2021 年臺灣國際科學展覽會
優勝作品專輯

作品編號　**190014**

參展科別　電腦科學與資訊工程

作品名稱　基於深度學習之服裝試衣系統

得獎獎項　大會獎 四等獎

就讀學校　國立新竹女子高級中學

指導教師　古佳怡、胡敏君

作者姓名　楊子誼、林維余、徐熙筠

關鍵詞　衣服試穿、深度學習、幾何匹配模型

作者簡介

我是楊子誼，就讀新竹女中二年級。很高興能跟兩位隊友們一起完成這個研究，也很榮幸能踏上國際科展的舞台。謝謝帶領我們的指導老師、教授、及學長，未來我會繼續努力！

我是林維余，就讀新竹女中二年級，喜歡故事、超脫於現實的想像，希望可以一直透過科學來達到生活中無法碰觸的幻想，很高興這次有這樣的機會參加國際科展。

我是徐熙筠，今年就讀新竹女中二年級。在這次的研究中，我才初步了解到深度學習如何實際運用，也體會到其中的難處和限制(還有 loss 一直高居不下的困難！！)。謝謝帶領我們從理解到建構的指導老師、教授、及學長，也謝謝父母的支持和組員的相互合作！

摘要

本研究以 AI 虛擬試衣系統(Virtual Try-on)為主題，透過深度學習技術，並結合幾何匹配模型，開發出試衣系統，可將使用者上傳的照片 ，模擬成穿著新衣的模樣。

首先，以深度學習模型將人物原始圖片取出骨架節點，並生成人體遮罩以及保留人物頭部，再結合以上三種資訊合成為高維特徵圖。接著將目標替換衣物生成出依照人體姿態扭曲後的衣物圖片。最後於 Virtual Try-on 模型中將人體高維特徵圖與扭曲衣物作為輸入，並經過深度學習網路合成出穿著目標衣物之人體圖像。本研究結果發現，人物站姿單純，且雙手緊貼身側，以及拍攝角度為正面、衣服款式為短袖、背景色彩對比度較高與衣服圖案單純的原始圖片 ，可得到較好的合成結果。

Abstract

This research mainly introduces AI virtual Try-on model, which is built via deep learning technique and geometric matching module. The Try-on system allows users to simulate the outlook of themselves dressed in the target clothes with the image uploaded by them.

First, we extract the keypoints of the original human images and generate the body mask as well as the image reservation. Then, we output the three generated images to fuse them into human representation image. Later, we warp the target clothes according to the human pose. Lastly, we input the human representation image along with the warped-cloth images into the Try-on module, resulting in the merged image between the human images and the target clothes. From the results, we find out that short sleeves, high color contrast ratio backgrounds, simple patterns, and simple poses such as facing forward with both hands laying flat, make better outputs which acquire a higher similarity.

壹、研究動機

現今網路發展日新月異，其中與生活最有相關性的便是網路商城。網路商城的便利之處在於：不須出門也能購買生活所需物品，其中又以購買衣服最為常見。但由於在網路上購衣的緣故，使用者無法試穿衣物，時常發生實際效果與原先想像不符，而需要退換貨的情形。在此狀況下不僅買方無法達到心裡的期望，賣家更可能在退換貨的過程中耗損不必要的資源。因此，我們想要運用所學的知識，透過深度學習模型實作出一套模擬試衣系統。

貳、研究目的

本研究的目的為建構虛擬試衣系統，可透過上傳的照片，模擬出使用者實際穿上新衣的模樣，協助使用者更為了解是否符合心理的預期。此外我們希望藉由輸入不同圖片進行比對，探討試衣模型之優缺點，以及當前使用試衣模型最合適的服裝形式和人物的姿勢。

參、研究設備及器材

一、硬體

(一) DELL OptiPlex 3060

CPU：Intel(R) Core(TM) i3-8100 CPU @ 3.60GHz 3.60GHz

GPU：NVIDIA GeForce GTX 1060 6GB

記憶體：8.0 GB

作業系統:Windows 10

二、軟體環境

(一)Anaconda 4.8.3

(二) Visual Studio Code 2015

(三) NumPy、PyTorch、OpenCV

三、網頁工具

(一) Contrast Ratio

肆、研究方法與過程

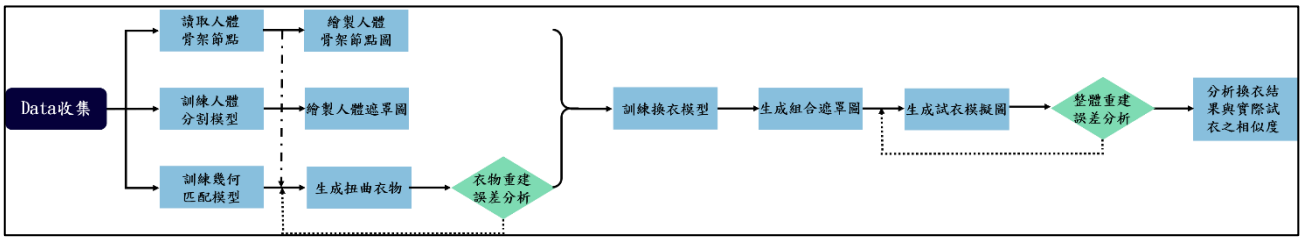


圖 4-1 研究流程圖(此圖為作者自製)

一、 文獻探討

以下將探討現今虛擬試衣技術的發展情形，以及介紹本研究中深度學習模型所使用的類神經網路。

(一) 虛擬試衣模型

最初虛擬試衣的實作是使用 3D 人體建模，再以剪貼的方式將目標衣物換上人物。這種方法對於圖像的掌握度最高也最精確，但卻需大量運算和精密的儀器對人體進行三維的掃描，使模擬過程耗費不少時間和金錢，因此無法被普及化。隨著深度學習技術的發展，基於圖像生成演算法的虛擬試衣系統也陸續被提出，其將輸入的衣服圖像進行扭曲和拼接到目標人物的圖像上，以產生目標的試衣結果。利用此方法的試衣模型非常多，其中最為經典的為 VITON 模型，其中包含了薄板樣條插值(Thin plate spline，簡稱 TPS)和多任務學習的技術。

VITON 模型共可分為三步驟執行，分別為：數據準備工作、扭曲衣物之生成、及最終結果圖片的合成。

首先為數據準備工作的部分，此步驟之目標為得到原始圖片中人體的部分特徵。而程式透過前端原始圖片的輸入，取得其人體骨架節點圖、人體遮罩圖、及人體頭部保留，並將三者結合為人體高維特徵圖。再者為扭曲衣物之生成步驟，先是將前步驟得到的人體高維特徵圖及目標衣物圖片作為輸入，進入類 Unet 的 encoder-decoder 模型，以生成衣物遮罩圖。而後將其再次與人體高維特徵圖進一步進行姿態匹配，生成扭曲衣物。最終為結果圖片的合成：將前步驟所得扭曲衣物和人體高維特徵圖一齊輸入 Unet 網路中，透過 Alpha 合成的技術生成穿著目標衣物之人體圖像。

