中華民國第 61 屆中小學科學展覽會

作品說明書

高級中等學校組　電腦與資訊學科

探究精神獎

052504

基於深度學習之服裝試衣系統

學校名稱：國立新竹女子高級中學

作者：	指導老師：
高二 楊子誼 高二 林維余 高二 徐熙筠	古佳怡

關鍵詞：衣服試穿、深度學習、幾何匹配模型

摘要

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

首先，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。而後將三個輸出合成為高維特徵圖。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共同輸入，並經過運算後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得出的結果圖片與測試圖片相似度較高。

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的發展。而近日新冠肺炎疫情肆虐，更加凸顯了網路商城的便利之處在於不須出門也可以購買生活所需之物品，又能夠有效地降低與人發生接觸的機會，而其中又以購買衣服最為熱門。但由於網路購衣的緣故，消費者無法直接試穿衣服。時常發生衣服效果與原先想像不符，而需要退換貨的情形。而漫長的退換貨週期和手續對於買家來說也是降低了體驗值，更在過程中造成資源浪費。因此，我們想要運用所學的知識，透過深度學習模型建構出模擬試衣系統。

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此衣服是否符合心理的預期。此外我們希望藉由輸入以相異變因分類出的圖片進行比對，藉此探討不同試衣模型在各類情形下之優缺點。

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

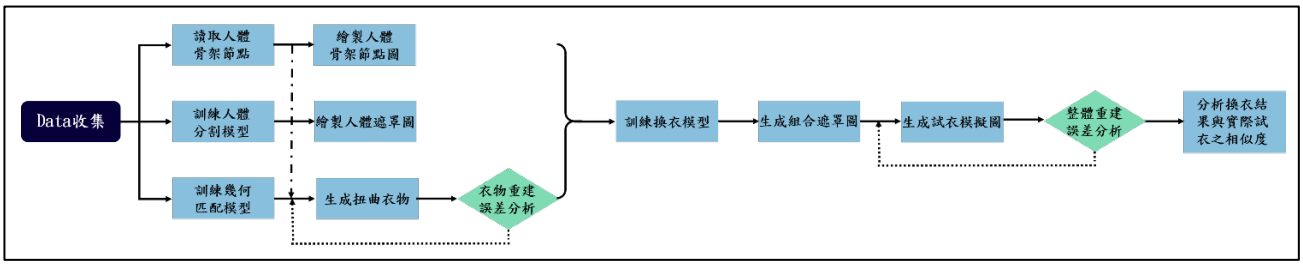


圖 4-1 研究流程圖(此圖為作者自製)

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) 虛擬試衣的發展

最初虛擬試衣系統的發展是使用 3D 人體建模的方式，透過剪貼將目標衣物換上目標人物。此方法對於圖像的掌握度高且精確，但其需要大量運算及精密的儀器對於人體進行三維的掃描，在建立資料庫及訓練模型的過程中往往耗費大量時間與金錢投入。因此此項技術多使用於服裝公司自行推出的虛擬試衣鏡，無法普及至生活當中。隨深度學習技術的發展，基於圖像生成演算法以進行虛擬試衣便成了另一種可能。其主要是透過將衣服圖像輸入模型中進行扭曲、拼接到目標人物的圖像上，以產生試衣結果。而使用此方法之最為經典者為 VITON 模型，其中包含了利用薄板

樣條插值（Thin plate spline，簡稱 TPS）和多任務學習的技術，本研究所參考的 CP-VTON 模型即是基於此基礎進行變化。

(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所處理的範圍。圖 4-2 展示了人體部位的影像分割範例，以左圖為輸入、右圖為輸出，且大小和原圖一樣、標註每個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 4-2 影像分割範例 取自[15]

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，相較於常見的分類問題，前者會輸出每個像素的類別資訊，但後者只有輸出圖片整體的類別資訊。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，以得到與原圖大小相同的分類結果。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，使用者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。

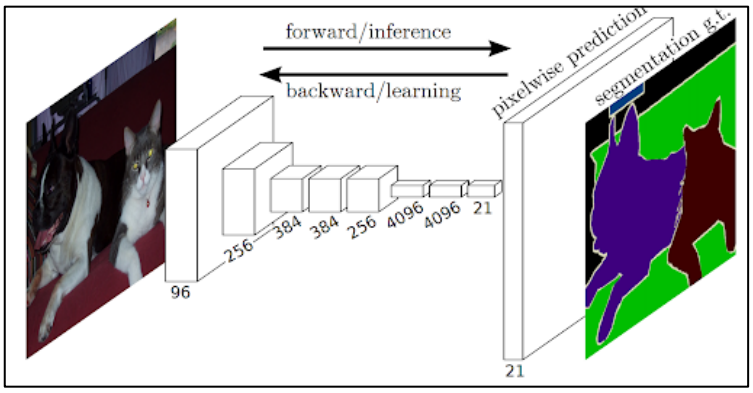


圖 4-3 全卷積網路模型架構 取自[17]

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 4-4 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。

