

# 獨立研究 作品說明書

科 別：電腦與資訊學科

組 別：高級中等學校組

作品名稱：生成人臉口罩去除圖像

關 鍵 詞：臉部合成、影像處理、類神經網路

# 摘要

新冠肺炎疫情期間，拍攝不少戴口罩的照片，但人們希望能保留未戴口罩照片，為了解決這個問題，本研究建構人臉口罩去除系統，給予無口罩覆蓋的來源圖片作為參考，透過擷取人臉和口罩輪廓的特徵點，進行人臉置換、圖像填補與色彩優化等步驟，對於戴口罩圖片的口罩區域進行人臉復原。

本研究透過校正戴口罩人臉特徵點的誤差，設計改良演算法精準地偵測口罩區域，利用輪廓偵測演算法擷取口罩輪廓特徵點框出區域，將來源圖片的對應區域置換到戴口罩圖片上，並使用圖像填補技術修復填補口罩置換後的殘影，再對圖片置換的拼接處進行色彩優化，讓整體臉部膚色具一致性，經成效評估驗證，成功自然地將戴口罩圖片復原為未戴口罩的真實樣貌。

## 壹、前言

### 一、研究動機

在新冠肺炎的影響下，在各個地方都會戴上口罩，使得拍照出來的人像照臉部往往被口罩覆蓋，只剩下眼睛可供辨別，不但難以辨識，照片也不甚美觀，而過去或未來在一些重要典禮及場合拍攝的照片，若是人像戴著口罩，不免有些可惜。

如果是給定另外一張沒有戴口罩的人像圖，是否可以擷取其臉部特徵，將人臉自然地置換到戴口罩的圖片上，使原本戴口罩的圖片還原成一張沒有戴口罩的圖片，如此一來，如果就算臉都被口罩覆蓋住，也可以藉由本研究把圖片中的口罩部分移除，。

### 二、研究目的

透過給予一張無口罩覆蓋的人像圖片作為參考，對於戴口罩的人像圖片進行修補，使其還原成沒有戴口罩且自然的人臉。研究目的可分為以下列點：

- (一) 整合改良 Grabcut 與 Region Growing 演算法，分割圖片中的口罩區域，並進行 Canny 輪廓偵測，擷取口罩輪廓特徵點。
- (二) 研究並撰寫人臉置換模型，參考特徵點將來源圖片與戴口罩圖片的口罩區域置換。
- (三) 研究圖像填補技術並訓練模型，輸入口罩殘影處遮罩，生成去除瑕疵的人臉圖像。
- (四) 使用 Reinhard 經典色彩遷移算法和泊松圖像編輯，優化人臉色彩，使得輸出圖片的膚色具一致性。

### 三、 影像處理相關理論

#### （一）生成對抗網路

生成對抗網路（Generative Adversarial Network, GAN）（許恆修，民 108），是由生成器（Generator）、判別器（Discriminator）兩個模型所組成，生成器  $G$  從樣本資料取得符合某一分布（均勻分布、高斯分布等）的雜訊  $Z$ ，生成盡可能逼真的樣本，騙過判別器  $D$ ，使生成樣本與真實樣本之間的差異最小化；判別器  $D$  輸入真實樣本以及生成器  $G$  的輸出，目的將生成器  $G$  生成的樣本從真實樣本儘可能分辨出來，使公式最大化（識別能力強）。透過兩個網路模型的相互對抗、不斷調整參數，使判別器無法判斷生成網路輸出結果是否真實。上述過程如下列公式及示意圖：

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))]$$

$G$  為生成器， $D$  為判別器；其中  $x$  為真實數據， $z$  代表了隨機輸入數據，求取期望  $E$ ， $P_{data}$  為真實數據分布， $P_z$  為潛在變量  $z$  數據分布。

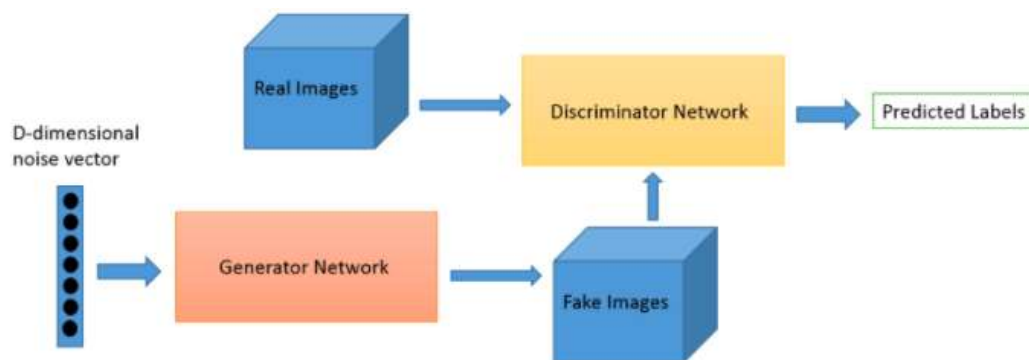


圖 1、生成對抗網路（GAN）模型示意圖（許恆修，民 108）

#### （二）卷積神經網路

卷積神經網路（Convolutional Neural Network, CNN）（Tommy Huang, 2018）是一種前饋神經網路，參數從輸入層向輸出層單向傳播，由一個或多個的卷積層（Convolution Layer）、全連通層（Fully Connected Layer）、跟權重相關的池化層（Pooling Layer）組成。

圖片在經過卷積層的卷積運算後，將各特徵投影到特徵圖上，再經由池化層的池化運算，降低特徵圖的維度，此時得到的特徵圖為一個 2-D 的圖片，轉為 1-D 的陣列後，將特徵輸入到全聯通層，進行仿射變換及分類。

卷積神經網路可以針對資料做特徵的萃取，增強了神經網路的效率，因此，相較於其他深度神經網路，在圖像辨識和語音辨識方面有更好的成效。

### 1. 卷積層

卷積層由許多卷積核（Kernel）組成，在輸入圖像上滑動不同的卷積核，並在每一個滑動的位置上，與輸入圖像對應元素之間進行卷積運算，乘積求和、重複動作以歸納出輸入圖片中各種可能的特徵，將每個卷積計算後的值和位置投影到一個元素，製作一個新的二維矩陣，而這新的二維矩陣又稱為特徵圖（Feature Map）。

### 2. 池化層

池化層會根據特徵圖將的結果去做池化運算，把輸入圖片最大池化，即將輸入圖片分成數個矩形，對每個子區域輸出最大值，亦可做平均池化或最小池化，目的為減少圖片資料量並保留重要資訊，得到降維的特徵圖。

### 3. 全連通層

做完卷積運算和池化運算後得到的特徵圖是一個 2-D 的圖片，先轉成 1-D 的陣列，再將卷積層與池化層輸出的特徵輸入到全連接層，透過仿射變換—乘以一個矩陣、加上一個偏移量（向量加上一個固定值或學習得到的偏差量），調整權重及偏差，得到分類的結果。

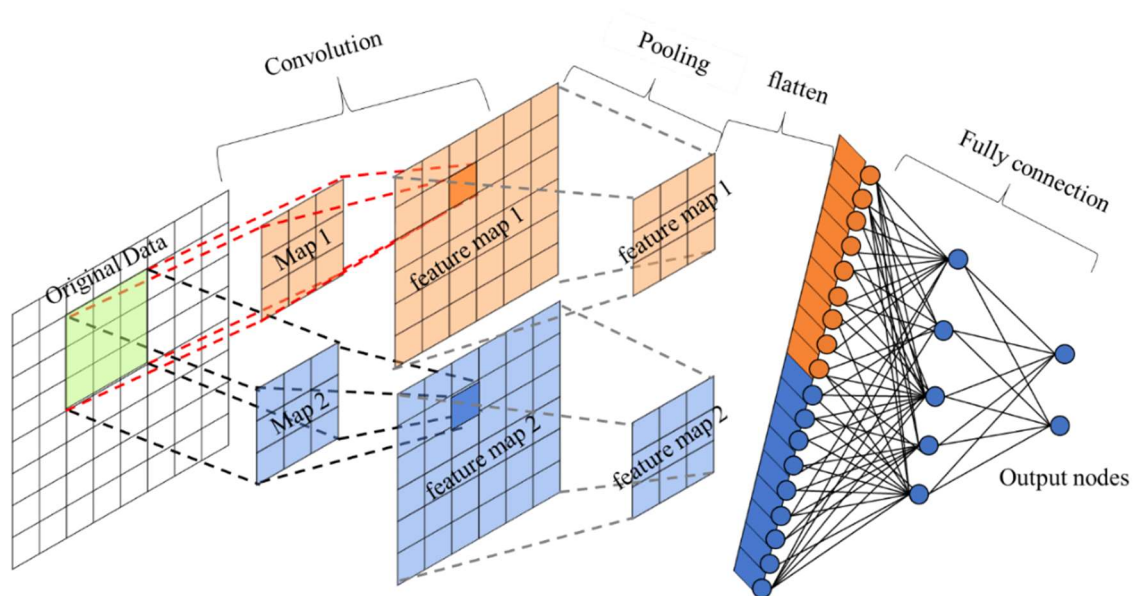


圖 2、卷積神經網路架構示意圖（Tommy Huang, 2018）

### (三) Image Fine-grained Inpainting

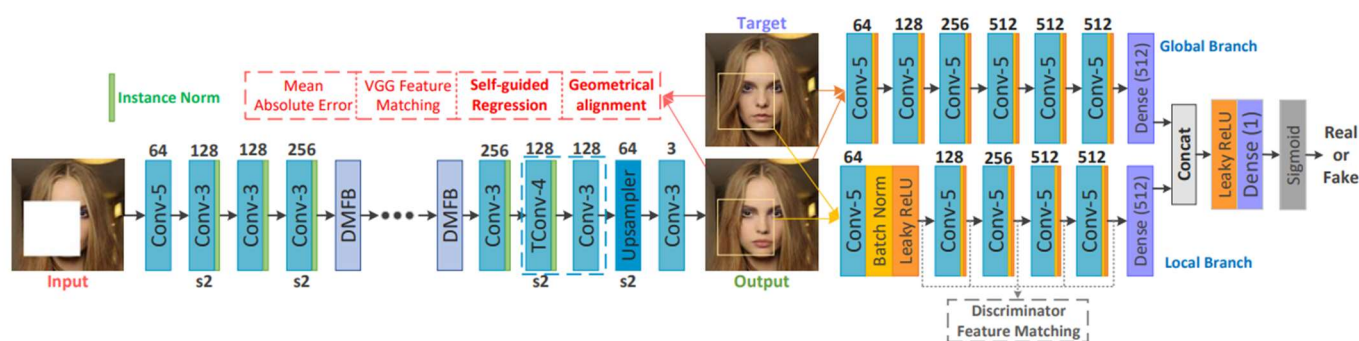


圖3、生成對抗網路架構圖 ( Hui, Li, Wang & Gao, 2020, p. 4 )