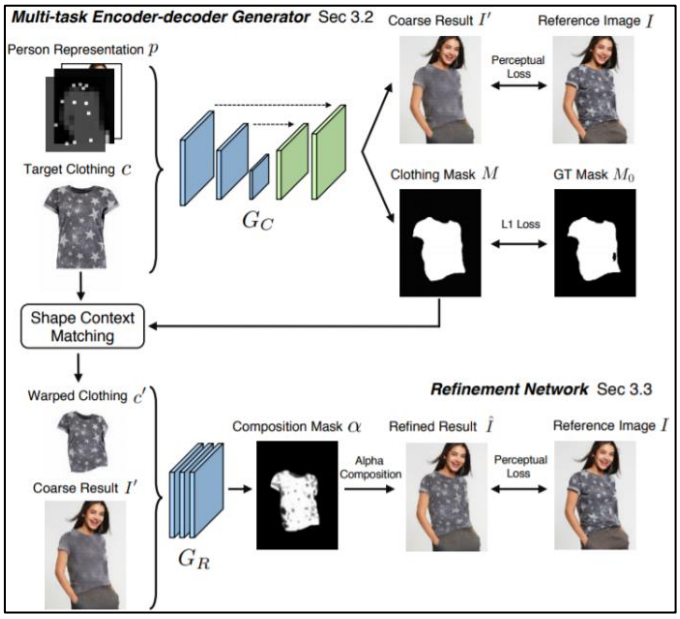


圖 4-2 VITON 程式流程圖 取自[2]

(二) 影像分割模型

虛擬試衣系統需參考圖像中人體部位的位置，而判斷位置資訊可視為影像分割的問題。圖 4-3 展示了人體部位的影像分割範例，以左圖為輸入，輸出跟原圖大小一樣標註每像素的身體部位分類資訊，以下我們對於影像分割模型進行文獻探討。

圖 4-3 影像分割範例 取自[14]

1. 全卷積網路(Fully Convolutional Networks, 簡稱 FCN))

全卷積網路為早期最經典的影像分割模型，相較於分類問題只有輸出圖片整體的類別資訊，影像分割需要輸出每像素的類別資訊，因此全卷積網路捨棄掉全連接層，將小尺寸的特徵圖進行上升採樣和卷積操作，以得到與原圖相同尺寸的分類結果。這個架構的優勢為輸出圖像的大小可以隨著輸入圖片改變，所以可對與訓練資料大小不同的圖片進行預測，在實際應用上有更大的彈性。

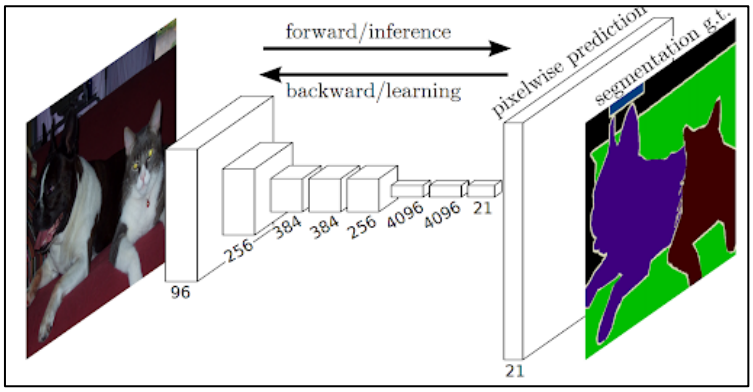


圖 4-4 全卷積網路模型架構 取自[16]

2. U-Net

在 FCN 模型中由於特徵經過多層的降採樣而失去了空間上的精度，因此在邊界上的分割結果較為不精準。U-Net 為了解決這個問題，其將編碼器不同尺度的特徵連接到解碼器上，如圖 4-5 所示。這樣的架構設計不僅可以得到不同尺度的特徵，也可以保留空間細節的精度，以在物體邊界上得到更為準確的分割結果。

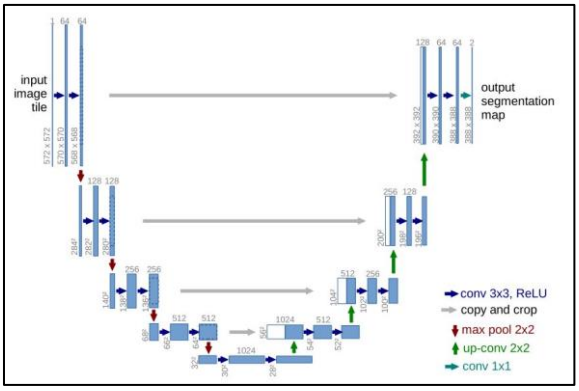


圖 4-5 Unet 原理圖 取自[14]

3. Deeplab v3+

在 Deeplab v3 中使用了不一樣的方法來保留空間的精度，其使用不同大小的擴張卷積(Atrous/Dilated Convolution)在同一層萃取不同尺度的資訊。其中擴張卷積的原理為在原先卷積之間填入一些零值，使視野變大而不降低解析度，如圖 4-6 所示。然而其對於結果的生成只使用單層的上升採樣一次放大回輸入圖像的尺寸，運算過程將耗費大量的運算資源。

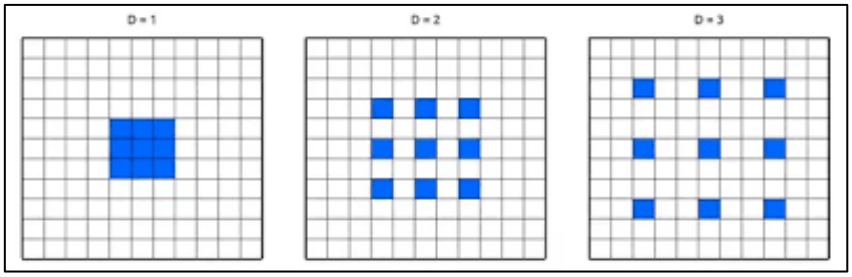


圖 4-6 Atrous Convolution 原理圖 取自[6]

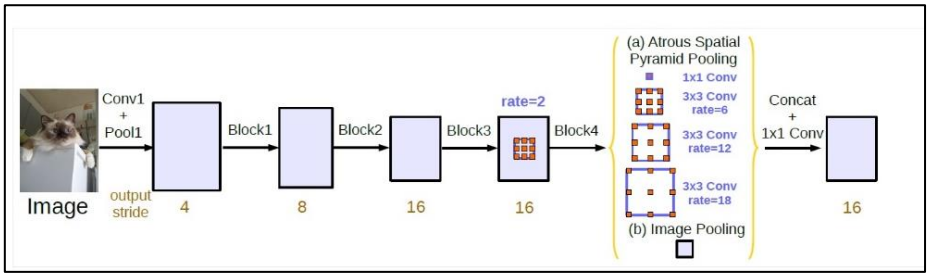


圖 4-7 Deeplab v3 架構圖 取自[6]

而 FCN 與 U-Net 中所使用的 encoder-decoder 架構雖然會失去空間的細節資訊，但是其會先進行降採樣再進行上升採樣而大大減少計算量。Deeplab v3+綜合上述兩項所提及的優點，將原本的 Deeplab v3 當作 encoder，結合 decoder 結構進行多層的升採樣得到最終結果，如圖 4-8 所示。

Deeplab v3+是由 spatial pyramid pooling 及 skip-connection encoder decoder 組合而成。前者是以不同大小的擴張卷積在同一層萃取不同尺度的資訊，比起 Unet 更能保留空間的精度。後者是將編碼器中較為精細的空間資訊傳遞到解碼器中，以得到更為精確的邊界分割結果。

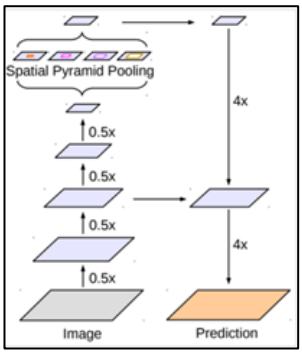


圖 4-8 Deeplab v3+改良原理圖 取自[4]