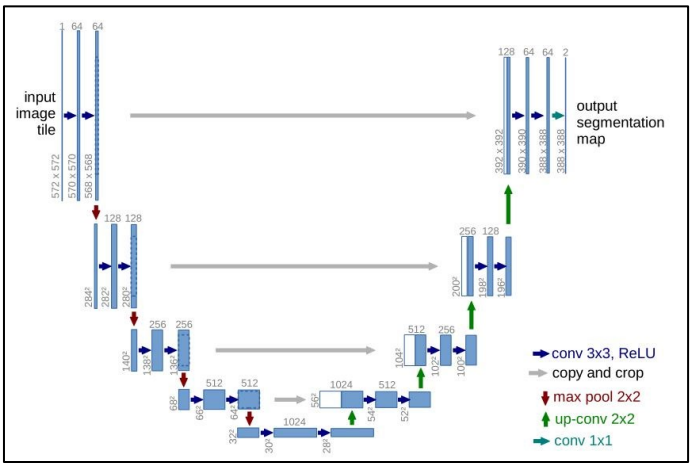


圖 4-4 Unet 原理圖 取自[13]

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 4-5 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。

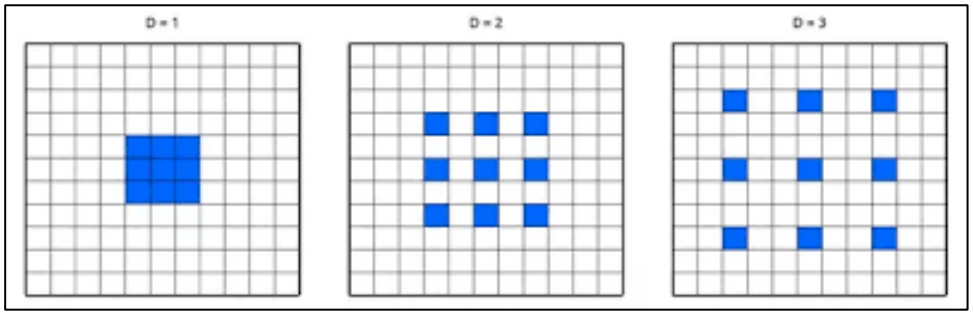


圖 4-5 Atrous Convolution 原理圖 取自[5]

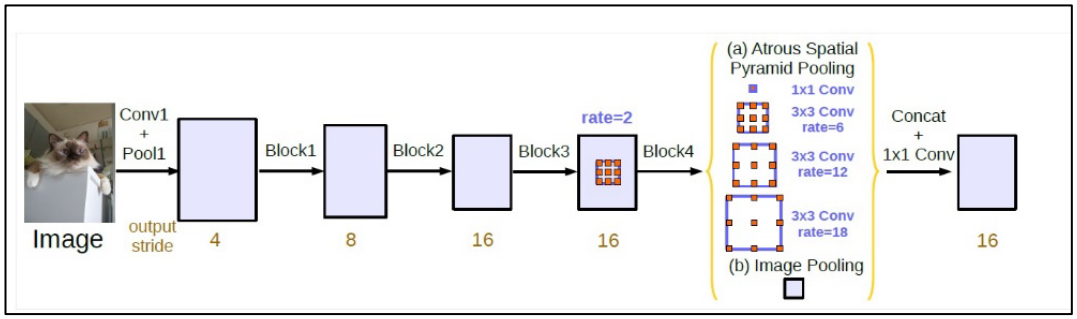


圖 4-6 Deeplab v3 架構圖 取自[7]

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 4-7 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。

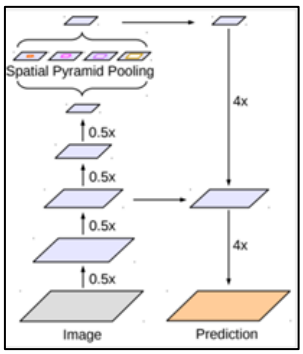


圖 4-7 Deeplab v3+改良原理圖 取自[3]

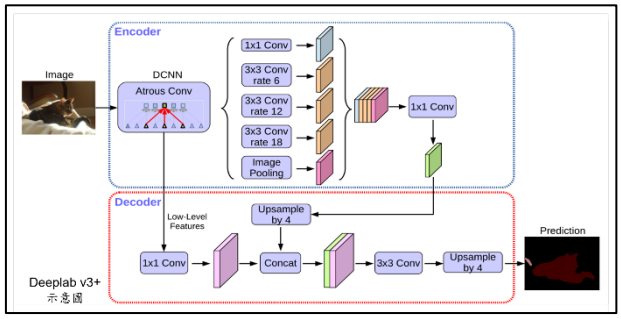


圖 4-8 人體遮罩圖生成原理圖 取自[5]

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。

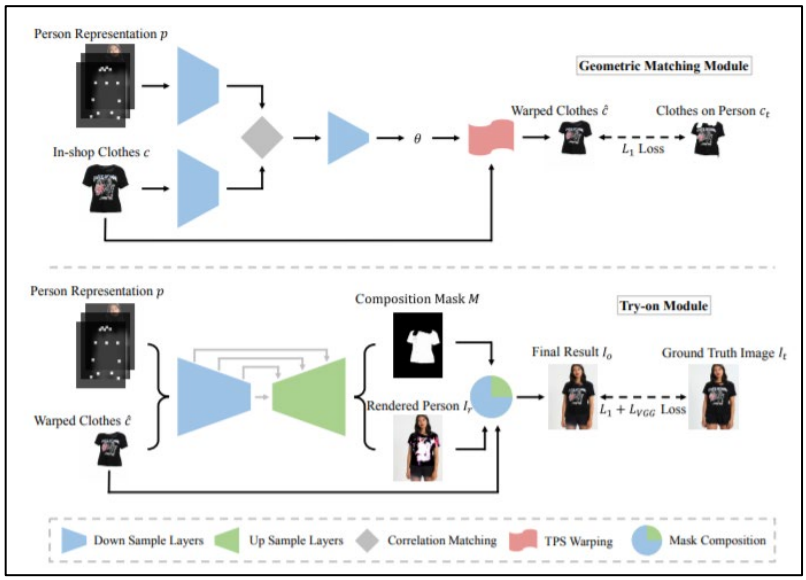


圖 4-9 CP-VTON 流程圖 取自[2]

