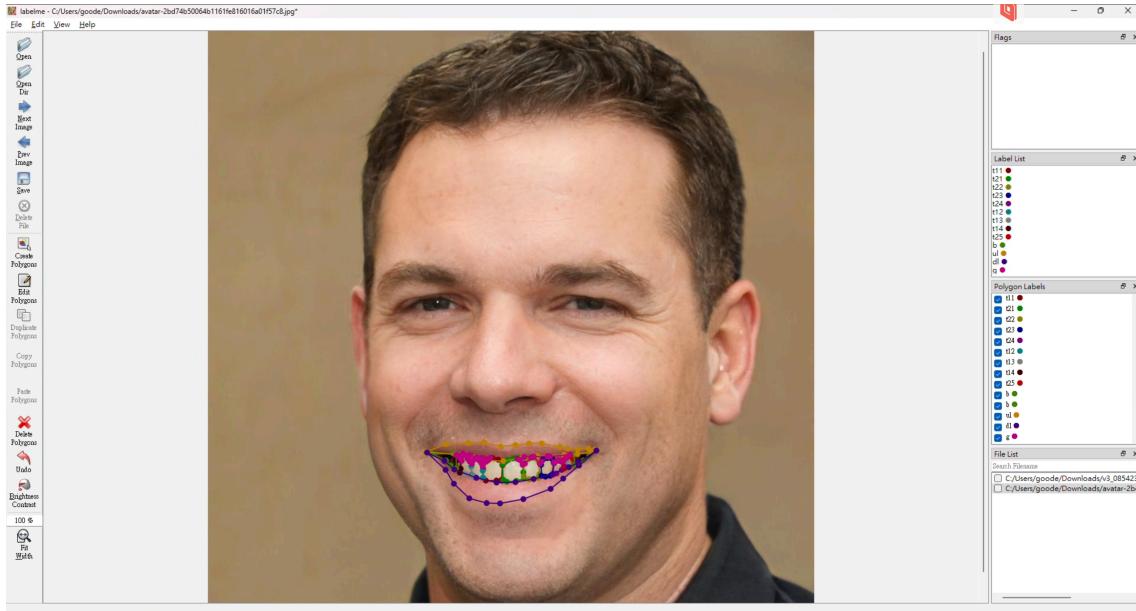


Fig 10: 使用 labelme 標註，照片來自 <https://this-person-does-not-exist.com/> 生成

## 3 模型訓練

研究中使用的 BlendMask 模型與 YOLACT++ 模型都是透過 python3, pytorch 10.0.1, Detectron2 實現。而 CNN 模型訓練通常不會從頭開始訓練，否則訓練時間太長且需要的資料集過大，因此從他人訓練好的權重開始訓練。

### 1) YOLACT++

因為希望獲得較高精度，選擇 Resnet101-FPN 作為架構 (backbone)<sup>[19]</sup>。根據 YOLACT++ 原論文作者的 github 介紹 (<https://github.com/dbolya/yolact>)，權重將繼承自作者提供的權重，輸入圖片大小為  $550 \times 550$ 。

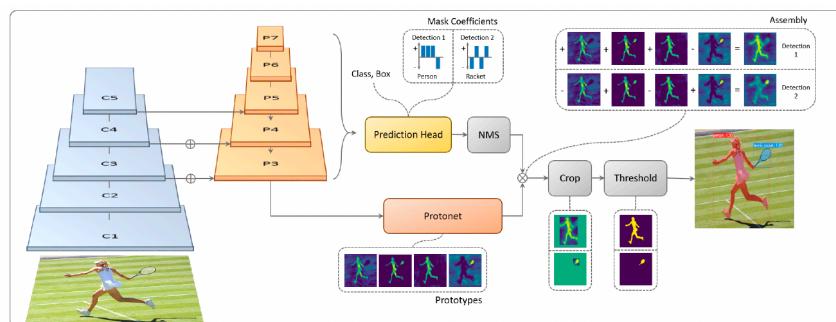


Fig 11: YOLACT++ 模型<sup>[19]</sup>

## 2) BlendMask

因為希望獲得較高精度，選擇 R\_101\_dcni3\_5x 作為架構 (backbone)<sup>[18]</sup>。根據 BlendMask 原論文作者的 github 介紹 (<https://github.com/aim-uofa/AdelaiDet/tree/master/configs/BlendMask>)，需要使用 AdelaiDet 在 Detectron2 執行訓練<sup>[22]</sup>。