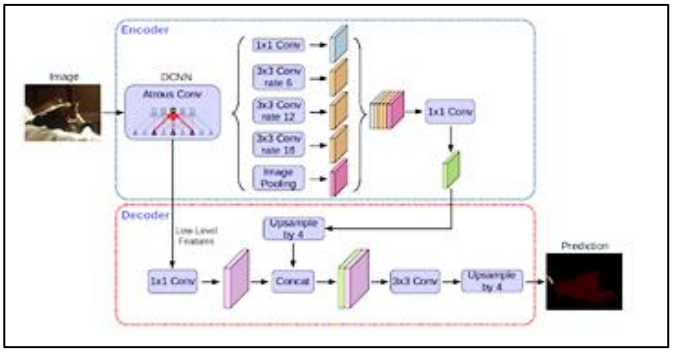


圖 4-9 人體遮罩圖生成原理圖 取自[4]

二、 研究方法

本研究主要參考的網路架構為 CP-VTON 模型(Characteristic- Preserving Virtual Try-on，基於圖像特徵保留的虛擬試衣網路)。此模型因處理架構的改變而使模型比起 VITON 模

型減少了複雜的計算量，提升了模型網路的效率，也提高了特徵保留的程度，故在處理特徵豐富的服裝或者有較大幅度的形變時，較不會產出模糊的結果。

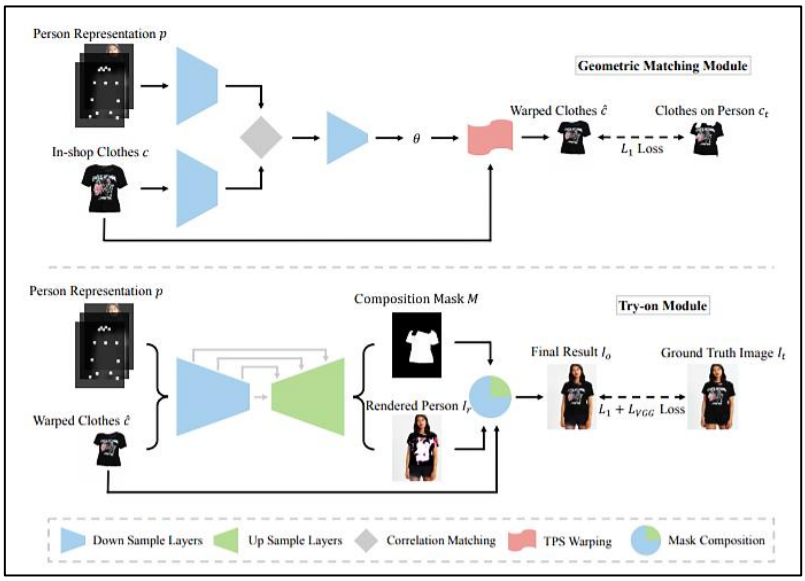


圖 4-10 CP-VTON 流程圖 取自[1]

本研究將會使用上述模型於公開資料集進行訓練，並將圖像前處理與人體特徵建構等步驟串連成完整的架構。以下的說明分為四部分：第一、二部分會分別說明所有輸入資料的格式、種類以及使用模型的方法原理，第三部分則會介紹其中最核心的模型概念，第四部分將以上三部分串接成完整程式且引入我們設計的使用者介面。

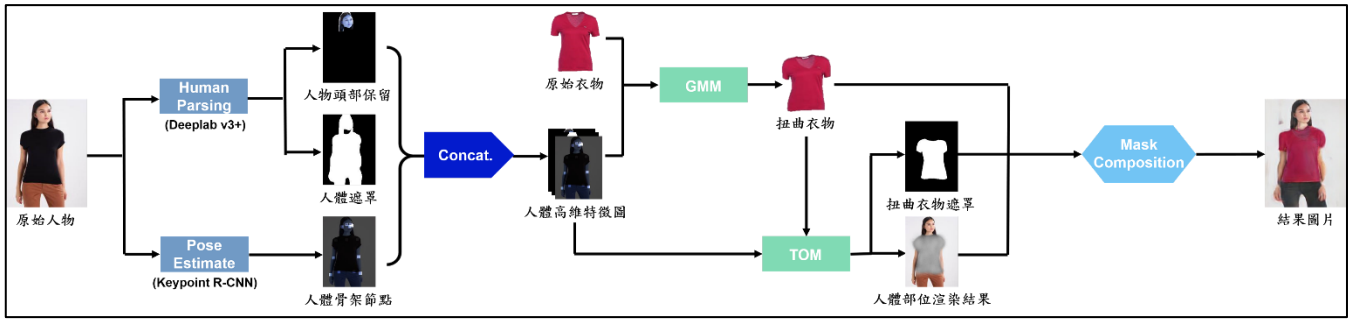


圖 4-11 程式流程圖(此圖為作者自製)

(一)建構人體高維特徵圖

人體高維特徵圖作為虛擬試衣系統的輸入包含了三種資訊，分別為人體骨架節點、人體遮罩及人物頭部。

1. 人體骨架節點：

我們使用 torchvision 函式庫中的 Keypoint R-CNN 作為骨架節點偵測模型，並套用函式庫所提供訓練完成的模型參數，其中 Keypoint R-CNN 可同時生成人體

位置與對應的骨架節點機率圖。在得到每個關節點的二維座標之後，我們依據下列步驟建構骨架節點的表示圖：

(1) 繪製底圖

使用 OpenCV 讀入原始圖片後取得圖片長寬，用以畫出 18 張大小相同的黑底 numpy 格式矩形，作為骨架節點的背景。

(2) 取出人體骨架節點

使用 Keypoint R-CNN 函數在一個人物上一共可抓到 17 個點的 x 座標與 y 座標，每個點的編號為由上到下排序。但虛擬試衣模型所使用的骨架節點編號與這裡的編排有所不同，所以我們在其中增加了一個 list 作為兩者的對照。另外缺少的 1 號節則利用肩膀處的兩點的連線中點作為替代。

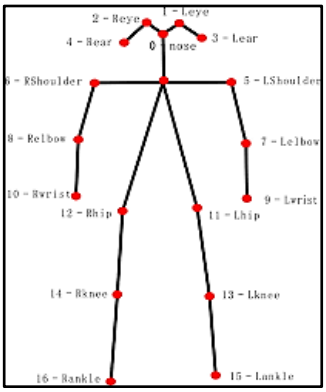


圖 4-12 17 個節點順序示意圖 取自[17]

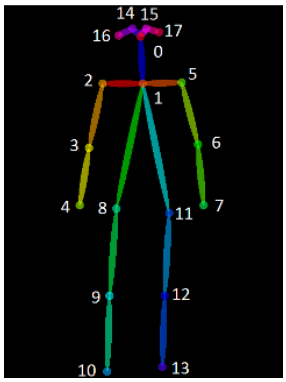


圖 4-13 18 個節點順序示意圖 取自[7]

(3) 骨架節點

在此步驟中，是透過先將骨架節點的資料轉為 numpy 格式，再使用 OpenCV 的繪圖功能，將每個骨架節點標示為 11pixel*11pixel 的白色矩形，分別繪製於黑色底圖上。

圖 4-14 測試原圖(取自網路)

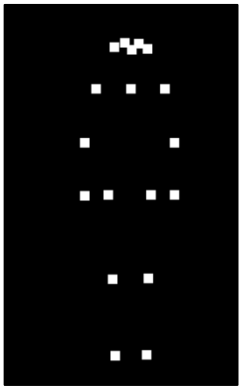


圖 4-15 骨架節點繪製圖(作者實作)

```python
img = cv2.imread("data/test_use.jpg")

print(img.shape)
cv2.imshow("Image", img)

list_new=img.shape

imgsingle=np.zeros([list_new[0],list_new[1],1])
print(imgsingle)
indexlist=[0,15,14,17,16,5,2,6,3,7,4,11,8,12,9,13,10]
img_list=[]
for i in range(18):
    img_new=np.zeros([list_new[0],list_new[1],1])
    img_list.append(img_new)
```