Hui et al. (2020)為生成對抗網路（Generative Adversarial Network）架構，在生成器中採用不同的擴張卷積（Dilated Convolution）提取的特徵的結合，獲得更好的多尺度特徵。擴張卷積是在標準的卷積核中注入空洞，以增加感受域（Reception Field）。透過訓練生成器與判別器相互對抗，生成盡可能真實的圖片。

以下是訓練後的生成圖片，圖左為輸入圖片，圖中為輸出圖片，圖右為目標圖片：



圖 4、訓練模型生成圖 ( Hui et al., 2020, p. 7 )

由於缺乏真實人像部分區域，導致容易缺失原人像的人臉身分訊息，本研究透過提供另一張圖像作為參考，保留人像的人臉身分特徵。

## 貳、研究設備與器材

### 一、 硬體設備：

筆記型電腦（i5-1035G4 CPU @ 1.10GHz 1.50 GHz、GPU、記憶體：8 GB）。

### 二、 軟體及工具環境：

- （一）python：具有高效能的高階資料結構，直譯式且物件導向的程式語言。
- （二）opencv-python：實現圖像處理和計算機視覺方面的應用，可用於解決物件辨識、擴增實境等領域，並提供視窗介面基本功能。
- （三）tensorflow：google 開發的開元機器學習工具，提供使用者進行開發與訓練模型。
- （四）dlib：包含機器學習、電腦視覺、影像辨識等的函式庫，廣泛使用於工業及學術界。
- （五）numpy：提供高效率的大量資料運算，並針對陣列運算提供大量的數學函式庫。
- （六）scipy：基於numpy的科學計算庫，含大量的 python 算法庫和數學工具包。
- （七）pillow：應用於影像處理，提供影像歸檔處理並支援廣泛的檔案格式。

## 參、研究方法及過程

### 一、研究構思與架構

圖 5 為本研究規劃的研究流程。經過相關理論探討後，進行圖片蒐集、資料前處理、擷取人臉特徵點、擷取口罩輪廓特徵點、人臉置換、圖像填補、色彩優化等步驟，最後形成完整的研究流程。

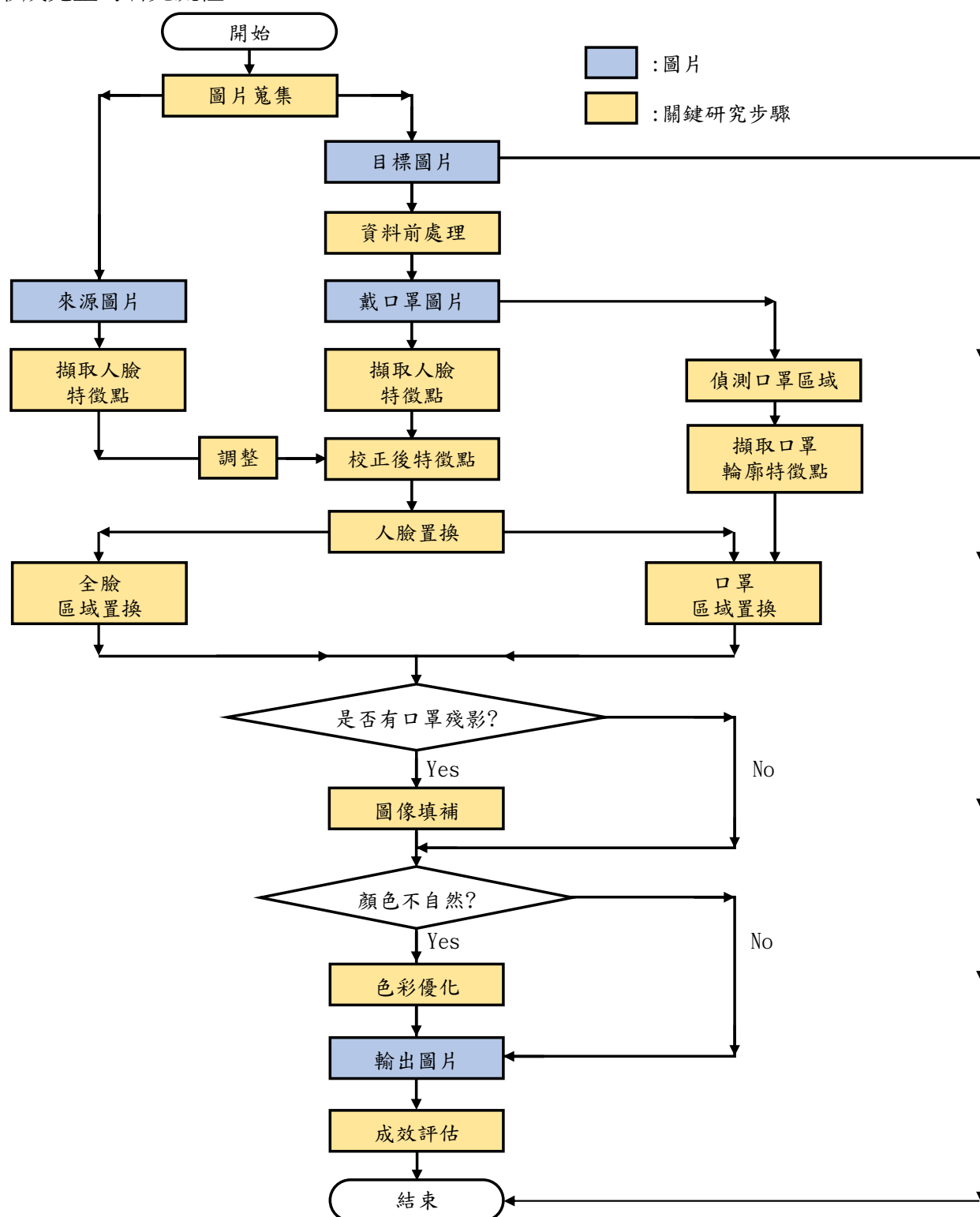


圖 5、研究流程示意圖



## 二、研究過程

### （一）圖片蒐集

#### 1. 圖片集

使用 Caltech Computational Vision Group Archive – Faces 圖片集

（Computational Vision at Caltech, n.d.），其中有 27 個獨特人物的 450 張正面圖像。

#### 2. 目標圖片及來源圖片

從圖片集中每個人物挑兩張背景和人臉位置都稍微不同的圖片作為目標圖片及來源圖片，以符合現實生活中，因為室外須戴上口罩拍照，而室內則可以脫下口罩，所造成的兩圖片背景和人臉特徵及位置的差異。

### （二）資料前處理

#### 1. 戴口罩圖片

本研究使用了 MaskTheFace (Anwar & Raychowdhury, 2020) 模型，其中具備了口罩資料庫，存放多種口罩在不同角度的圖片。MaskTheFace 模型透過偵測目標人臉的關鍵特徵，再依據不同的人臉傾斜角度，選擇了合適的口罩模板貼合目標人臉，其中也提供了不同樣式的口罩，以供本研究分析及討論。