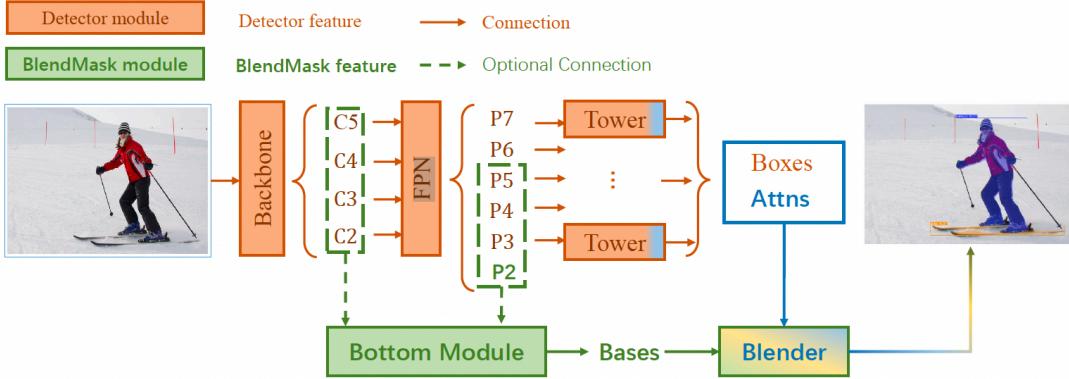
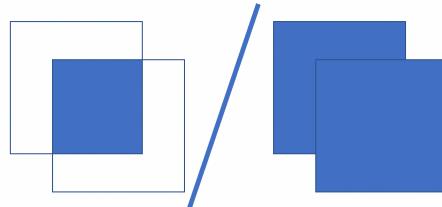


Fig 12: BlendMask 模型<sup>[18]</sup>

## 4 統計評估

### Definition 4.1. IoU

$$\text{IoU} = \frac{\text{交集面積}}{\text{連集面積}}$$



### Definition 4.2. 精確率 (Precision)、召回率 (Recall)

若以信度大於 95% 作為閾值， $\text{IoU} > 0.5$  作為正確 (True)：

	信度大於 0.95 的辨識框	信度不大於 0.95 的辨識框
$\text{IoU} > 0.5$	TP	FP
$\text{IoU} \leq 0.5$	FN	TN

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

### Definition 4.3. 精度 (AP)

將辨識出來的候選框依信度由高至低依序作為信度閾值，可得到一系列精確率、召回率。繪製精確率-召回率圖形 (Fig 13 藍線) 即可由曲線下面積得到。在計算時會使用橘色線 ( $P_{smooth}$ ) 近似。