圖 4-16 讀取圖片與建立底圖程式碼(此圖為作者自製)

```python
for i in range(result["labels"].shape[0]):
    if result["scores"][i] > 0.9:
        kp = result["keypoints"]
        print(kp[i])
        neck = (int((kp[i,5,0] + kp[i,6,0])/2), int((kp[i,5,1] + kp[i,6,1])/2))
        hip = (int((kp[i,11,0] + kp[i,12,0])/2), int((kp[i,11,1] + kp[i,12,1])/2))
        cv2.line(img_pose,(int(neck[0]), int(neck[1])), (int(kp[i,1,0]), int(kp[i,1,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,1,0]), int(kp[i,1,1])), (int(kp[i,2,0]), int(kp[i,2,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,5,0]), int(kp[i,5,1])), (int(kp[i,6,0]), int(kp[i,6,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(neck[0]), int(neck[1])), (int(hip[0]), int(hip[1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,11,0]), int(kp[i,11,1])), (int(kp[i,12,0]), int(kp[i,12,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,11,0]), int(kp[i,11,1])), (int(kp[i,13,0]), int(kp[i,13,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,13,0]), int(kp[i,13,1])), (int(kp[i,15,0]), int(kp[i,15,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,12,0]), int(kp[i,12,1])), (int(kp[i,14,0]), int(kp[i,14,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,14,0]), int(kp[i,14,1])), (int(kp[i,16,0]), int(kp[i,16,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,5,0]), int(kp[i,5,1])), (int(kp[i,7,0]), int(kp[i,7,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,7,0]), int(kp[i,7,1])), (int(kp[i,9,0]), int(kp[i,9,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,6,0]), int(kp[i,6,1])), (int(kp[i,8,0]), int(kp[i,8,1])), (255,0,0), 2)
        cv2.line(img_pose,(int(kp[i,8,0]), int(kp[i,8,1])), (int(kp[i,10,0]), int(kp[i,10,1])), (255,0,0), 2)
        for j in range(17):
            cv2.circle(img_pose,(int(kp[i,j,0]), int(kp[i,j,1])), int(3), (int(color_list[j,0]),int(color_list[j,1]),int(color_list[j,2])), 2)
            cv2.rectangle(img_list[indexlist[j]],(int(kp[i,j,0]+5),int(kp[i,j,1]+5)),(int(kp[i,j,0]-5),int(kp[i,j,1]-5)),(255,255,255),-1)
        cv2.rectangle(img_list[1]  ,(int(neck[0]+5),int(neck[1]+5)),(int(neck[0]-5),int(neck[1]-5)),(255,255,255),-1)
```

圖 4-17 讀取節點與繪製骨架節點圖程式碼(此圖為作者自製)

2. 生成人體遮罩及保留人物頭部

我們使用 encoder-decoder 架構的模型預測輸入人物圖像的部位分割結果。其中所使用到的 encoder-decoder 模型可解釋為編碼解碼器。一般其影像辨識模型的架構是由卷積層為主的特色提取器，而後才為全連接層為主的分類器。對於輸入圖片先進行抽取特徵，直到取得足夠的特徵量才進行下一步的分類。

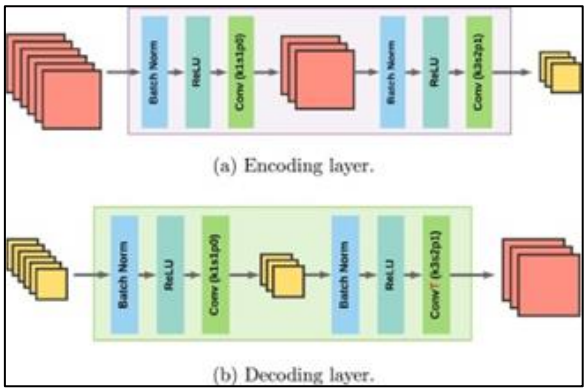


圖 4-18 encoder-decoder 模型原理圖 取自[20]

其中人體遮罩圖模型的形成是透過 LIP 數據集進行訓練。LIP 數據集中共含約 50000 張圖片，其中內容具有 19 項人體標籤及帶 16 個骨架節點的 2D 人體姿勢。於訓練的前置處理，分別將圖片資料根據後續用途分為：Testing、Training、及 Validation 三個資料集，透過不同資料集中的圖片及文字檔的訓練，以提升人體遮罩圖模型辨識影像的效果。得到分割部位的結果後，我們將頭部取出作為遮罩套在原圖上得到人物頭部的保留影像。另外我們將背景之外的人體部位進行疊加，以得到完整的人體遮罩。

(二)生成扭曲的衣物

圖片扭曲最常見的方法是使用 TPS 的函數進行計算，將圖片制定有限點作為控制點，以此將平面圖形進行扭曲。其為一種徑向基函數，藉由尋找一個通過所有控制點且彎曲程度最小的光滑曲面，以使平面進行彎曲時的彎曲能量達到最小。各控制點皆有其高度，而 TPS 參數便是改變其高度以達到扭曲的效果，擬合出結果圖。簡而言之，TPS 所變更的是圖片中控制點的 y 方向高度，而不會使 x 座標所影響的長及寬有任何變動。

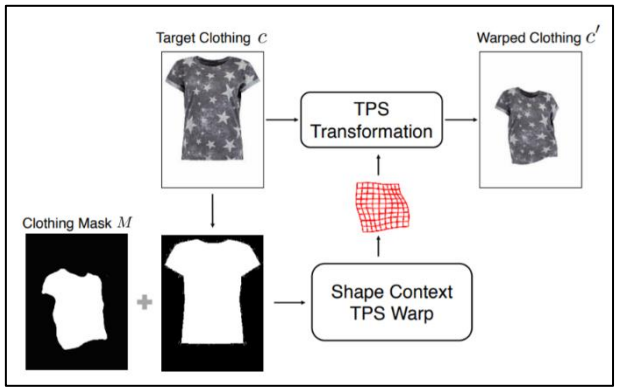


圖 4-19 TPS 模型示意圖 取自[19]

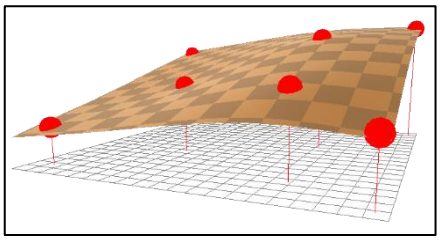


圖 4-20 含有一組控制點的二維平面圖 取自[19]

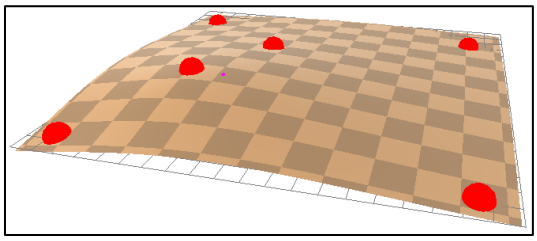


圖 4-21 經扭曲後的二維平面圖 取自[19]

在 VITON 模型中，使用了類 Unet 的 encoder-decoder 模型使其生成衣物遮罩圖，再利用此遮罩判斷 TPS 的扭曲參數，也就是在模型的第二階段中以第一階段的結果預測扭曲函數的參數。但本研究所參考的 CP-VTON 模型中將這部分獨立成一個幾何匹配模型(Geometric Matching Module，簡稱 GMM)，使用卷積網路直接將輸入的人體高維特徵和目標衣物進行相關匹配，使其合併為一個張量，作為 TPS 的變換參數去扭曲目標衣物，並以 L1 loss 對比已扭曲的衣物進行訓練。因此，此模型不再使用預訓練過的模型進行訓練，而是從同開始訓練。