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。以下的說明分為六部分：第一、二部分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；第三部分為介紹研究中最核心的模型概念；第四至六部分為將以上三部分串接成完整程式，引入我們設計的使用者介面，並優化輸出結果以及導入結果分析方式。

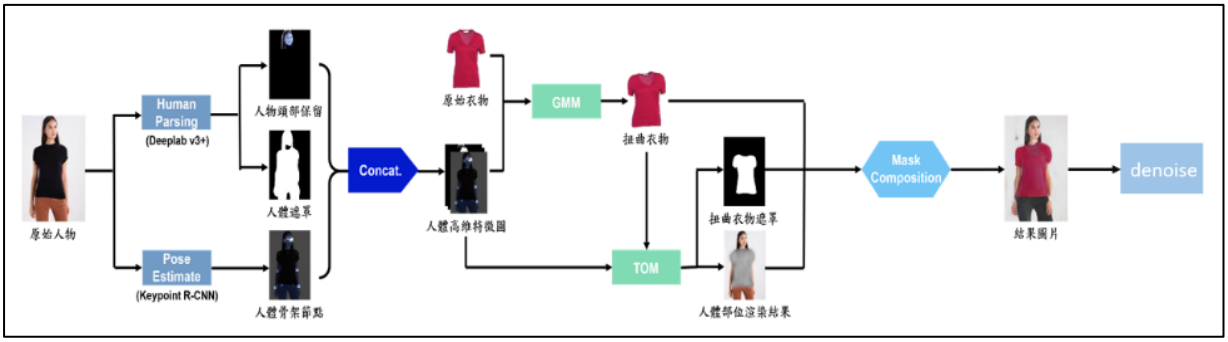


圖 4-10 程式流程圖(此圖為作者自製)

(一)建構人體高維特徵圖

人體高維特徵圖為虛擬試衣模型的輸入，其中包含了三種資訊，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。

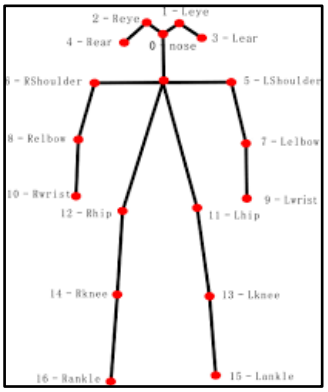


圖 4-11 17 個節點順序示意圖 取自[18]

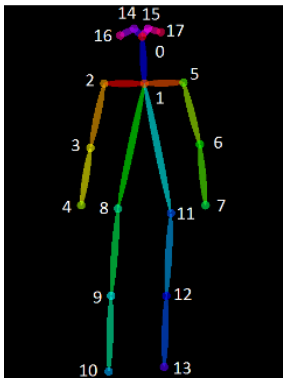


圖 4-12 18 個節點順序示意圖 取自[9]

(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 4-13 測試原圖(取自網路)

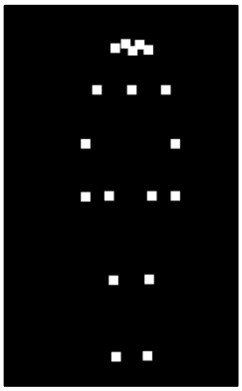


圖 4-14 骨架節點繪製圖(作者實作)

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。

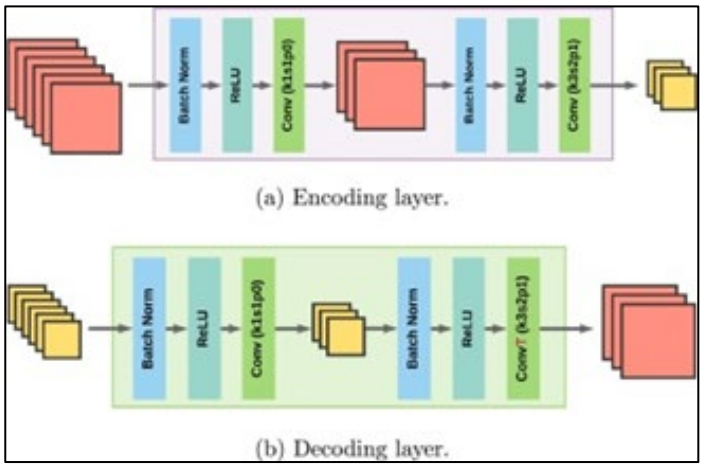


圖 4-15 encoder-decoder 模型原理圖 取自[21]

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，內容含有 19 項人體標籤及帶 16 個骨架節點的 2D 人體姿勢。於訓練的前置處理中，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，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，再以此將平面圖形進行扭曲。TPS 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。簡述之，TPS 所變更的是圖片中控制點的 y 方向高度，不會使 x 座標所影響的長及寬有