將目標圖片輸入 MaskTheFace 模型，並生成戴口罩圖片，如圖 6 所示。而沒有口罩覆蓋的目標圖片恰可做為檢測輸出圖片成效的對照圖片。

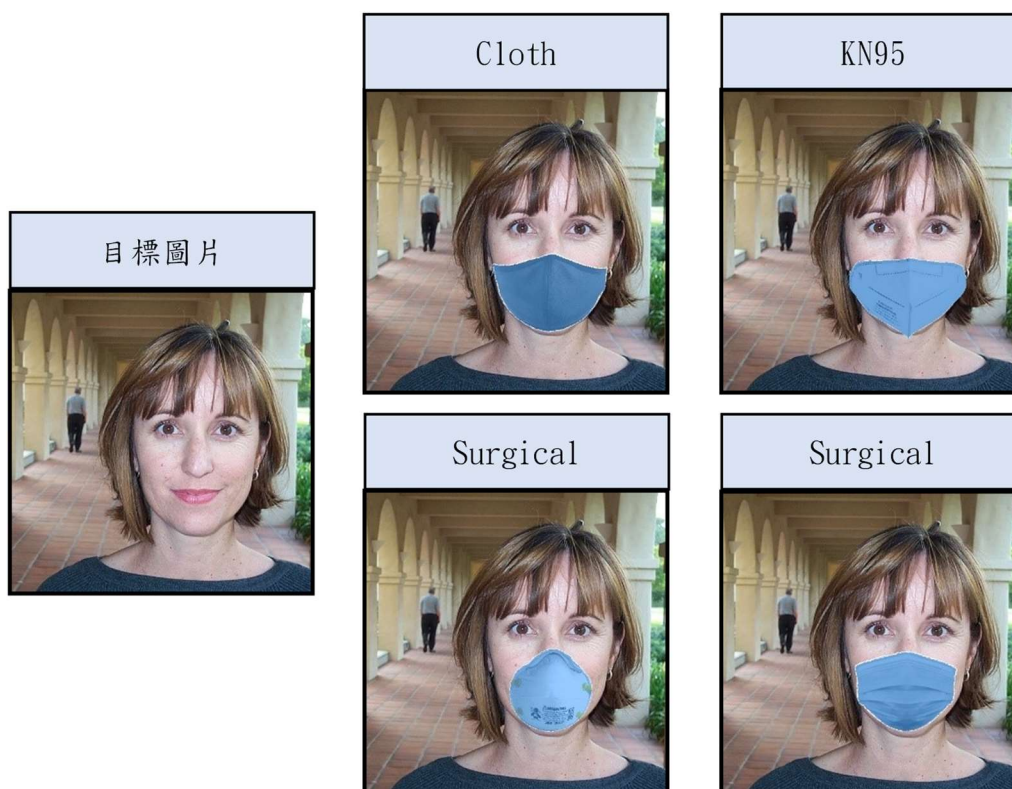


圖 6、不同樣式的戴口罩圖片示意圖



### （三）擷取人臉特徵點

#### 1. 68 個臉部特徵點

在進行人臉處理時需要提取人臉上的特徵訊息，而本研究透過擷取人臉特徵點，可以找出人臉五官及輪廓的位置，並作為圖片中互相比對人臉特徵位置的參考。本研究利用 Python 內部的 Dlib 函式庫以取得兩張圖片中人臉的 68 個臉部特徵點（Facial 68 points landmarks）。

##### （1）Dlib 函式庫

Dlib 為一套包含了機器學習、計算機視覺、圖像處理等的函式庫。本研究將戴口罩圖片及來源圖片輸入 Dlib 官方訓練好的模型 shape\_predictor\_68\_face\_landmarks.dat (Rosebrock, 2017)，對人臉進行偵測並取得圖片中的 68 個臉部特徵點（如圖 7）。

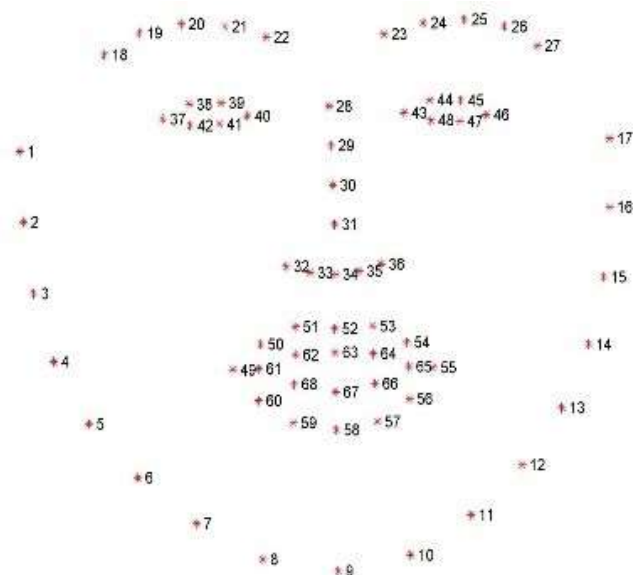


圖 7、Facial 68 points landmarks

##### （2）應用

藉由臉部 68 個特徵點（如圖 8）的座標，進行人臉位置的比對和計算，在之後的步驟中應用。

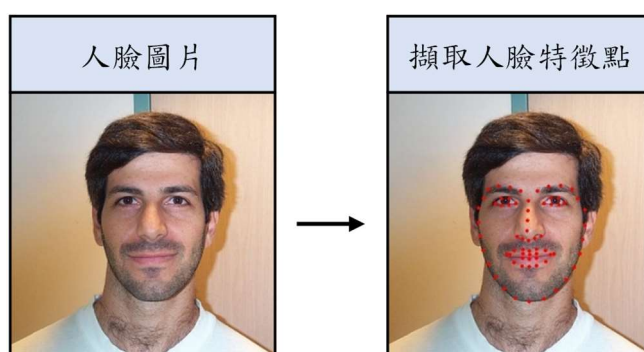


圖 8、擷取人臉特徵點示意圖

## 2. 校正戴口罩圖片的特徵點

### (1) 原因

在擷取戴口罩圖片的臉部特徵點時，會因為口罩覆蓋五官及臉部輪廓，造成圖片中臉部口罩覆蓋處的特徵點偏差，於是本研究參考了來源圖片的臉部特徵點，以校正戴口罩圖片的特徵點。

但因為戴口罩圖片與來源圖片中的人臉姿勢、角度、尺寸不同，即便使用了相同的模型進行特徵點擷取，仍然會產生同一編號特徵點的位置差異，導致口罩覆蓋處的特徵點無法直接採用來源圖片相同位置擷取到的特徵點，而特徵點偏差也會影響後續人臉處理成效。

### (2) 方法

由於戴口罩圖片中，多數特徵點被口罩覆蓋，本研究選用沒有被覆蓋的眼睛作為參考點，分別對兩張圖片的左右眼特徵點取平均，得到眼睛座標（點  $A$ ），再計算兩眼座標的中心點（點  $B$ ），可得到眼睛座標與中心點的距離（ $\overline{AB}$ ），比較兩張圖片眼睛座標與中心點的距離（ $\overline{AB}$ ）的倍率，透過矩陣運算，算出兩張圖片各特徵點（點  $C$ ）與眼睛中心點的距離（ $\overline{BC}$ ）的縮放倍率，進行縮放及位移，校正戴口罩照片的特徵點，如圖 9。

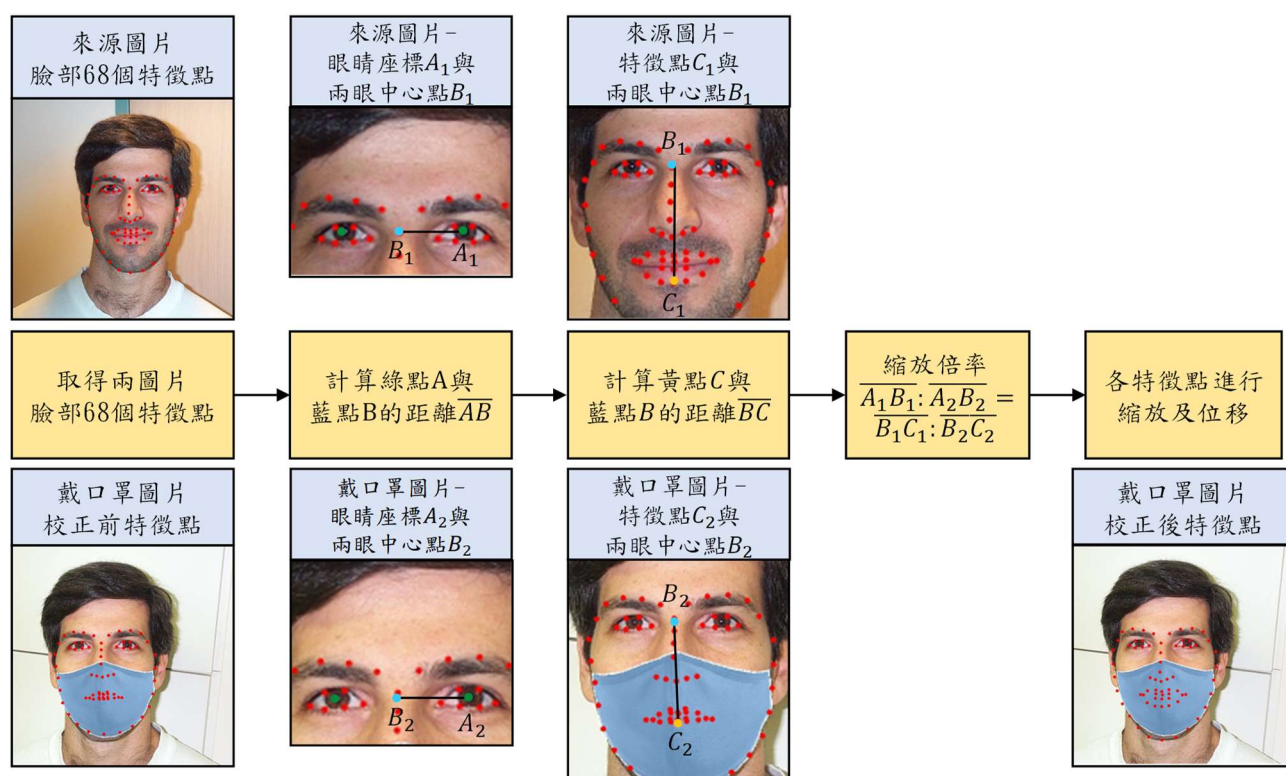


圖 9、戴口罩圖片特徵點校正流程圖

#### （四）擷取口罩特徵點

為了後續針對口罩區域進行兩張圖片的置換，本研究需要偵測口罩區域並擷取口罩輪廓的特徵點。本研究結合 Grabcut 演算法以及Region Growing 演算法，生成口罩的遮罩，再對口罩遮罩進行 Canny 偵測，擷取口罩輪廓特徵點，說明如下：

##### 1. 偵測口罩區域

###### （1）Grabcut 演算法

在偵測過程中，將圖片中目標物體區域視為前景，其餘則是背景。定義目標物體區域的矩形後，矩形外視為背景，而矩形內基於周邊像素顏色上的相似度，將未定義的像素標記為前景或背景，圖片中的每一個像素視為節點，通過虛擬邊與周圍像素連線。在每個節點完成連線後，若節點之間的邊屬於不同終端（一個節點屬於前景，另一個節點屬於背景），則會切斷他們之間的邊，將目標物體區域分割出來。