1. 將人體遮罩圖、人體骨架節點圖、區域保留圖結合為人體高維特徵圖，和目標衣物的原始圖輸入至 GMM 中。
2. 接者引入衣物扭曲參數，使程式輸出一個張量，合成扭曲衣物圖。
3. 最後，再引入 L1 loss，進行輸出和模板之間相對應的元素相減後的總和。

$$\mathcal{L}_{GMM}(\theta) = ||\hat{c} - c_t||_1 = ||T_\theta(c) - c_t||_1$$

(三)Try-on 模型

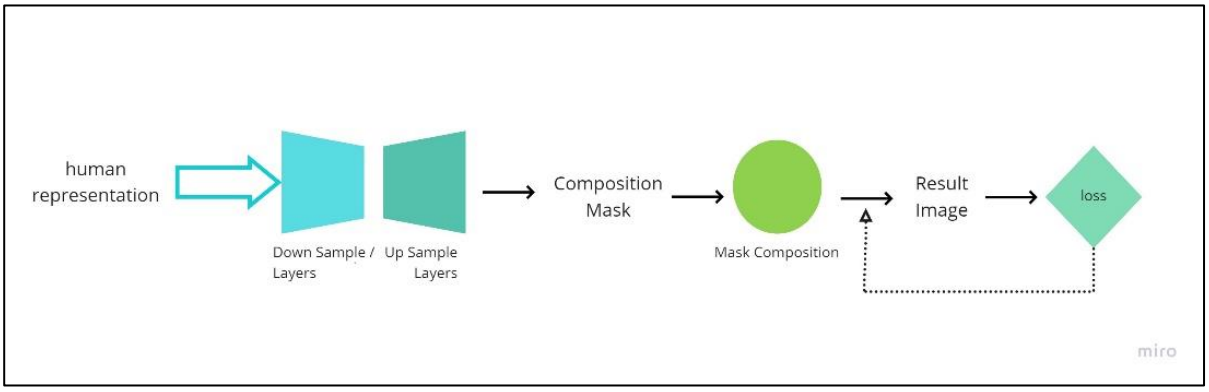


圖 4-22 Try-on 模型程式流程圖(此圖為作者自製)

Try-on 模型(Try-on Model，簡稱 TOM)中，共可分為兩大部分：合成衣物遮罩圖與生成結果圖片，其中與 VITON 試衣模型最大相異之處在於前者。CP-VTON 模型前端輸入是將人體高維特徵圖與扭曲衣物一併送至 Unet 中，直接得到粗糙衣物遮罩圖，

相較 VITON 可有效地減少其中計算所花費的時間。由圖 4-22 可知其生成結果圖片步驟為下述：

1. 由前步驟所得之人體高維特徵圖與扭曲衣物共同輸入 Unet 中，並得到粗糙衣物遮罩圖。
2. 使用人體部位渲染結果與扭曲衣物圖透過衣物遮罩圖融合成目標結果圖片 。
3. 引入 loss 函數，分別計算結果圖片及整體 Try-on 模型誤差。

對於結果圖片與測試圖片進行比對，再進一步求得 loss 值，並返回修改原先模型的偏差，即可使往後得到效果更佳的試衣模擬結果。而下式為計算整體 Try-on 模型的損失函數：

$$\mathcal{L}_{TOM} = \lambda_{L1}\|I_o - I_t\|_1 + \lambda_{vgg}\mathcal{L}_{VGG}(\hat{I}, I) + \lambda_{mask}\|1 - M\|_1$$

其中第一項為生成圖片跟真實圖片的 L1 誤差，第二項為將生成圖片與真實圖片送進 VGG 預訓練網路取其中一層的特徵圖計算 L1 誤差:

$$\mathcal{L}_{VGG}(I_o, I_t) = \sum_{i=1}^{\cup} \lambda_i \|\phi_i(I_o) - \phi_i(I_t)\|_1$$

最後一項是為了限制遮罩不變成全黑的補正項。

(四)模型整合

以上所提到的取得人體骨架節點、生成人體遮罩及保留人物頭部、生成扭曲的衣物與 Try-on 模型皆為獨立的模型。在此步驟中，我們串聯所有模型的輸入與輸出，統一成一主程式，並導入使用者介面中，成為互動系統。在介面中我們設置了兩大區域。左側區域中有兩個功能，分別為添加使用者的照片及選擇目標衣物，並顯示於介面中。而當使用者按下轉換鍵後，結果圖片將出現於右側區域中。此外我們提供了存檔的功能並設有評論表單供使用者提供改善意見。

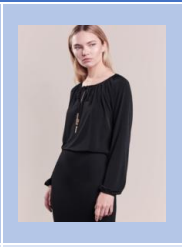
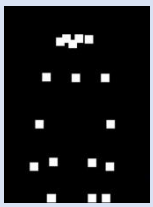
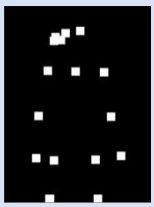
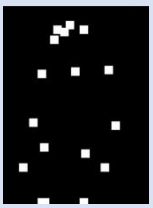
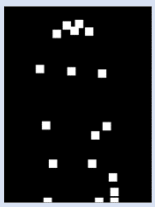
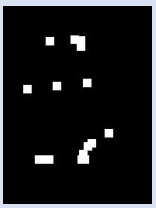
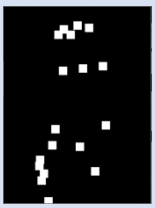
圖 4-23 使用者介面(此圖為作者自製)

伍、研究結果與討論

一、人體特徵表示

(一) 人體骨架節點取得結果：

表 5-1 原始圖片及人體骨架節點繪製圖比對表

	A	B	C	D	E	F
原始圖片						
骨架節點繪製						

由表 5-1 中可看出原始圖片與節點繪製圖的對應關係。其中表 5-1 中的 E 與其餘圖片差異在於原始人物圖中腿部是否明確。由 18 個骨架節點順序示意圖(圖 4-13)與實作出的節點繪製圖對比後可發現。當輸入影像不包含完整人體腿部時，原本屬於腿部的節點 (編號 9、10、12、13) 將無法正確被偵測，會出現不規則分布的情況。

(二) 人體遮罩圖生成結果：

表 5-2 原始圖片與人體遮罩圖比對表

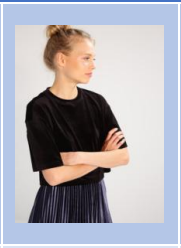
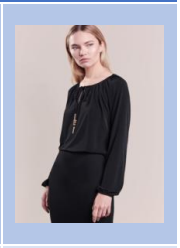
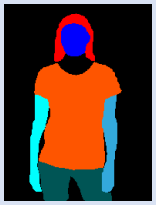
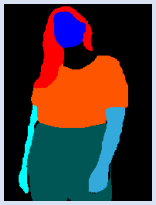
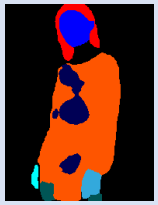
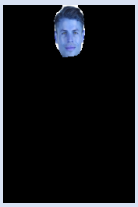
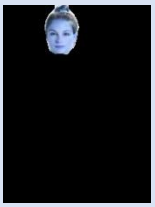
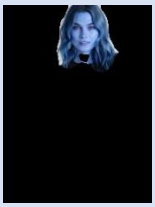
	A	B	C	D	E	F
原始圖片						
人體遮罩圖						

表 5-2 分別為原始圖片及透過 Human parsing 模型取得之人體遮罩圖。透過比對後，可知 LIP 數據及中並不包含頸部區域的標註。且根據其原始圖片的站姿及角度，對於其辨識遮罩結果皆有所影響。舉表中 F 行為例，原始圖片中人物身體方向並非面向正前方，導致其人體遮罩圖有破洞。此外我們觀察 D、E 兩行發現，儘管有重疊遮擋的地方，依舊能有效的辨識。