$$AP = \int_0^1 P_{smooth}(r)dr$$

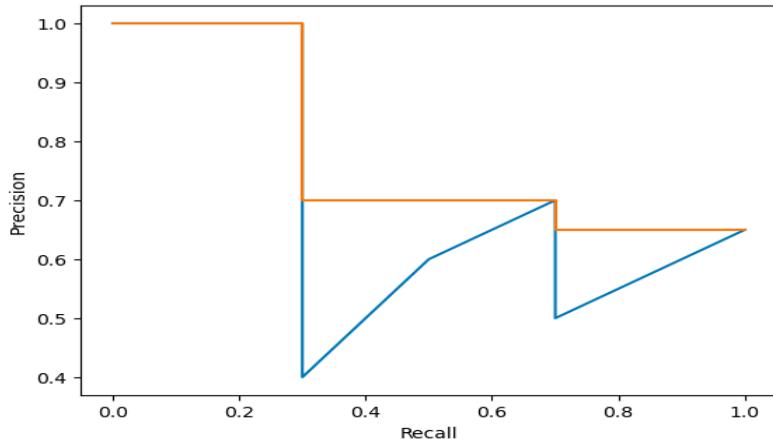


Fig 13: 精確率-召會率圖

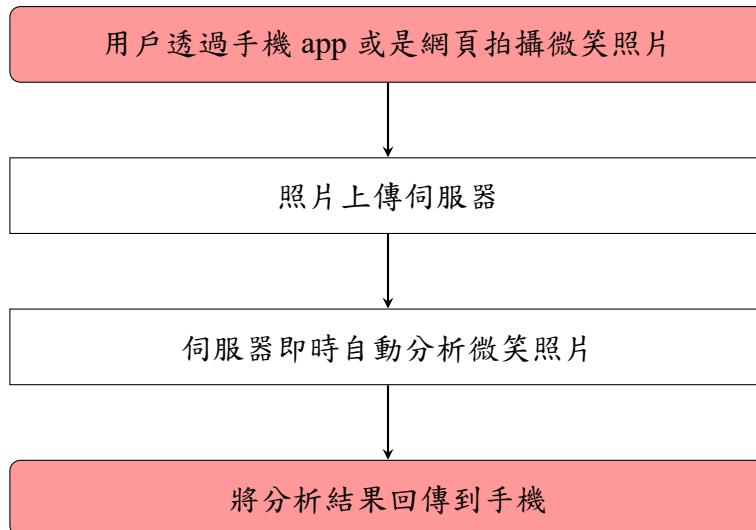
## (五) 預期結果

### 1 分割模型

我們將會針對正面與側面照 (如 Fig 2, Fig 3)，使用 YOLACT++ 和 BlendMask 算法訓練實例分割模型，並比較分析其辨識結果。希望能比 Lee, Seulgi & Kim, Jong-Eun 得到更好的精度 (AP)<sup>[16]</sup>

### 2 微笑分析系統

在完成模型後，我們希望開發一套手機拍照後即可自動對微笑照片進行分割，再計算相關美觀參數的系統。該系統將使用 Python 與 Flask 在伺服器架設 api，用戶拍照後即上傳照片到伺服器分析。該系統將用於即時向患者視覺化呈現微笑訓練、矯正的結果，也能輔助醫師進行診斷與治療。



## (六) 需要指導教授指導內容

### 1 微笑照片的搜集

我沒有搜集醫學資料的經驗，對於該如何搜集資料與搜集資料該注意的問題，如該如何取得患者同意等，我並不熟悉。因此在搜集微笑照片上，需要指導教授的協助。

### 2 臨床知識

因為本計畫目標將辨識成果應用在臨牀上，但臨床知識是我所缺乏的。因此這方面需要指導教授的幫助，指導我怎麼樣的使用情境是臨牀上需要的，而我又該達到怎麼樣的辨識精度或該實現怎麼樣的功能。