任何變動。

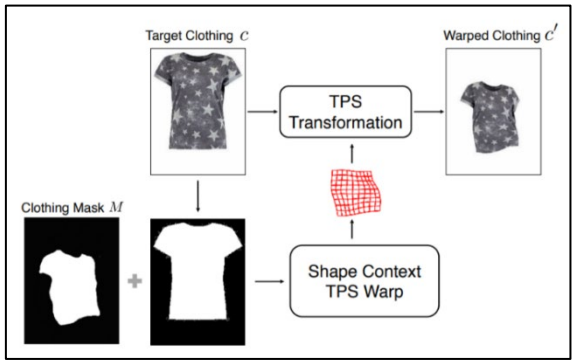


圖 4-16 TPS 模型示意圖 取自[19]

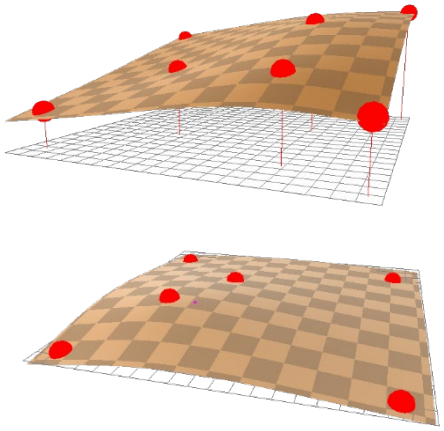


圖 4-17 經扭曲後的二維平面圖 取自[19]

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資料進行訓練，而是從原始數據開始訓練。建構扭曲衣物模型之步驟為下述：

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. 最後引入 L1 loss，計算輸出和模板之間相對應的元素相減後的總和。下式為 GMM 模型所使用的 loss 函數。

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

(三)Try-on 模型

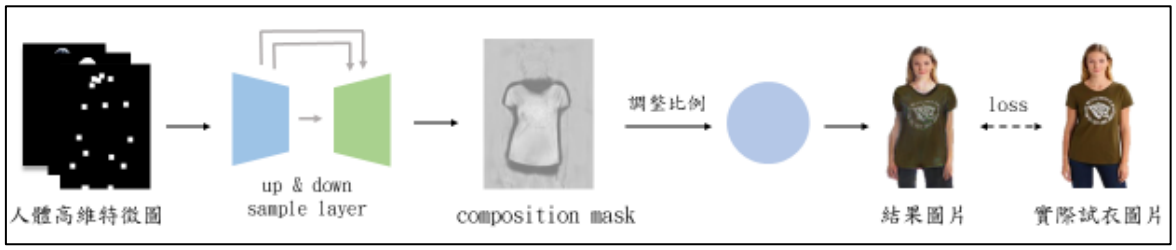


圖 4-18 Try-on 模型程式流程圖(此圖為作者自製)

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 4-18 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. 調整粗糙衣物遮罩圖與區域保留圖保留比例，使兩者合成目標結果圖片 。
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：

$$\mathcal{L}_{TOM} = \lambda_{L1}\|I_0 - I_t\| + \lambda_{VGG}\mathcal{L}_{VGG}(\hat{I}, I) + \lambda_{mask}\|1 - M\|_1$$

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，計算方式依照下列算式。而最後一項是為了限制遮罩不變成全黑的補正項。

$$\mathcal{L}_{VGG}(I_0, I_t) = \sum_{i-1}^{5} \lambda_i \|\emptyset_i(I_0) - \emptyset_i(I_t)\|_i$$

(四)去除背景雜訊

此系統的輸出圖片常會因不規則的雜訊而在背景出現混濁色塊。為了解決此一

問題，我們於系統中引入另一模型以確保最後取得的結果圖片之背景為乾淨的。在此步驟中，我們使用了 Pytorch 以 COCO 2017 訓練集上的其中一子集所預訓練過的 Deeplabv3-ResNet101。其步驟為下述：

1. 從此模型中取得結果圖片的遮罩並將其乘上結果圖片，以保留原始圖片的人物區塊。
2. 將原始圖片和遮罩圖相減，取得原始圖片的背景遮罩。再將此遮罩乘上原始圖片，取得原始圖片的背景區塊。
3. 將前述兩步驟所得的人物和背景區塊相加，得到完整的模擬試衣圖片。

(五)相似度計算

為取得試衣系統所模擬出的結果圖片和目標圖片之相似度，我們引入了 tensorflow 中的 SSIM 計算函數作為指標的計算，以分析模型於不同分類下進行衣物置換之結果的優劣。SSIM 的輸入為兩張圖，輸出值為一個 0 和 1 之間的數值，數值越大表示結果圖片和原始圖片的相似度越高，即試衣模擬的置換結果越好。此研究中均是以原始衣物作為目標置換衣物，因此相似度是以原始圖片和結果圖片進行計算。下式為 SSIM 的計算函數：

$$SSIM(x,y) = [1(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma}$$

第一項$1(x,y)$在 SSIM 的數學定義中是亮度比較，第二項$c(x,y)$是對比度比較，第三項$s(x,y)$則是結構比較。另外 SSIM 具有對稱性，即$SSIM(x,y) = SSIM(y,x)$。

(六)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，並將程式導入使用者介面中。在介面中我們設置了兩大區域。第一個區域中有兩個功能，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，結果圖片將出現於第二區域中。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。