(三) 原始圖片人物頭部保留結果：

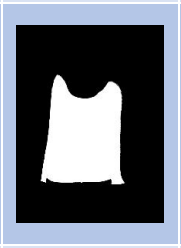
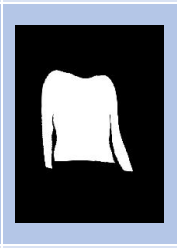
表 5-3 原始圖片及人體頭部保留圖比對表

	A	B	C	D	E	F
原始圖片						
人體頭部保留						

從運算出的結果中可發現，原始圖片中的人物保留辨識選取僅包含頭部，而未含頸部的部分。人物保留和人體遮罩圖的辨識選取皆未包含頸部。導致在後續模擬目標人物時，頸部會與實際圖片產生誤差。此外，經比對後發現，其頭髮無法被程式精細的區分出，使最終得到的結果圖會產生多餘的空白區域導致誤差。

二、生成扭曲衣物及遮罩

表 5-4 GMM 模型的運算結果比較表

	A	B	C	D	E	F
原始衣物						
扭曲衣物						
衣物遮罩						
目標圖片						

(一) 扭曲衣物生成結果：

此步驟將原始衣物及人體高維特徵圖一同輸入 GMM 模型中，使其預測 TPS 參數以生成衣物扭曲圖。而表 5-4 為實驗中所輸出的結果。比對表中 C、D、E 行後可發現：原始衣物為長袖者，在後續 GMM 模型轉換的步驟中，效果比原始衣物為短袖者稍遜。對此我們推論出 GMM 模型中原始衣物與扭曲衣物的款式會影響扭曲衣物轉換的準確度，且因長袖需在短距離中做大幅度的扭曲而導致其準確度難以提升。

(二) 衣物遮罩圖之合成：

由表 5-4 中所呈現圖片中可發現：原始衣物扭曲圖與衣物遮罩圖會互相影響。若有其中一個的結果較差，另一個也會有所誤差。

三、試衣合成模擬

本研究中所使用之資料包含：公開數據集之訓練資料、測試資料、及實作圖片。訓練資料與測試資料為 Try-on 模型使用的公開數據集，而實作則為實際拍攝之照片或自行收集資料。

三者之間輸入的資料集差異在於：訓練資料於輸入時，會同時輸入原始圖片與正確答案，因此可以透過模擬結果與正確答案間的比對，進行模型的調整與修正。而測試階段，僅輸入原始圖片，並對所生成的模擬試衣結果與正確答案比對、進行誤差值之計算。實作圖片輸入時只有原始圖片，且無正確答案可以比對，即模擬真實使用者的使用情況。

本研究中試衣模擬之偏差度是透過將結果圖片與原始圖片的色彩做點對點的相減再取平均值。為避免人物以外 RGB 為 0 的白色區域影響到平均值，故在相減之後乘以人體遮罩圖，僅取出人體的部分求平均。除實作圖片以外的訓練資料與測試資料的偏差度，皆為取約 300 張相同條件的資料取平均的結果。

(一) 探討原始衣物圖案對於模擬試衣結果的影響

表 5-5 相異衣服圖案對模擬試衣結果影響比對表

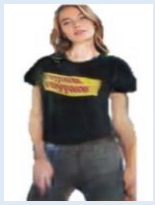
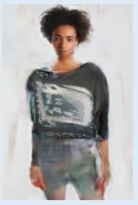
	A:純色	B:複雜	C:純色	D:複雜	E:純色	F:複雜
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
偏差度	0.4018195	0.4305497	0.4407360	0.4607401	0.2614644	0.4554980

表 5-5 資料列舉了訓練、測試、實作資料集中，以純色與複雜圖案為變因，所得到的結果範例和偏差度。由表中資料可彙整成圖 5-1，分別將各項資料進行比對後可發現，各類資料集的皆為純色之偏差度小於複雜圖案者。根據前述個步驟結果我們分析，會造成複雜圖案被過度扭曲、模糊，是因為細節的材質較難被重建。於是據此結果推論，當輸入之原始衣物越為單純，其模擬結果越佳，且衣服為純色者最佳。

據此實驗所得，以下實驗皆以純色衣服為主進行操作。

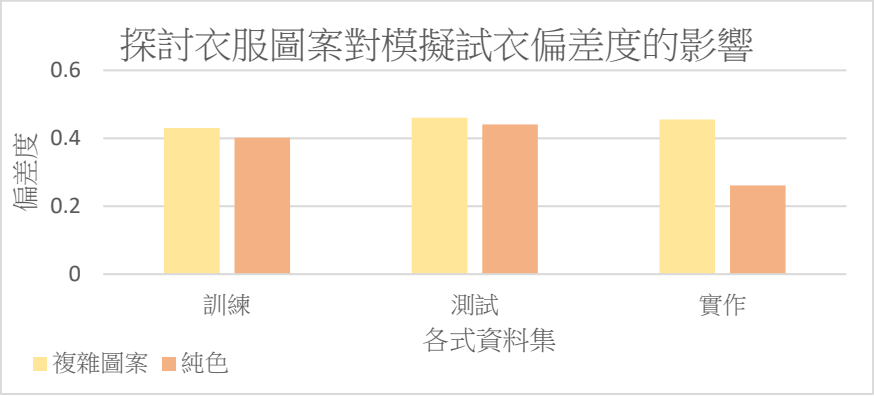
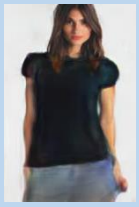


圖 5-1 衣服圖案對模擬試衣偏差度的影響比較圖(此圖為作者自製)

(二) 探討原始衣物色彩對比度對於模擬試衣結果的影響

表 5-6 相異衣服色彩對比度對模擬試衣結果影響比對表

	A:差異大	B:差異小	C:差異大	D:差異小	E:差異大	F:差異小
來源	訓練資料		測試資料		實作圖片	
對比度	18.589052	1.069871	17.730379	1.080712	12.072189	1.001295
原始圖片						
原始衣物						
結果圖片						
偏差度	0.4043193	0.6106393	0.4172408	0.6232005	0.2614644	0.7087366

本實驗透過 Contrast Ratio 比較原始衣物與其背景之顏色對比度。顏色對比度是指兩相鄰顏色之間的亮度或發光強度差異值，此比值介於 1 到 21 之間。且數字越大表示其對比度越高，以純黑白對比其比值為 21。

表 5-6 資料列舉了訓練、測試、實作資料集中，以原始衣物顏色與其背景色彩對比度差異大或小為變因，所得到的結果範例與偏差度。根據表 5-6 中可整理出圖 5-2，比對後可發現：當原始衣物與其背景色彩對比度越高，其試衣結果與實際結果之偏差度越低。我們根據生成扭曲衣物的結果分析，會使淺色衣物出現過度扭曲或缺漏是因 GMM 在生成扭曲衣物時無法圈選出正確的衣服位置。於是綜合上述實驗結果推論，模型輸入的原始衣物顏色與其背景色彩對比度越高，尤以黑色為佳，其模擬結果效果越好。