本研究藉由來源圖片中的人臉下半部輪廓特徵點，來框出包含口罩範圍的矩形，以下巴位置的最低點、臉頰兩側最寬處兩點及鼻樑中心點框出矩形的位罝。

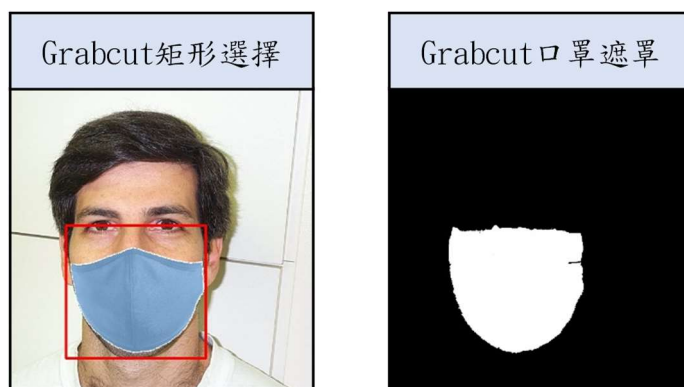


圖 10、Grabcut 演算法之矩形選擇與口罩遮罩

###### （2）Region Growing 演算法

從圖片中選取種子點（Seed point），本研究利用人臉 68 個特徵點中的臉輪廓點的平均值作為種子點，其恰好會落在口罩範圍內，每個種子點都分別往外擴充，持續疊代（Iteration）生長，若周圍的像素與種子點的像素相似度差小於某個閾值（Threshold），則此像素屬於相同的區域，將其設為 1（白色），並將這個像素視為種子點，繼續從向外延展，直到不滿足條件，反之，若周圍的像素與種子點的像素相似度差大於閾值，則此像素不屬於相同的區域，將點設為 0（黑色），並停止擴充。

在處理每個像素後，將圖片依據相似度分成數個區域。相似度表示在灰度級中，觀察在兩個空間鄰接像素之間或像素集合的平均灰度級間的最小差分，判斷像素是否屬於相同的區域。而定義的閾值則影響像素中的最小區域大小。

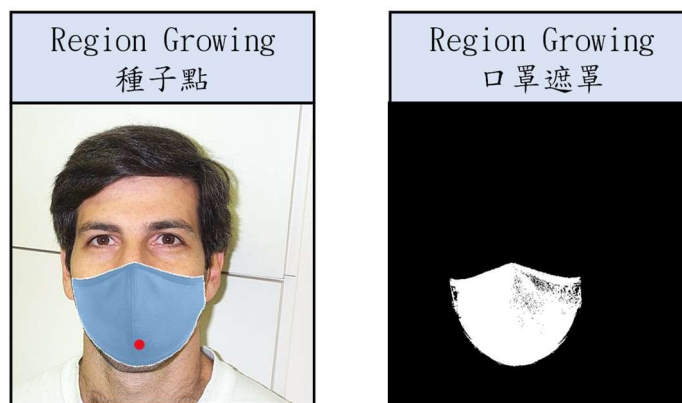


圖 11、Region Growing 演算法之種子點選擇與口罩遮罩

## 2. Grabcut 演算法與 Region Growing 演算法的改良與結合

兩者皆有誤判的情形，因此本研究結合兩者，先利用影像膨脹去除 Region Growing 演算法分割區域內部的雜訊，再將兩圖取交集，以生成精準的口罩遮罩。

### (1) 影像膨脹

影像膨脹將遮罩中的白色區域（像素值為 1）進行擴張，運算後比原圖的白色區域大，可用來填滿隙縫，而膨脹的寬度取決於卷積的大小，卷積沿著圖像滑動並計算，若卷積的範圍內有一個以上的像素值為 1，則此卷積範圍的所有像素值都會變為 1，藉此使得白色區域膨脹，並填滿內部縫隙，以去除雜訊。

本研究經過測試，將卷積大小設為 5x5 時，能最有效的去除口罩遮罩的内部雜訊。

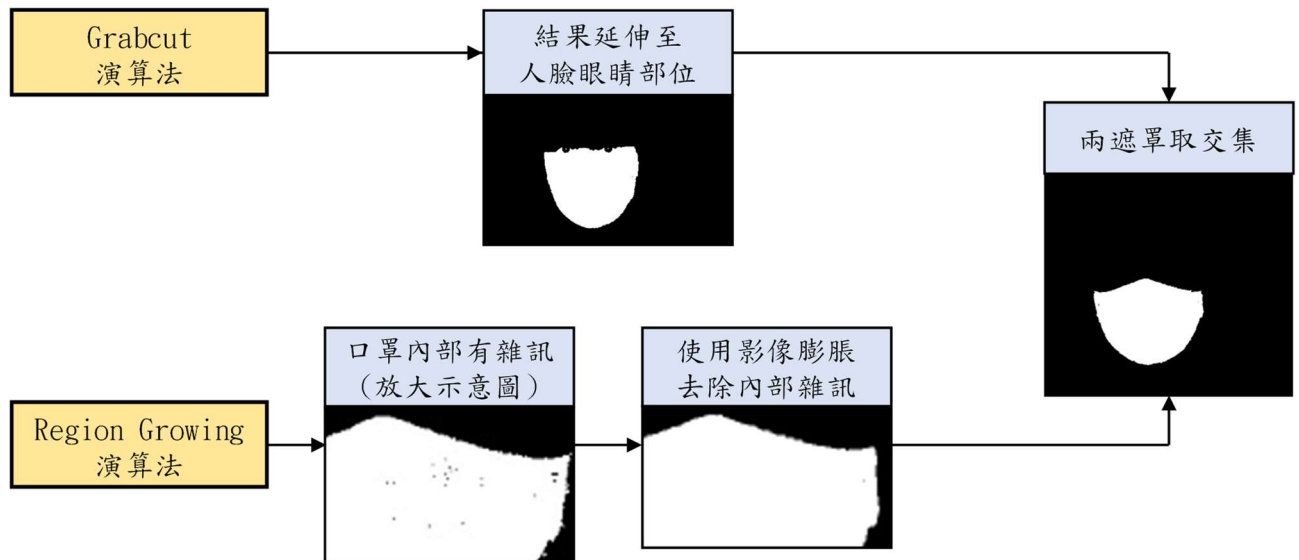


圖 12、Grabcut 與 Region Growing 演算法之口罩遮罩結合流程圖

### 3. 擷取口罩輪廓特徵點

#### (1) Canny 輪廓偵測法擷取特徵點

Canny 輪廓偵測法利用高思濾波以去除圖像中的雜訊，再透過計算像素光度的一階導數（梯度向量）進行邊緣檢測。

相較其他偵測法，Canny 輪廓偵測輸出結果邊緣連續完整，符合本研究需求，應用 opencv 提供的函數，擷取口罩輪廓的特徵點。

### (五) 人臉置換

本研究利用來源圖片中的人臉置換（Face Swapping）（Canu, 2019）到戴口罩圖片上，步驟分成下列部分：

#### 1. 對來源圖片及戴口罩圖片進行三角剖分

##### (1) 原因

考慮到若是將來源圖片的人臉部分直接貼合在戴口罩圖片上，會因為處理面積大而導致置換後人臉不自然、有明顯貼合痕跡。本研究將來源圖片及戴口罩圖片中的人臉區域分割成小面積後，再去進行貼合。

##### (2) 方法

透過口罩區域的特徵點把臉部分割成數個三角形，並使用德勞內三角化（Delaunay Triangulation）（Tseng, 2019），此算法能把平面上的點，分割成具有良好性質的三角形，而良好性質的三角形是指分割的三角形內角最



大化，避免出現「極瘦」的三角形。經由德勞內三角化分割的三角形，可以減少在變形和線性差值時的誤差。

如圖 13、圖 14，兩個全等的四邊形中，圖 13 的  $\overline{BD}$  所分割出的三角形比圖 14 的  $\overline{EG}$  分割出的三角形還要好，而德勞內三角化就會形成圖 13 的分割，使得第四點不會處於三角形外接圓的內部（如圖 15）（點  $C$  在外接圓  $ABC$  外，點  $A$  也在外接圓  $BCD$  外），相反的，不具良好性質的三角形（如圖 14），其第四點處於三角形外接圓的內部（如圖 16）（點  $H$  在外接圓  $EFG$  外，點  $F$  也在外接圓  $EGH$  外）。

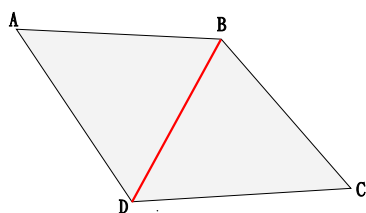


圖 13、具良好性質的三角形

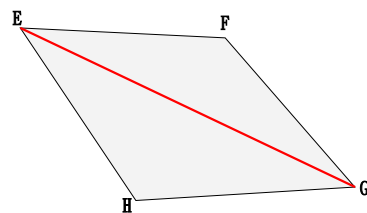


圖 14、不具良好性質的三角形

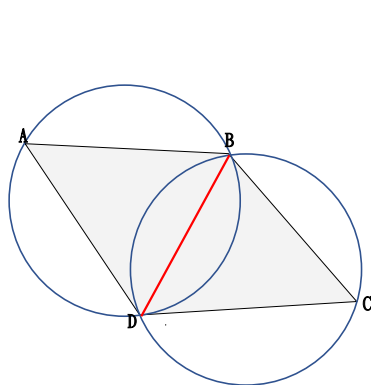


圖 15、具良好性質的三角形  
與其外接圓

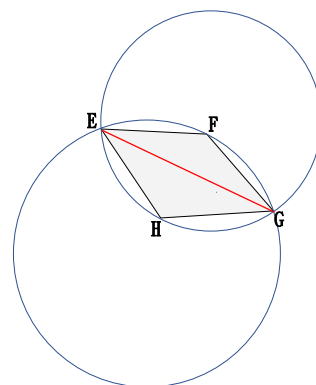


圖 16、不具良好性質的三角形  
與其外接圓

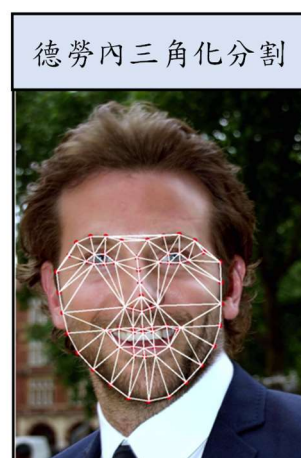
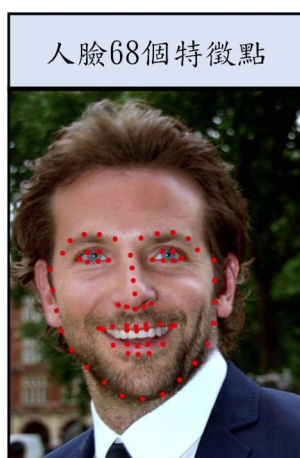