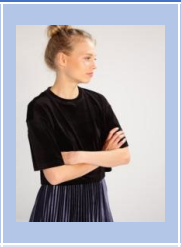
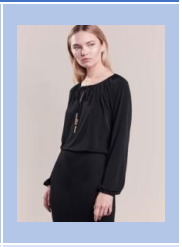
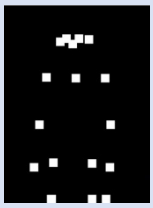
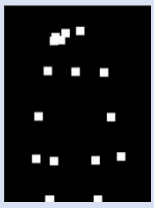
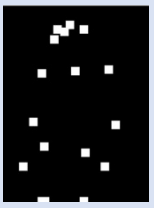
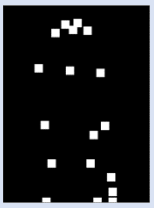
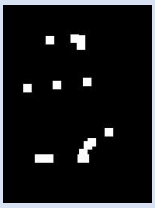
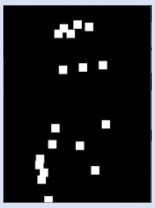
圖 4-19 使用者介面(此圖為作者自製)

伍、研究結果

一、人體特徵表示

(一) 人體骨架節點取得結果：

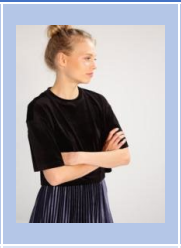
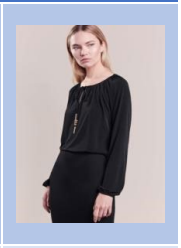
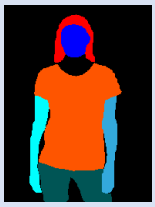
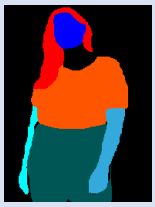
表 5-1 原始圖片及人體骨架節點繪製圖比對表

	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。其中可見腿部的 A 至 D 可明確繪製出節點，E、F 則在腿部節點的區塊出現不規則分布的現象。

(二) 人體遮罩圖生成結果：

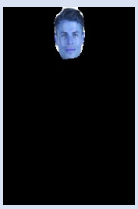
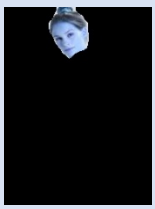
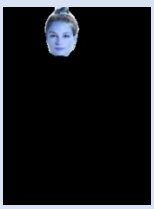
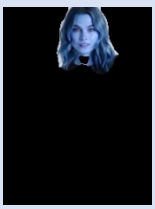
表 5-2 原始圖片與人體遮罩圖比對表

	A	B	C	D	E	F
原始圖片						
人體遮罩圖						

由表 5-2 中可看出原始圖片與人體遮罩圖的對應關係。其中可見腿部的 A 至 D 可明確區分出人體各部位；側身站立的 E、F 中，E 可明確區分出人體各部位，F 則出現破裂的情形。

(三) 原始圖片人物頭部保留結果：

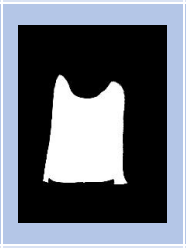
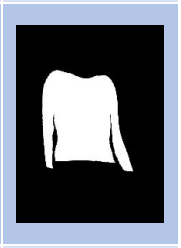
表 5-3 原始圖片及人體頭部保留圖比對表

	A	B	C	D	E	F
原始圖片						
人體頭部保留						

由表 5-3 中可看出原始圖片與節點繪製圖的對應關係。其中平頭的 A 及束起頭髮的 B、C 之人體頭部保留皆較貼合頭部輪廓且白邊較少；沒有束髮的 D、E、F 之人體頭部保留皆擷取到較多白邊。

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						

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與衣物遮罩圖。比對表中 C、D、E 行後可發現：原始衣物為長袖者，效果較原始衣物為短袖者為差。

三、試衣合成模擬

表 5-5 相異原始圖片輸入比對表

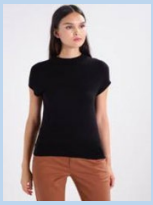
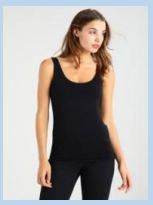
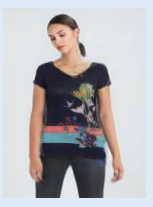
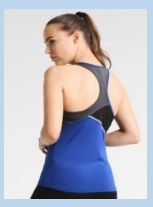
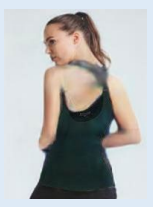
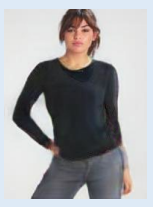
	A	B	C	D	E	F
原始圖片						
結果圖片						
	G	H	I	J	K	L
原始圖片						
結果圖片						

表 5-5 中列舉了我們將各式原始圖片作為輸入所得到的結果圖片，由上表中可以發現透過此系統各式原始衣物進行模擬皆能得到不錯的結果圖片，例如上表中的 D、E 行中進行長短袖的衣物交換，並無發生缺漏或不正常覆蓋的現象；又 J 行中，雖然原始圖片人物拍攝姿勢並非面向正前方，但透過類似衣物款式進行模擬亦能得到不錯的試衣結果。因此我們想加以探討不同變因對於此虛擬試衣系統模擬成效的影響，且藉此提高結果圖片與實際試衣結果圖片相似度。