據此實驗所得，以下實驗皆以原始衣物與背景色彩對比度較大之原始衣物為主進行探討。

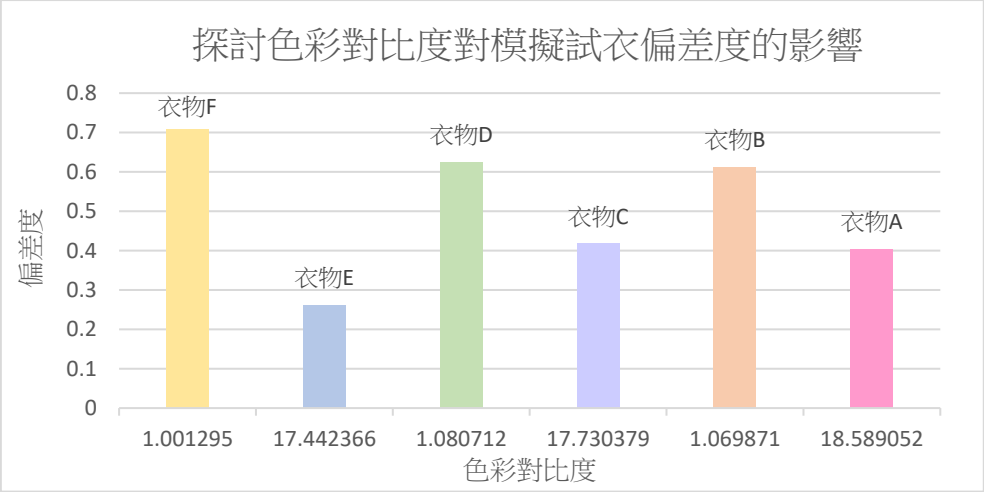


圖 5-2 色彩對比度對模擬試衣偏差度的影響比較圖(此圖為作者自製)

(三) 探討相異原始衣物款式對於模擬試衣結果的影響

表 5-7 相異衣服款式對模擬試衣結果影響比對表

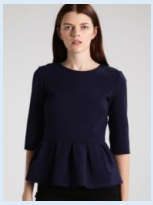
	A:短袖	B:長袖	C:短袖	D:長袖	E:短袖	F:長袖
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
偏差度	0.4031307	0.4073664	0.4486451	0.4579029	0.2614644	0.3609284

表 5-7 列舉了訓練、測試、實作資料集中，以兩種相異衣服款式的為變因，分別為短袖衣物及長袖衣物，所得到的結果範例與偏差度。根據表 5-7 中資料可整理出圖 5-3，比對後可發現：原始衣物款式為短袖者，其得到結果圖片偏差度小於款式為長袖者。我們根據生成扭曲衣物及遮罩的結果分析，是因為長袖手臂扭曲幅度很大導致效果模糊或不正確的扭曲。綜合上述實驗結果推論，當輸入的原始衣物款式為短袖者，在模型後端得到之結果圖片效果皆會優於長袖者。

據此實驗所得，以下以下實驗皆以短袖為主進行實驗

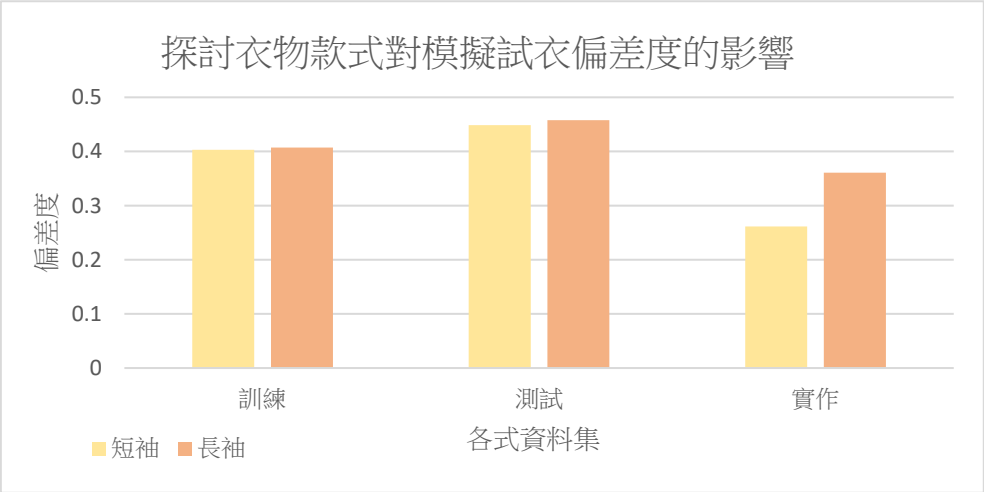


圖 5-3 衣物款式對模擬試衣偏差度的影響比較圖(此圖為作者自製)

(四) 探討相異原始圖片人物拍攝姿勢對於模擬試衣結果的影響

表 5-8 相異拍攝姿勢對模擬試衣結果的影響

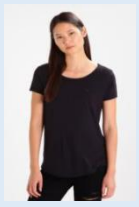
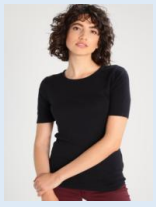
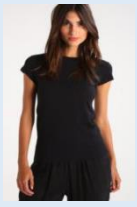
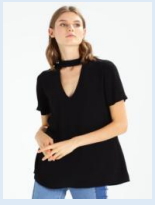
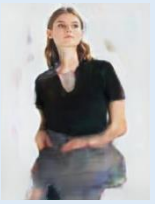
	A:基本	B:變化	C:基本	D:變化	E:基本	F:變化
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
偏差度	0.4077984	0.4235286	0.4358860	0.4520450	0.2614644	0.4330302

表 5-8 資料列舉了訓練、測試、實作資料集中，以原始圖片中人物的不同拍攝姿勢為變因，所得到的結果及其偏差度。根據表 5-8 中資料可統整出圖 5-4，透過比較可發現：不論為何資料集，其人物站姿越為單純，結果偏差度越低。我們根據結果分析，不標準站姿會出現缺漏以及覆蓋，是因訓練 GMM 以及 TOM 的資料集中這樣的姿勢

較少，導致不論是扭曲衣物還是最終結果都有誤差。於是據上述可推論，當站姿越單純，例如手部貼身且不交叉，其模型模擬出結果圖片的效果較為佳。

據此實驗所得，以下其他實驗以原始圖片人物站姿為單純站姿為主進行探討。

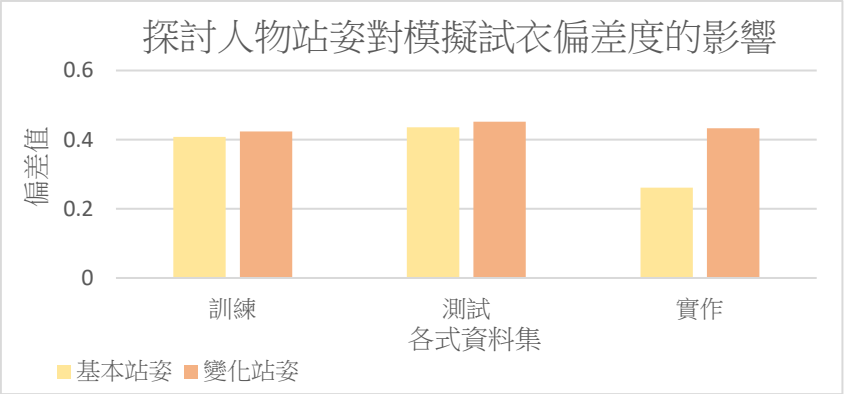


圖 5-4 人物站姿對模擬試衣偏差度的影響比較圖(此圖為作者自製)

(五)探討相異原始圖片拍攝角度對於模擬試衣結果的影響

表 5-9 相異拍攝角度對模擬試衣結果影響比對表

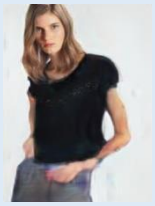
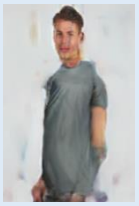
	A:正面	B:側面	C:正面	D:側面	E:正面	F:側面
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
偏差度	0.4042671	0.4122761	0.4412514	0.7676513	0.2614644	0.352851