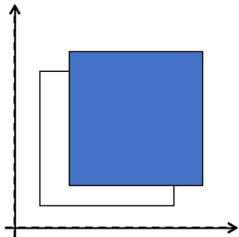
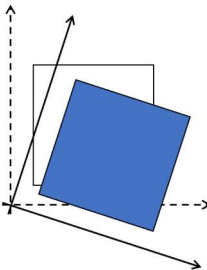
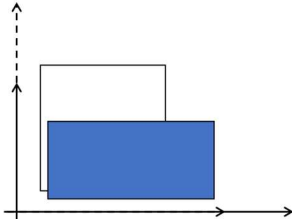
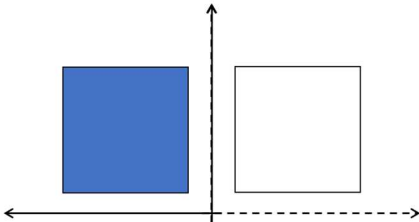
圖 17、德勞內三角化示意圖（Canu, 2019）

## 2. 對三角形做變形

本研究使用仿射變換（Affine Transformation）（Eating Lee, 2019）對三角形做變形，使來源圖片所分割的三角形吻合貼合在戴口罩圖片分割的三角形。其中仿射變換是指在向量空間中進行一次線性變換和一次平移，變換到另一個向量空間的過程。其數學表達式及齊次座標（用  $N+1$  維來代表  $N$  維座標）矩陣形式如下所示：

$$\begin{cases} u = a_1x + b_1y + c_1 \\ v = a_2x + b_2y + c_2 \end{cases} \quad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

數學表達式
齊次座標

	示意圖	矩陣
1. 平移		$\begin{bmatrix} 1 & 0 & m \\ 0 & 1 & n \\ 0 & 0 & 1 \end{bmatrix}$
2. 旋轉		$\begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$
3. 縮放		$\begin{bmatrix} w & 0 & 0 \\ 0 & h & 0 \\ 0 & 0 & 1 \end{bmatrix}$
4. 翻轉		$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$



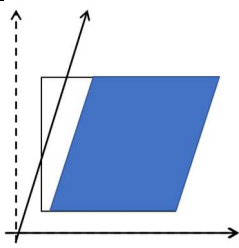
5. 錯切		$\begin{bmatrix} 1 & dy & 0 \\ dx & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
-------	---	---

圖 18、五種仿射變換型別示意圖與之矩陣

由於仿射變換比較複雜，很難找到仿射變化的 M 矩陣，本研究藉助 opencv 的 GetAffineTransform() 函數得到仿射矩陣 M。

`M = cv2.GetAffineTransform(dst, src)`

dst：原始圖像中的三角形三個點的點座標

src：變換後對應的三個點的點座標

M：求出的仿射變換矩陣

## （六）圖像填補

圖像填補（Image Inpainting）重建圖像中遺失或損壞部分的過程，是介於圖像編輯和圖像生成之間的一種技術，通過填充圖像指定區域的像素，使其與周圍的像素融為一體，達到和諧的成效。本研究用以填補口罩小面積殘留處，使結果可以消除口罩殘影，達到完全去除口罩的結果。

### 1. 模型介紹

圖像填補的架構引用自 Wadhwa, Dhall, Murala, & Tariq (2021) 研究中的模型，透過自行建構訓練集，經過重新訓練模型後測試。其主要由類似 CNN 架構結合 PatchGAN 中的判別器做為架構，生成圖像的方式是先運用一個神經網路形成模糊圖像，再利用另一個神經網路形成精確的邊緣，作為生成器輸出，再經過判別器、生成器的互相對抗調整參數以訓練模型。以下對神經網路架構中使用到的超圖卷積以及 PatchGAN 作介紹：

### (1) 超圖卷積

相比一般的簡單圖，每條邊與兩頂點連接，超圖一條邊則可以連接任一數量的頂點，超圖卷積 (Hypergraph Convolution) (popozyl, 2021) 可以提取特徵數據間的高階關係、擴展到多模態數據。如圖 19，超圖可表示為  $H = (X, E)$ ，其中  $X$  是點的集合， $X = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ ， $E$  是邊的集合， $E = \{e_1, e_2, e_3, e_4\} = \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$ 。

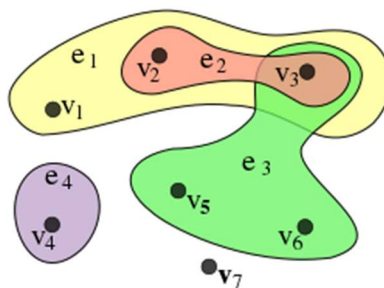


圖 19、超圖示例 (popozyl, 2021)

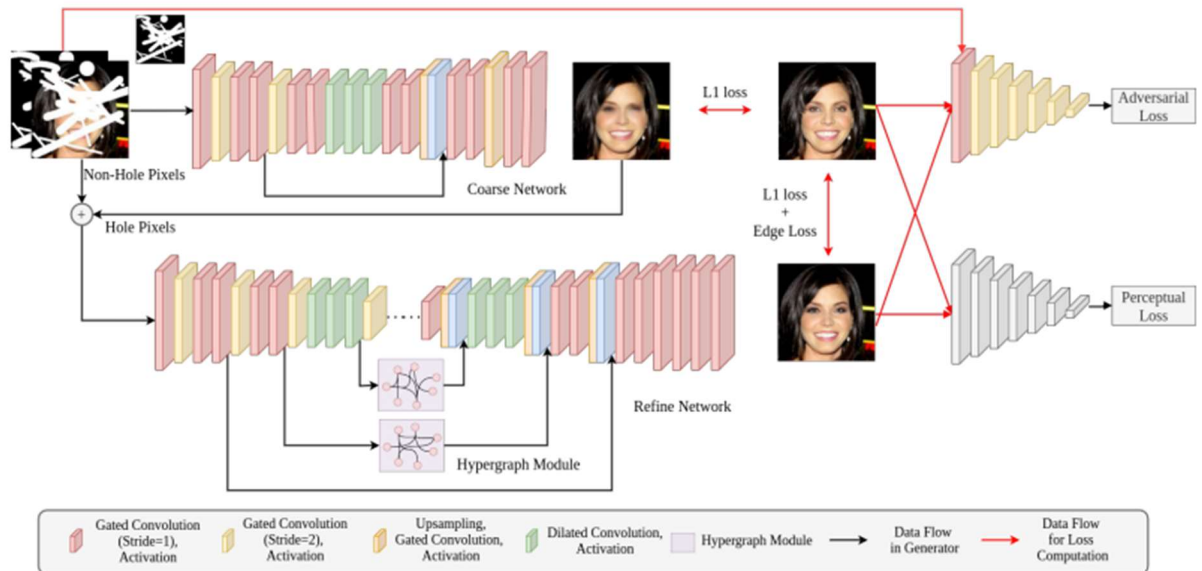
### (2) PatchGAN

一般 GAN 的判別器輸出是一個介於 0 ~ 1 之間的數值，代表真實圖片的概率，而 PatchGAN (xiaoxifei, 2019) 是輸出一個  $N \times N$  的矩陣  $X$ ，矩陣中的每一個數值皆表示一個原與圖片中的一小區塊（也就是 patch 的含意）的輸出，最終對各區塊數值求平均來代表真實圖片的機率，作為判別器最終輸出。而相較於一般的 GAN，其更能關注到圖像的細節。

## 2. 模型訓練

### (1) 選擇訓練集

考慮圖像填補區域為口罩，主要為填補人臉而非背景，因此本研究選擇 FFHQ 資料集，FFHQ 是一個高質量的人臉數據集，有 70,000 張人臉圖像，在年齡、種族和圖像背景上豐富多樣且差異明顯，在人臉屬性上也擁有非常多的變化，擁有不同的年齡、性別、種族、膚色、表情、臉型、髮型、人臉姿態等。



**Figure 3:** Overview of our proposed network for Image inpainting. The Coarse network roughly completes the missing holes. Later, the hypergraph convolution based Refine network generates the final high quality completed image.