而在以下研究中所使用之資料共可分為三類：分別為訓練資料、測試資料、及實作圖片。訓練資料與測試資料為 Try-on 模型使用的資料集，而實作則為實際拍攝之照片或取自網路之圖片。

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。

(一) 探討原始衣物圖案對於模擬試衣結果的影響

表 5-6 相異衣服圖案對模擬試衣結果影響比對表

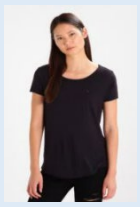
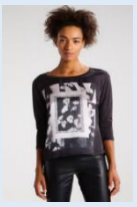
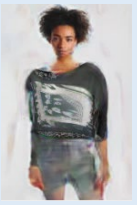
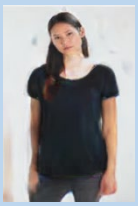
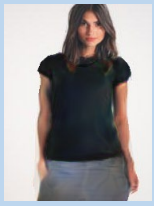
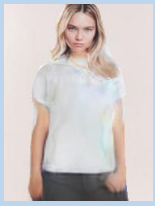
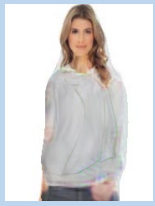
	A:純色	B:複雜	C:純色	D:複雜	E:純色	F:複雜
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
相似度	0.7435926	0.6852192	0.7388035	0.6441410	0.7143156	0.67436737
去除雜訊圖片						
去雜訊相似度	0.7824041	0.7222175	0.7795442	0.6824085	0.7698996	0.7308249

表 5-6 資料列舉了訓練、測試、實作資料集中，以純色及有圖案之原始衣物進行實驗，所得到的其中數百項結果計算其結果圖片與測試圖片之平均相似度。由表中可見純色資料所得之結果皆較有圖案之資料為佳。

(二) 探討原始衣物色彩對比度對於模擬試衣結果的影響

表 5-7 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						
相似度	0.7371262	0.6970588	0.7280142	0.6808669	0.7143156	0.6977324
去除雜訊圖片						
去雜訊相似度	0.7762095	0.7381903	0.7674871	0.7223954	0.7698996	0.7257591

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 5-7 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， 所得到的其中數百項結果計算其結果圖片與測試圖片之平均相似度。由表中可見對比度大之資料所得之結果皆較對比度小之資料為佳。

(三) 探討相異原始衣物款式對於模擬試衣結果的影響

表 5-8 相異衣服款式對模擬試衣結果影響比對表

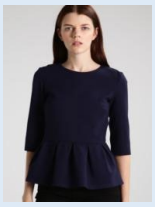
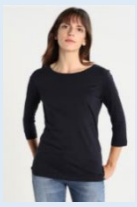
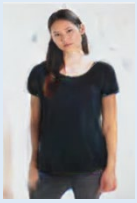
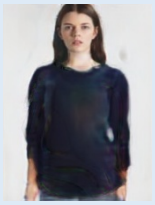
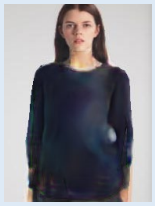
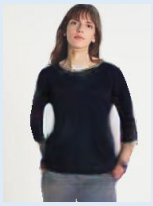
	A:短袖	B:長袖	C:短袖	D:長袖	E:短袖	F:長袖
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
相似度	0.7399091	0.7338381	0.7198432	0.7148836	0.7143156	0.6091847
去除雜訊圖片						
去雜訊相似度	0.7791063	0.7748552	0.7589462	0.7556799	0.7698996	0.6336080

表 5-8 列舉了訓練、測試、實作資料集中，以長袖衣物和短袖衣物為變因，分別為短袖衣物及長袖衣物，所得到的其中數百項結果計算其結果圖片與測試圖片之平均相似度。由表中可見短袖資料所得之結果皆較長袖資料為佳。

(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 5-9 相異拍攝姿勢對模擬試衣結果的影響

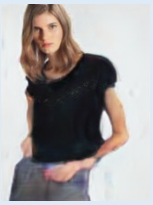
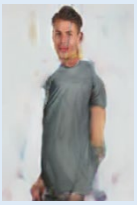
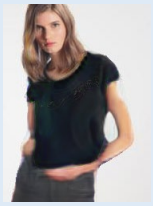
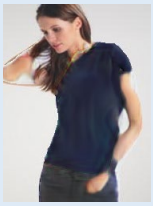
	A:正面	B:側面	C:正面	D:側面	E:正面	F:側面
來源	訓練資料		測試資料		實作圖片	
原始圖片						
原始衣物						
結果圖片						
相似度	0.7352293	0.7299168	0.7126221	0.6934602	0.7143156	0.6845921
去除雜訊圖片						
去雜訊相似度	0.7750565	0.7682734	0.7533926	0.7338854	0.7698996	0.726942

表 5-9 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及角度為輸入變因，所得到的其中數百項結果計算其結果圖片與測試圖片之平均相

似度。由表中可見站姿標準(手貼大腿面向前方)之資料所得結果皆較站姿不標準之資料為佳。

陸、討論

一、人體特徵表示

(一) 人體骨架節點取得結果：

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，當模型無法偵測到人體腿部時，原本屬於腿部的節點(如圖 4-11 中編號 9、10、12、13)將無法正確標示，會出現不規則分布的情況。