hi

## 參考資料

- [1] Guilherme Janson, Nuria Castello Branco, Thais Maria Freire Fernandes, Renata Sathler, Daniela Garib, and José Roberto Pereira Lauris. Influence of orthodontic treatment, midline position, buccal corridor and smile arc on smile attractiveness: A systematic review. *The Angle Orthodontist*, 81(1):153–161, 2011.
- [2] Parichart Pasukdee, Johnson Hsin-Chung Cheng, and De-Shing Chen. Smile preferences of orthodontists, general dentists, patients, and the general public in three-quarter and lateral views. *American Journal of Orthodontics and Dentofacial Orthopedics*, 159(4):e311–e320, 2021.
- [3] John P Frush and Roland D Fisher. The dynesthetic interpretation of the dentogenic concept. *The Journal of prosthetic dentistry*, 8(4):558–581, 1958.
- [4] Charles M Hulsey. An esthetic evaluation of lip-teeth relationships present in the smile. *American journal of orthodontics*, 57(2):132–144, 1970.
- [5] David Sarver and Ronald S Jacobson. The aesthetic dentofacial analysis. *Clinics in plastic surgery*, 34(3):369–394, 2007.
- [6] Vincent O Kokich Jr, H Asuman Kiyak, and Peter A Shapiro. Comparing the perception of dentists and lay people to altered dental esthetics. *Journal of Esthetic and Restorative Dentistry*, 11(6):311–324, 1999.
- [7] Hideki Ioi, Sangwook Kang, Takahiro Shimomura, Seong-sik Kim, Soo-byung Park, Woo-sung Son, and Ichiro Takahashi. Effects of buccal corridors on smile esthetics in japanese and korean orthodontists and orthodontic patients. *American Journal of Orthodontics and Dentofacial Orthopedics*, 142(4):459–465, 2012.
- [8] Zeba Jafri, Nafis Ahmad, Madhuri Sawai, Nishat Sultan, and Ashu Bhardwaj. Digital smile design—an innovative tool in aesthetic dentistry. *Journal of oral biology and craniofacial research*, 10(2):194–198, 2020.

- [9] Christian Coachman, Marcelo Alexandre Calamita, and Newton Sesma. Dynamic documentation of the smile and the 2d/3d digital smile design process. *Int J Periodontics Restorative Dent*, 37(2):183–193, 2017.
- [10] Carole Charavet, Jean-Claude Bernard, Cyril Gaillard, and Michel Le Gall. Benefits of digital smile design (dsd) in the conception of a complex orthodontic treatment plan: A case report-proof of concept. *International orthodontics*, 17(3):573–579, 2019.
- [11] Hongqiang Ye, Kuan-Paul Wang, Yushu Liu, Yunsong Liu, and Yongsheng Zhou. Four-dimensional digital prediction of the esthetic outcome and digital implementation for rehabilitation in the esthetic zone. *The Journal of Prosthetic Dentistry*, 123(4):557–563, 2020.
- [12] Kele Xu, Dawei Feng, and Haibo Mi. Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules*, 22(12):2054, 2017.
- [13] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics & vision (ICARCV)*, pages 844–848. IEEE, 2014.
- [14] Jiahong Qian, Ming Cheng, Yubo Tao, Jun Lin, and Hai Lin. Cephanet: An improved faster r-cnn for cephalometric landmark detection. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 868–871. IEEE, 2019.
- [15] Dima M Alalharith, Hajar M Alharthi, Wejdan M Alghamdi, Yasmine M Alsenbel, Nida Aslam, Irfan Ullah Khan, Suliman Y Shahin, Simona Dianišková, Muhanad S Alhareky, and Kasumi K Barouch. A deep learning-based approach for the detection of early signs of gingivitis in orthodontic patients using faster region-based convolutional neural networks. *International Journal of Environmental Research and Public Health*, 17(22):8447, 2020.
- [16] Seulgi Lee and Jong-Eun Kim. Evaluating the precision of automatic segmentation of teeth, gingiva and facial landmarks for 2d digital smile design using real-time instance segmentation network. *Journal of Clinical Medicine*, 11(3):852, 2022.
- [17] Johnson Hsin-Chung Cheng, Mingming Jareerat Luechapanichkul, and Tracy Yi-Hsuan Lee. The relationship between dentofacial morphology and smile characteristics in lateral and oblique views. *Journal of Dental Sciences*, 16(1):37–44, 2021.
- [18] Hao Chen, Kunyang Sun, Zhi Tian, Chunhua Shen, Yongming Huang, and Youliang Yan. Blendmask: Top-down meets bottom-up for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8573–8581, 2020.

- [19] Chong Zhou. *Yolact++ Better Real-Time Instance Segmentation*. University of California, Davis, 2020.
- [20] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [21] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact: Real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9157–9166, 2019.
- [22] Zhi Tian, Hao Chen, Xinlong Wang, Yuliang Liu, and Chunhua Shen. AdelaiDet: A toolbox for instance-level recognition tasks. <https://git.io/adelaidet>, 2019.