圖 20、圖像填補之模型架構圖 (Wadhwa et al., 2021, p. 4)

### 3. 遮罩選取

為了填補口罩超出人臉輪廓的區域，本研究利用口罩區域的遮罩 A，以及人臉特徵點所圍出的遮罩 B，取遮罩 A 在遮罩 B 中的相對差集  $A - B$ ，即可以自動找出口罩殘影遮罩 C，進而輸入圖像填補模型進行修復。

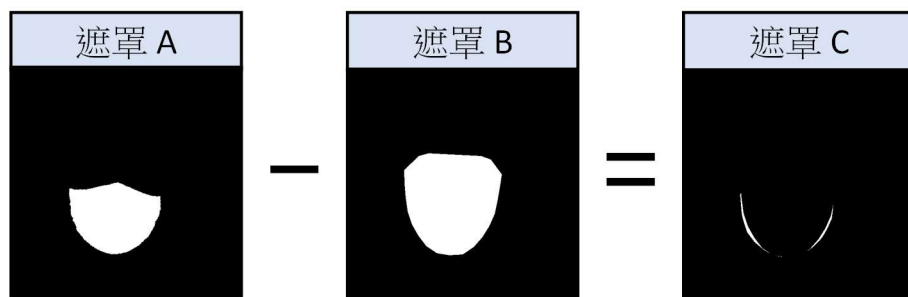


圖 21、圖像填補之填補遮罩選取示意圖

### (七) 人臉置換與圖像填補之結合

圖片經過人臉置換後，雖然特徵點無太大的偏移，但會因為口罩面積大於臉部輪廓，導致置換後仍會產生口罩的殘影。

在圖像填補的部分，若是把整個口罩部分作為遮罩，因為遮罩面積大，圖片經過填補後，即使訓練了大量圖像，在沒有來源圖片的參考之下，仍會產生極不自然的人像，也因為沒有特徵點的參考，圖片中的五官會造成極大的偏差，並且缺漏許多重要的身分訊息。

於是本研究結合了兩種模型，先透過人臉置換把來源圖片的人臉置換到戴口罩圖片上，再藉由框出口罩殘影的區域去作圖像填補，如圖 22，結合圖像填補後可讓人臉更自然。



圖 22、人臉置換、圖像填補與結合之比較

#### (八) 人臉色彩優化

由於拍攝來源圖片及戴口罩圖片時，拍攝場景、光線，以及人的姿勢、角度的不同，造成圖片中人臉的膚色、亮度差異，導致在進行人臉置換時會有清楚的明暗差異，使得貼合輪廓顯著。以下是優化方法：

##### 1. Reinhard 經典色彩遷移算法

根據 Reinhard 經典色彩遷移算法 (Reinhard, Adhikhmin, Gooch & Shirley, 2001) 調整來源圖片的亮度，使得來源圖片與戴口罩圖片中不包含口罩區域的亮度趨近一致。此算法是根據  $l, \alpha, \beta$  顏色空間中各通道不相連的特點（其中  $l$  為亮度， $\alpha$  為從綠色到紅色的分量， $\beta$  為從藍色到黃色的分量），將圖片各顏色分量的統計分析，算出一個線性變換，使得戴口罩圖片和來源圖片在  $l\alpha\beta$  顏色空間中，各顏色分量具有相同的平均值和標準差。其演算法流程為下列分項：

- (1) 將來源圖片及戴口罩圖片從原本的  $RGB$  色彩空間轉換成  $l\alpha\beta$  色彩空間，求得  $l\alpha\beta$  空間的各通道數據。
- (2) 分別計算來源圖片及目標圖片的平均值與標準差。
- (3) 帶入線性方程式：  
目標圖片  $l\alpha\beta$  空間的像素 =  
(目標圖片標準差 / 來源圖片標準差)  $\times$  (來源圖片像素 - 來源圖片平均值) + 目標圖片平均值
- (4) 將來源圖片轉回到  $RGB$  空間。

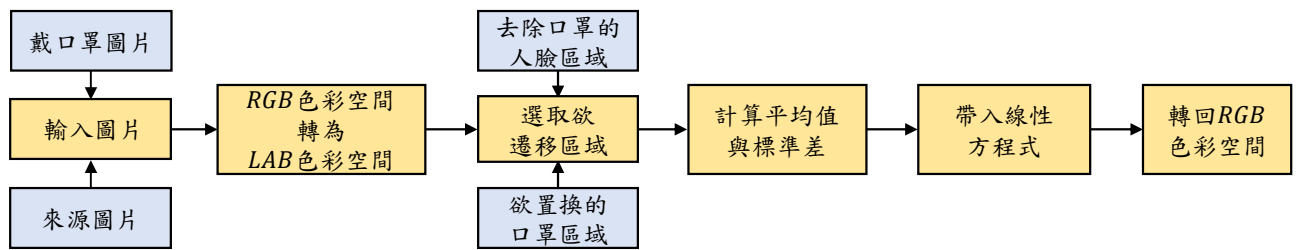


圖 23、Reinhard 經典色彩遷移算法流程圖

## 2. 泊松圖像編輯

在對圖像進行合成的過程中，如果源圖像  $S$  以及目標圖像  $\Omega$  有不同的紋理特徵，在合成圖片中就會產生明顯的邊界  $\partial\Omega$ ，為了讓拼接處變得更自然，本研究採用了泊松圖像編輯（Poisson Image Editing）(Pérez, Gangnet & Blake, 2003)，其以源圖像的梯度場做為指導，將融合邊界上的目標圖像和源圖像的差異平滑地擴散到融合圖像中，使融合後的圖像邊界顏色、亮暗差異不明顯，以達到無縫融合。

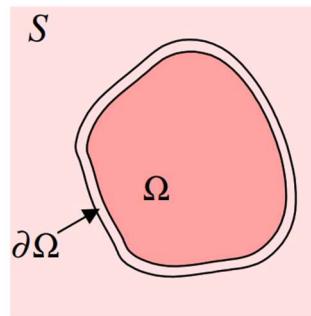


圖 24、泊松圖像編輯過程 (Pérez, Gangnet & Blake, 2003)



圖 25、色彩優化前後比較圖

## (九) 成效評估

### 1. 峰值信噪比

$$PSNR = 10 \log_{10} \left( \frac{R^2}{MSE} \right), \quad MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \quad (\text{Jason Chen, 2019})$$

峰值信噪比（Peak Signal To Noise Ratio，PSNR）比較每像素的顏色相似度，值越大越相似，其數值最高為 50 dB。

$R$  為圖中像素  $RGB$  的最大值（ $RGB$  最大值為 255）。 $MSE$  為均方誤差距離（Mean Square Error），其中  $m, n$  為像素的行列值； $M, N$  為圖像長與寬大小； $I_1$  與  $I_2$  為比較兩張圖。單位為分貝（dB），數值的級距意義如下：

- (1)  $\geq 40$  dB：非常接近原始圖像。
- (2) 30 ~ 40 dB：人眼很難察覺壓縮後和原始影像的差異。
- (3) ~ 30 dB：人眼就可以察覺出圖像的差異。
- (4) 10 ~ 20 dB：人眼可看出圖像的原始結構，且判斷兩張圖像不存在很大的差異。
- (5)  $< 10$  dB：人類很難用肉眼去判斷兩個圖像是否為相同。

## 2. 結構相似度指標

結構相似度指標（Structural Similarity Index, SSIM index）是用以衡量兩張數位影像相似程度的指標。當兩張影像其中一張為無失真影像，另一張為失真後的影像，兩者的結構相似度可以看成是失真影像的影像品質衡量指標。相較於傳統所使用的影像品質衡量指標，像是峰值信噪比，結構相似度在影像品質的衡量上更能增強對影像品質的判斷。其數值範圍落在  $[0, 1]$  之間，數值越高，影像失真程度越小。

分別給予兩張圖片  $x$ 、 $y$ ， $x$  為目標圖片， $y$  為來源圖片：

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (\text{Xavier Jiezou, 2021})$$

其中  $l(x, y)$  比較  $x$  和  $y$  的亮度； $c(x, y)$  比較  $x$  和  $y$  的對比度； $s(x, y)$  比較  $x$  和  $y$  的結構，結構相似度指標的值越大，代表兩個信號的相似度越高； $\alpha$ 、 $\beta$ 、 $\gamma$  為調整  $l(x, y)$ 、 $c(x, y)$ 、 $s(x, y)$  相對重要性的參數； $\mu_x$ 、 $\mu_y$  為  $x$  和  $y$  的像素平均值， $\sigma_x$ 、 $\sigma_y$  為  $x$  和  $y$  的標準差， $\sigma_{xy}$  為  $x$  和  $y$  的共變數， $C_1$ 、 $C_2$ 、 $C_3$  為維持  $l(x, y)$ 、 $c(x, y)$ 、 $s(x, y)$  的常數。

### 3. 人為評估

由於峰值信噪比與結構相似度指標未必能呈現完整的視覺感知，因此本研究設計了一套評估標準，找來 100 位受試者進行人為評估，並給予受試者目標圖片作為評估參考，對輸出圖片進行給分，評估標準如下：

表 1、人為評估標準

評估	口罩面積殘留	自然度	色彩落差程度
0	口罩有殘留，面積大	有明顯瑕疵，且不自然面積大	色彩有落差，程度大
1	口罩有殘留，面積小且明顯	有明顯瑕疵，但不自然面積尚小	色彩有落差，程度中
2	口罩有殘留，面積小且不明顯	有少許瑕疵，但不影響整體感觀	色彩有落差，程度小
3	幾乎看不到口罩殘留	幾乎沒有瑕疵，沒有合成的感覺	幾乎沒有色彩落差

#### (十) 實驗設計說明

在前述的研究方法中，本研究針對相關的影響因素，設計實驗進行探討，實驗設計規劃如下：

1. 實驗一：Region Growing 演算法閾值參數的選擇，對口罩分割成效。
2. 實驗二：Grabcut 與 Region Growing 演算法之個別與整合改良，對口罩分割成效。
3. 實驗三：人臉置換結合圖像填補技術，對口罩邊緣殘影的消除之成效。
4. 實驗四：Reinhard 經典色彩遷移演算法和泊松圖像編輯的使用，對人臉色彩優化之成效。
5. 實驗五：MaskTheFace 模型中不同類型口罩的使用，對人臉復原的成效。
6. 實驗六：全臉區域與口罩區域的圖片置換，對人臉復原的成效。