(二) 人體遮罩圖生成結果：

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 數據集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，儘管手部有交叉重疊的地方，依舊能有效的辨識。

(三) 原始圖片人物頭部保留結果：

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。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，因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

(一) 扭曲衣物生成結果：

比對表 5-4 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，另一個也會因此產生誤差。

三、 探討相異原始衣物款式對於模擬試衣結果的影響

(一) 探討原始衣物圖案對於模擬試衣結果的影響

由表 5-6 中資料可發現，原始衣物上圖案的單純複雜與否會影響其模擬試衣結果的相似度值，且不論為任一類資料集，其衣服圖案皆為純色之相似度大於複雜圖案者。而我們推論其成因在於 GMM 在生成扭曲衣物時難以對於細節材質進行控制，導致在後續進行試衣模擬時，容易發生過度扭曲或缺漏的現象。又據此結果推論，當輸入之原始衣物的圖案越為單純，對於其模擬試衣結果圖片的效果越佳，且衣服圖案為純色者最佳。

(二) 探討原始衣物色彩對比度對於模擬試衣結果的影響

根據表 5-7 中資料可發現，原始衣物之色彩對比度與模擬試衣結果之相似度有關，且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相似度越大，兩者呈負相關關係。而我們推論其成因在於與背景色彩對比度較低的衣物 GMM 生成扭曲衣物時難以圈選出正確的衣物位置，導致後續進行試衣模擬時，容易發生過度扭曲或缺漏的現象。又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，並以深色、黑色為佳，其模擬試衣結果圖片的效果越好。

(三) 探討相異原始衣物款式對於模擬試衣結果的影響

根據表 5-8 中資料可發現，原始圖片之衣服款式與模擬試衣結果之相似度有關，且不論在模型前端輸入支援使衣物種類為長袖或短袖，在其輸出結果圖片之相似度的比較中，結果皆為原始衣物款式為短袖者所得之結果圖片相似度大於款式為長袖者。而我們推論其成因長袖衣物在 GMM 中需要在短距離內進行大幅度的扭曲，導致後續進行試衣模擬時，容易發生過度扭曲或缺漏的現象。又綜合以上結果可以進行推論，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佳於長袖者。

(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

由表 5-9 中資料中比對可發現，對於各類資料集，其結果相似度皆為標準拍攝姿勢大於不標準者。我們根據結果推論，是因不標準的拍攝姿勢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，拍攝角度為正面首貼大腿者，其輸出結果會佳於其餘姿勢。

四、 去除背景雜訊

從表 5-6、5-7、5-8、5-9 中的數據比較可以看出經過去除背景雜訊後的圖片相似度均提升了約 0.04。但由於結果圖片的遮罩辨識並不完美，導致邊緣處依舊有雜訊殘留。且經過大量資料測試後發現雜訊不只出現於背景部分，於衣服和生成出的手臂等地方也會有明顯的雜訊。

柒、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。我們參考了 CP-VTON，分解出每個步驟的輸出輸入整合成完整程式，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標準者，其試衣結果成效越佳。

捌、參考資料

[1] 本文作者 (民國 110 年 2 月)。基於深度學習之服裝試衣系統。2021 國際科學展覽會。

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

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

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

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

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

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

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

[9] Github Look-Into-Person-v2

(https://github.com/foamliu/Look-Into-Person-v2)

[10] Thin plate splines 薄板樣條插值個人理解

(https://www.twblogs.net/a/5b8de0022b7177188341385c)

[11] Autosport labs Choosing between MAP or TPS

(https://www.twblogs.net/a/5b8de0022b7177188341385c)

[12] Diver 薄板樣條插值

(https://hideoninternet.github.io/2019/11/06/d3c15ac3/)

[13] 鍊聞 CHAINNEWS 圖像分割中的深度學習：U-Net 體系結構

(https://www.chainnews.com/zh-hant/articles/369337679143.htm)

[14] Medium Deeplab v3+

(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)

[15] 人體解析數據集（human parsing）及近期論文

(https://www.twblogs.net/a/5c310044bd9eee35b21ca00c?lang=zh-cn)

[16] Neural Convolutional layers

(https://m-alcu.github.io/blog/2018/01/13/neural-layers/)

[17] Implememnation of various Deep Image Segmentation models in keras

(https://pythonawesome.com/implememnation-of-various-deep-image-segmentation-models-in-keras/)

[18] 姿態估計之 COCO 數據集骨骼關節 keypoint 標註對應

(https://www.stubbornhuang.com/525/)

[19] Bayesian deep convolutional encoder – decoder networks for surrogate modeling and uncertainty quantification

(https://www.sciencedirect.com/science/article/pii/S0021999118302341)

[20] Manual Registration with Thin Plates

(https://profs.etsmtl.ca/hlombaert/thinplates/)

[21] Bayesian deep convolutional encoder – decoder networks for surrogate modeling and uncertainty quantification

(https://www.sciencedirect.com/science/article/pii/S0021999118302341)

【評語】052504

本件作品以深度學習技術開發試衣系統，模擬人體穿著不同衣服的模樣，作品具實用性。研究方法使用的也是研究文獻中公開的模型及資料集進行訓練與實驗，因此也有不錯的研究成果。不過試衣相關研究不少，甚有些購物網站也有推出相似之系統，建議可把此研究之實驗結果與其他研究或相關系統進行比較，以凸顯本研究的創新性或成效。