表 5-9 資料列舉了訓練、測試、實作資料集中，以原始圖片的拍攝角度為變因，所得結果與偏差度。由表 5-9 中資料中可整理成圖 5-5，分別將圖中資料進行比對可發現，對於各類資料集，其結果偏差值皆為正面拍攝角度低於側面拍攝。我們根據結果分析，非正面的拍攝角度會使衣服無法隨角度轉換並正常覆蓋人體，是因為資料集中缺乏非正面角度拍攝的影像。由上述結果可推論：拍攝角度為標準正面者其輸出結果會佳於側面、俯視或仰角等不標準的角度。

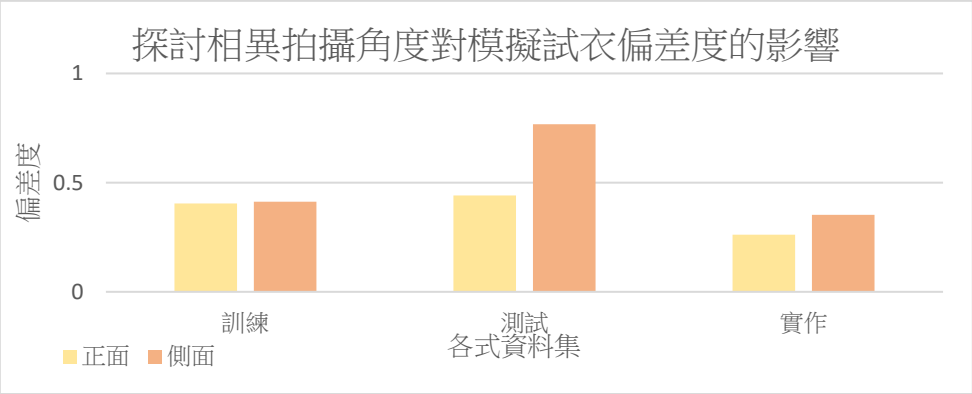


圖 5-5 拍攝角度對模擬試衣偏差度的影響比較圖(此圖為作者自製)

陸、結論

本研究是以數個深度學習模型，模擬圖片中人物穿著目標衣物。經過實作、分析與討論後，在這次的研究中我們參考了 CP-VTON 分解出裡面的每個步驟的輸出輸入與不同功能的模型，整合成完整程式。並於實作測試中找出這個模型可能會有的侷限，在未來可進一步的修改。以下是我們經過大量資料實驗後所得出的結論：

(一) 原始圖片人物身體站姿、角度會影響扭曲衣物及人體遮罩圖的成效，進而影響到模擬試衣的結果。且原始圖片人物拍攝角度為正面者成效會優於側身。

(二) 原始衣物的衣服款式會影響扭曲衣物的成效並進而影響到模擬試衣的結果。且原始衣物款式為短袖者成效會優於長袖。

(三) 原始衣物圖案越為單純，其試衣結果越佳。

(四) 原始衣物與背景色彩對比度越高，其試衣結果成效越佳。

(五) 原始圖片人物站姿較為單純者，其試衣結果成效越佳。

柒、未來展望

我們希望未來的研究中可考量到更多不同影響模擬試衣結果的因素，並將研究之成果真正的應用於網路商城上，對此我們列舉了以下四點分別敘述。

一、去除多餘雜訊

目前的結果圖像中會出現許多雜訊，未來可透過建立深度學習模型以分辨何為雜訊、找到雜訊和去除雜訊。並使圖片顏色平滑，色彩拼接自然。生成的手部部分在形狀與顏色上都不太自然，未來可以在加上生成對抗的訓練來輔助生成接近原圖的效果

二、應用於多人的照片

目前所使用的模型只適用於單人照片，希望未來可以藉由模型進一步的修改，使其得以在多人的照片中為多位使用者替換其他衣物。

三、針對特殊姿勢

未來可以透過增加特殊姿勢的訓練資料，以解決模型容易在面對手部姿勢重疊、及身體角度的改變等不同姿勢時發生局部模糊不清的現象。期望未來可透過模型的調整讓使用者在更多不同姿勢下，皆可利用此模型測試衣物是否適合自己。

四、適用於更多種類的服飾

本研究的主要研究方向為衣物的置換，但現今網路商城販售的衣物品項更加的五花八門，例如:帽子、圍巾、襪子等。期望未來可以透過打破模型中區域保留僅包含頭部的限制，使此模型的應用範圍可延伸的更加全面。

五、跟隨社會發展趨勢

現代社會中穿著衣物的已經不僅限於人類，隨著愈來愈多寵物的飼養，試衣系統理應可以發展至更多不同方面。由於本篇研究的模型僅適用人體骨架的試穿，希望未來可以透過模型的更改使其符合不同物種的骨架，讓此模型不僅僅適用於人類，更可以替貓、狗等不同動物換上各式衣物。

捌、參考資料

[1] Bochao Wang , Huabin Zheng , Xiaodan Liang , Yimin Chen , Liang Lin , and Meng Yang.(2018) Toward Characteristic-Preserving Image-based Virtual Try-On Network.

[2] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, Larry S. Davis.(2017) VITON: An Image-based Virtual Try-on Network.

[3] Amit Raj , Patsorn Sangkloy, Huiwen Chang, James Hays , Duygu Ceylan, and Jingwan Lu.(2018) SwapNet: Image Based Garment Transfer.

[4] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation.

[5] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.

[6] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation.

[7] Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields.

[8] Github Look-Into-Person-v2 Retrieved from https://github.com/foamliu/Look-Into-Person-v2.

[9] Thin plate splines 薄板樣條插值個人理解 Retrieved from https://www.twblogs.net/a/5b8de0022b7177188341385c

[10] Autosport labs Choosing between MAP or TPS Retrieved from https://wiki.autosportlabs.com/Choosing_between_MAP_or_TPS

[11] Diver 薄板樣條插值 Retrieved from https://hideoninternet.github.io/2019/11/06/d3c15ac3/

[12] 鍊聞 CHAINNEWS 圖像分割中的深度學習：U-Net 體系結構 Retrieved from https://www.chainnews.com/zh-hant/articles/369337679143.htm

[13] Medium Deeplab v3+ Retrieved from https://medium.com/%E8%BD%89%E8%81%B7%E8%B3%87%E5%B7%A5%E8%BF%B7%E9%80%94%E8%A8%98/deeplab-v3-3a105519a0cf

[14] 人體解析數據集（human parsing）及近期論文 Retrieved from https://www.twblogs.net/a/5c310044bd9eee35b21ca00c?lang=zh-cn

[15] Neural Convolutional layers Retrieved from
https://m-alcu.github.io/blog/2018/01/13/neural-layers/

[16] Implememnation of various Deep Image Segmentation models in keras Retrieved from
https://pythonawesome.com/implememnation-of-various-deep-image-segmentation-models-in-keras/

[17] 姿態估計之 COCO 數據集骨骼關節 keypoint 標註對應 Retrieved from
https://www.stubbornhuang.com/525/,

[18] Bayesian deep convolutional encoder – decoder networks for surrogate modeling and uncertainty quantification Retrieved from https://www.sciencedirect.com/science/article/pii/S0021999118302341

[19] Manual Registration with Thin Plates Retrieved from https://profs.etsmtl.ca/hlombaert/thinplates/

[20] Bayesian deep convolutional encoder – decoder networks for surrogate modeling and uncertainty quantification Retrieved from https://www.sciencedirect.com/science/article/pii/S0021999118302341

【評語】190014

本作品結合深度學習技術與生活經驗進行發想，研究自動化的服裝試衣系統，並且透過程式開發，實際驗證該系統的效能，在研究方法與研究步驟上皆展現良好的科學方法與精神。本作品若能就各項引用的技術，更進一步深入討論適切性與可能的優化方法，同時加強與其他類似系統的比較與分析，將能更凸顯本研究的價值貢獻，也更能展現本研究的科學內涵與精神。