## 肆、研究結果

### 一、研究結果

（一）實驗一：Region Growing 演算法閾值參數的選擇，對口罩分割的成效。

由於 Region Growing 演算法中，不同閾值所分割出的口罩成效不一，經由反覆測試，選出適當閾值進行分割。

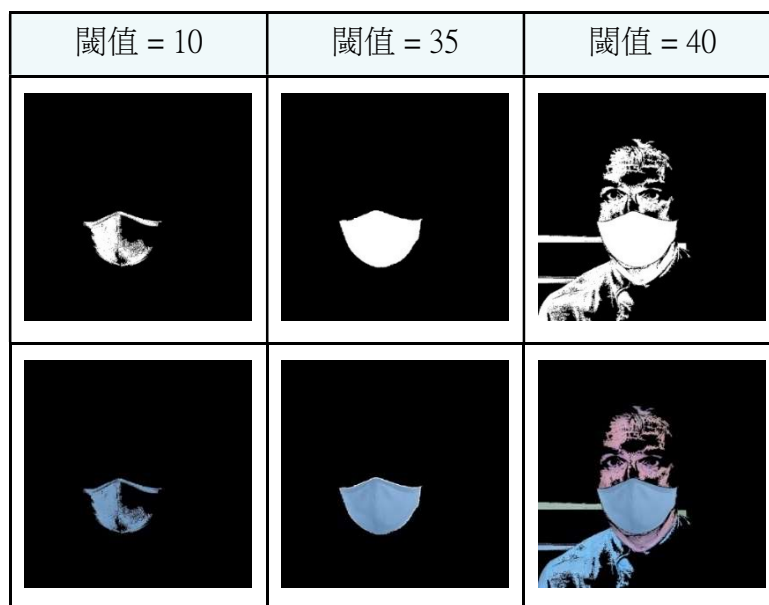


圖 26、Region Growing 不同閾值比較示意圖

（二）實驗二：Grabcut 與 Region Growing 演算法之個別與整合改良，對口罩分割成效。

在本研究中，Grabcut 和 Region Growing 演算法分別會有過度延伸、內部雜訊等問題出現，因此本研究結合兩者優點，來生成精準的口罩遮罩。

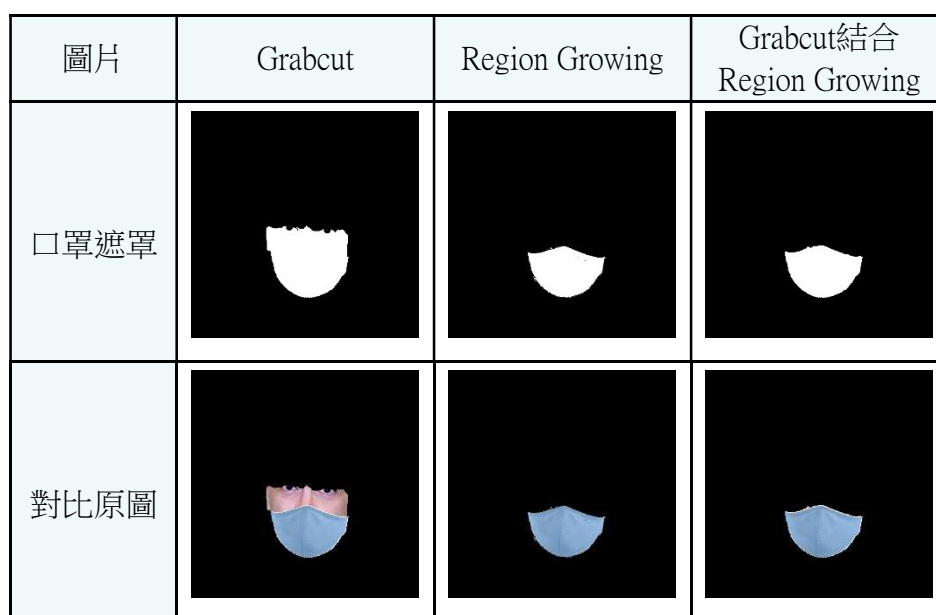


圖 27、口罩偵測輸出結果

(三) 實驗三：人臉置換結合圖像填補技術，對口罩邊緣殘影的消除之成效。

本研究利用將人臉置換結合圖像填補，以修補口罩殘影處，以下為結合圖像填補前後的比較：

圖片	圖片一	圖片二	圖片三
目標 圖片			
來源 圖片			
戴口罩 圖片			
純人臉置換 輸出圖片 (A)			
填補遮罩			
人臉置換 結合 圖像填補 輸出圖片 (B)			

圖 28、人臉置換結合圖像填補前後成效

表 2、純人臉置換結評估分數




評估	圖片一	圖片二	圖片三
PSNR	<b>31.5271</b>	29.5454	23.8555
SSIM	<b>0.9694</b>	0.9438	0.8972
人為評估 - 口罩面積殘留	1.1683	1.1089	<b>1.4356</b>
人為評估 - 自然度	1.0198	0.9208	<b>1.2772</b>

表 3、人臉置換結合圖像填補結果評估分數

評估	圖片一	圖片二	圖片三
PSNR	28.6468	<b>28.8020</b>	23.5507
SSIM	<b>0.9148</b>	0.9079	0.8714
人為評估 - 口罩面積殘留	<b>2.0792</b>	1.9604	1.8614
人為評估 - 自然度	<b>1.7129</b>	1.6535	1.5149

(四) 實驗四：Reinhard 經典色彩遷移演算法和泊松圖像編輯的使用，對人臉色彩優化之成效。

由於戴口罩圖片和來源圖片為不同時間與地點拍攝的圖片，除了背景不同，光線與亮度也有很大的差異，以下是人臉經過色彩優化的前後比較：

圖片	圖片一	圖片二	圖片三
目標 圖片			

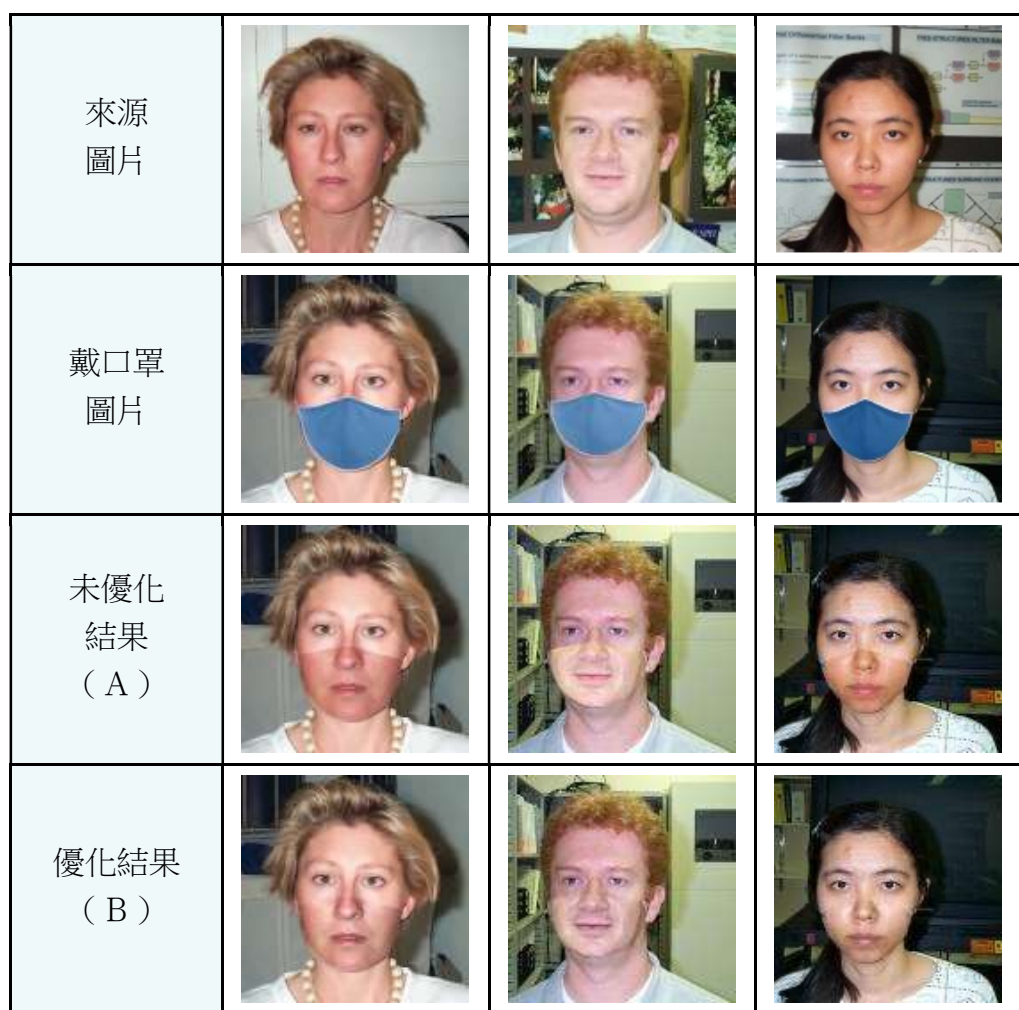


圖 29、色彩優化前後的成效

表 4、未色彩優化結果評估分數

評估	圖片一	圖片二	圖片三
PSNR	24.5088	28.6395	28.8134
SSIM	0.9134	0.9151	0.9250
人為評估 - 色彩落差程度	0.6040	0.7624	0.4950
人為評估 - 自然度	0.6040	0.6832	0.5149

表 5、色彩優化結果評估分數

評估	圖片一	圖片二	圖片三
PSNR	27.2279	29.2814	30.4369
SSIM	0.9568	0.9159	0.9287
人為評估 - 色彩落差程度	1.6535	1.4554	2.3069
人為評估 - 自然度	1.5050	1.3861	2.2277

(五) 實驗五：MaskTheFace 模型中不同類型口罩的使用，對人臉復原的成效

不同類型的口罩，大小、形狀、覆蓋面積都不同，影響了模型合成的成效：

















圖片	N95	KN95	Surgical	Cloth
目標 圖片				
來源 圖片				
戴口罩 圖片				
輸出 圖片				

圖 30、不同類型口罩的成效



表 6、圖 30 輸出圖片的評估分數

評估	N95	KN95	Surgical	Cloth
PSNR	31.9978	30.7861	31.6063	31.1406
SSIM	0.9375	0.9321	0.9389	0.9314

(六) 實驗六：全臉區域與口罩區域的圖片置換，對人臉復原的成效



圖 31、臉部不同區域置換效果

## 伍、討論

(一) 探討 Region Growing 演算法閾值參數的選擇，對口罩分割的成效：

從結果可知，若閾值過小，由於口罩在圖像中因光線、角度不同有顏色差異，輸出結果無法完整切割口罩區域；若閾值過大會使分割的成效不彰，像素差異小於閾值的衣服、人臉部分無法與口罩區域進行有效的分割，故本研究選擇了適當大小的閾值，也就是閾值等於 35，以達到較好的分割成效。

(二) 探討 Grabcut 與 Region Growing 演算法之個別與整合改良，對口罩分割成效：

從研究結果可知，使用 Grabcut 演算法、Region Growing 演算法偵測口罩區域各有優缺點。Grabcut 演算法分別在左、右、下側有較好的分割成效，但因人臉特徵點擷取的限制，矩形上側無法貼齊口罩頂點位置，造成結果延伸至人臉眼睛部位，

無法精準地分割口罩與人臉。Region Growing 演算法的分割結果則會產生內部的小點以及外部毛邊的突出。經過二者結合後，可以將分割的區域侷限在口罩範圍。

### （三）比較人臉置換結合圖像填補技術，對口罩邊緣殘影的消除之成效：

從 PSNR 和 SSIM 來看，人臉置換結合圖像填補後的數值微幅下降但並不顯著，這是由於峰值信噪比為比對圖片間的像素，對誤差敏感度高，而結構相似度指標的數值在高分範圍時，影像間的差異也很難被人眼察覺，因此分數無法完整地呈現人的視覺感受，會造成肉眼認為是自然的輸出圖片，卻得到較低的分數。

從視覺上來看，輸出圖片經過圖像填補後，成功讓人臉置換後產生的口罩拼接痕淡化，使圖片中的人像也更為自然。人為評估上可以得知口罩殘餘減少；人臉的自然度評估分數雖然大部分落在 1 到 2 分之間，圖片有部分瑕疵，但經過圖像填補後仍有略微提升。由此可見人臉置換結合圖像填補後可以有效的提升圖片的品質。

### （四）探討人臉色彩優化之成效：

由於來源圖片與戴口罩圖片來自不同張照片，受到亮度影響導致臉部膚色差異，若沒有經過色彩優化，會使得輸出圖片膚色不均，在拼接處產生明顯的邊界，雖然色彩優化後的光線與原圖片有所不同，但使得輸出圖片整體具有近似的膚色，從成效評估中也能看到色彩優化後的分數明顯提高，人為評估中的兩個項目皆從 0~1 分提升至 1 分以上，說明色彩優化有良好的改善成效。

### （五）探討不同類型的使用，對人臉復原的成效：

從研究結果發現，口罩面積大會因為超出臉部輪廓的面積較多，在圖像處理過程造成視覺上的瑕疵，成效評估有較低的分數。像是 KN95 口罩面積超出臉部輪廓甚多，人臉置換和圖像填補過後超出臉輪廓處仍會略為不自然，從峰值信噪比和結構相似度指標中可以看出，KN95 分數最低。

### （六）全臉區域與口罩區域置換結果分析

口罩區域置換的在拼接處仍有些許不自然，而全臉區域的置換雖然把眼睛部分一併置換，但只要兩圖片眼睛角度、神韻相似，便不會受到影響，且因為拼接在臉部較靠近邊緣處，相較於口罩區域的置換，其輸出圖片整體視覺上更為自然。



## 陸、結論

- 一、 在擷取人臉 68 個特徵點時，會因為口罩的遮擋造成人臉特徵點偵測的偏差，而本研究針對特徵點進行調整，提升了人臉有覆蓋物時的特徵擷取，並應用在戴口罩圖片與來源圖片的特徵位置比對。
- 二、 現有的研究著重在偵測是否有配戴口罩，少有對於口罩區域偵測的研究，本研究利用 Grabcut 結合 Region Growing 演算法，成功自動偵測口罩區域並擷取口罩輪廓特徵點。
- 三、 純人臉置換後，在口罩面積超過臉部輪廓處容易會有口罩殘留；而圖像填補技術對於大面積的填補成效差，若把口罩區域遮罩當作填補遮罩，會有極度不自然的人臉輸出。在本研究成功地整合傳統方法的人臉置換與深度學習的圖像填補，在人臉置換後，利用圖像填補技術填補口罩殘影，使得生成後的圖像有自然的人臉。
- 四、 本研究進行色彩優化，利用 Reinhard 經典色彩遷移算法及泊松圖像編輯，輸出圖片的人臉膚色具一致性，拼接處顏色銜接也更自然，有助於提升整體人臉自然度。
- 五、 未來展望：
  - （一）在色彩優化的部分，經過優化後還是會有部分不自然處，希望未來在色彩優化方面有更細節的研究，使兩圖片置換的拼接處更為自然。
  - （二）希望將研究整合成一個系統，未來可製作成簡易的 App，將戴口罩圖片與參考圖片輸入，便可得到輸出結果。
- 六、 未來應用：
  - （一）將本研究應用於犯案人員的進行臉部復原與追蹤，將戴口罩的犯案人員圖片輸入人臉口罩去除系統，進行口罩修復及人臉特徵比對，可協助檢調機關破案。
  - （二）將本研究運用於其他領域，例如：醫學美容，透過置換美化的人臉區域，可以藉此輸出容貌改變後的成效，以供人參考。

## 柒、參考文獻資料

許恆修（民 108 年 6 月 17 日）。何謂 Generative adversarial Networks (GANS) 生成式對抗網路。取自：<https://r23456999.medium.com/%E4%BD%95%E8%AC%82-generative-adversarial-networks-gans-%E7%94%9F%E6%88%90%E5%BC%8F%E5%B0%8D%E6%8A%97%E7%B6%B2%E8%B7%AF-4500a06dd0be>。

Tommy Huang（民 107 年 5 月 27 日）。卷積神經網路(Convolutional neural network, CNN) — CNN 運算流程。取自：<https://chih-sheng-huang821.medium.com/%E5%8D%B7%E7%A9%8D%E7%A>

5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF-convolutional-neural-network-cnn-cnn%E9%81%8B%E7%AE%97%E6%B5%81%E7%A8%8B-ecaec240a631 。

Hui, Z., Li, J., Wang, X., & Gao, X. (2020). Image fine-grained inpainting.arXiv preprint arXiv:2002.02609, 2020.

Computational Vision at Caltech (n.d.). Caltech Computational Vision Group Archive – Faces. Retrieved from <http://www.vision.caltech.edu/html-files/archive.html>.

Anwar, A., & Raychowdhury, A. (2020). Masked face recognition for secure authentication.arXiv preprint arXiv:2008.11104.

Adrian Rosebrock (2017, April 3). Facial landmarks with dlib, OpenCV, and Python. Retrieved from <http://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>.

Sergio Canu (2019, May 8).Face swapping (explained in 8 steps )-Opencv with Python. Retrieved from <https://pysource.com/2019/05/28/face-swapping-explained-in-8-steps-opencv-with-python/>.

CH.Tseng (2019, August 7). Delaunay Triangulation & Voronoi Diagrams. Retrieved from <https://chtseng.wordpress.com/2019/08/07/delaunay-triangulation-voronoi-diagrams/>.

Eating Lee ( 民 108 年 3 月 19 日 ) 。 圖像幾何變換之仿射變換原理及實現 。 取自 : [https://blog.csdn.net/qq\\_40369926/article/details/88672855](https://blog.csdn.net/qq_40369926/article/details/88672855) 。

Wadhwa, G., Dhall, A., Murala, S., & Tariq, U. (2021). Hyperrealistic image inpainting with hypergraphs. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3912-3921).

popozyl ( 民 110 年 2 月 9 日 ) 。 超圖神經網路 ( Hypergraph Neural Networks, HGNN ) 。 取自 : <https://www.cnblogs.com/popodynasty/p/14391442.html> 。

xiaoxifei ( 民 108 年 1 月 16 日 ) 。 關於PatchGAN的理解 。 取自 : <https://blog.csdn.net/xiaoxifei/article/details/86506955> 。

Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. IEEE Computer graphics and applications, 21(5), 34-41.

Pérez, P., Gangnet, M., & Blake, A. (2003). Poisson image editing. In ACM SIGGRAPH 2003 Papers (pp. 313-318).

Jason Chen ( 民 108 年 1 月 23 日 ) 。 【影像處理】峰值信噪比 Peak Single-to-Noise Ratio 。 取自 : <https://jason-chen-1992.weebly.com/home/-peak-single-to-noise-ratio> 。

Xavier Jiezou ( 民 110 年 4 月 6 日 ) 。 【python】ssim原理簡介及代碼實現 。 取自 : [https://blog.csdn.net/qq\\_42951560/article/details/115463083](https://blog.csdn.net/qq_42951560/article/details/115463083) 。