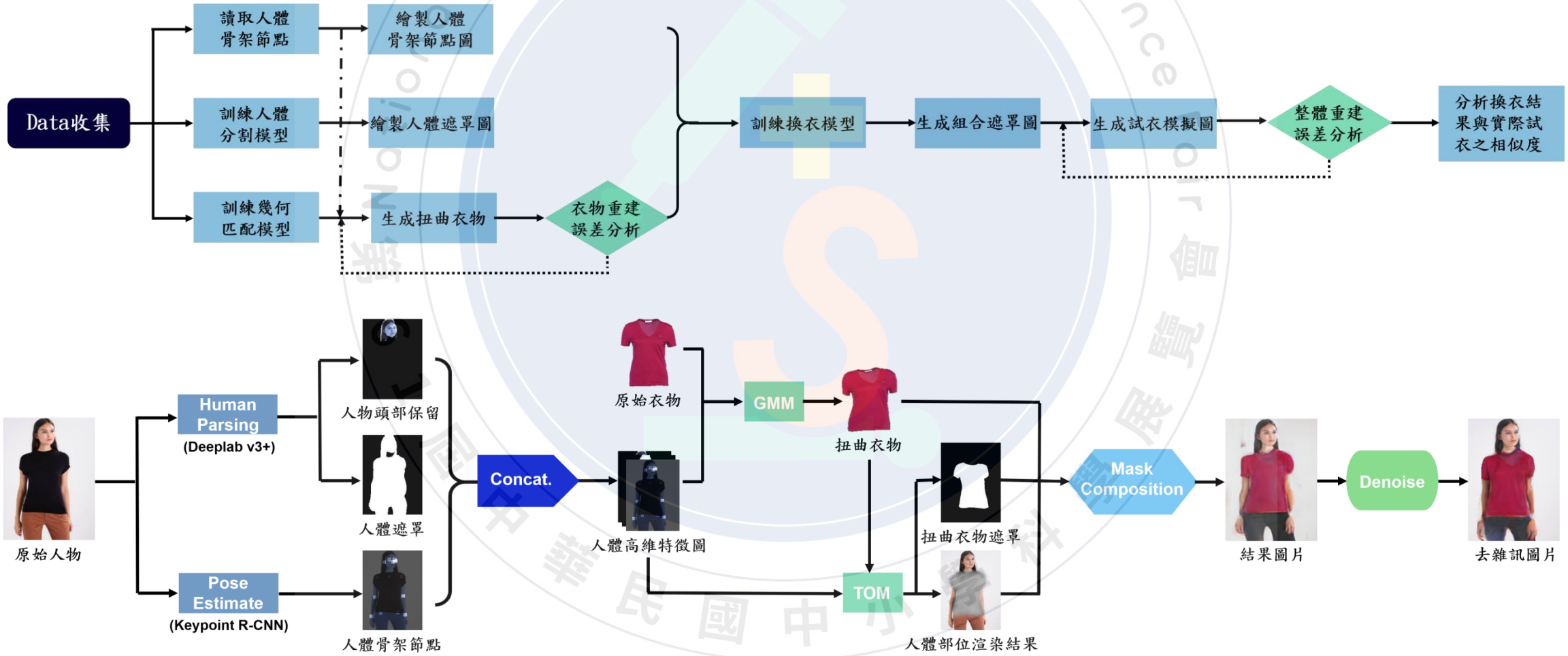
作品簡報

基於深度學習之服裝試衣系統

高中組 電腦與資訊學科 052504

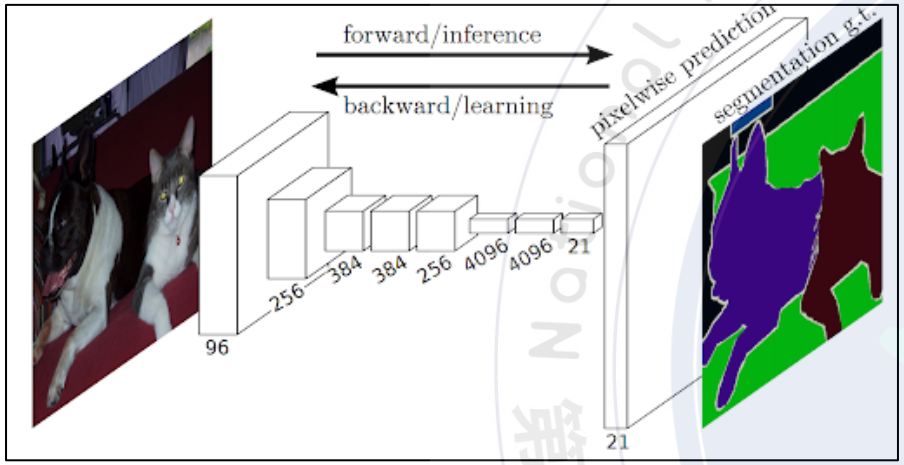
研究動機

- 網路購衣興起→消費者無法直接試穿衣服→效果與想像不符，需退換貨→體驗值降低、資源浪費
- 希望以深度學習模型建構出虛擬試衣系統
- 由各步驟之細節探討在不同情形下使用的優缺點，提出後續優化研究方向



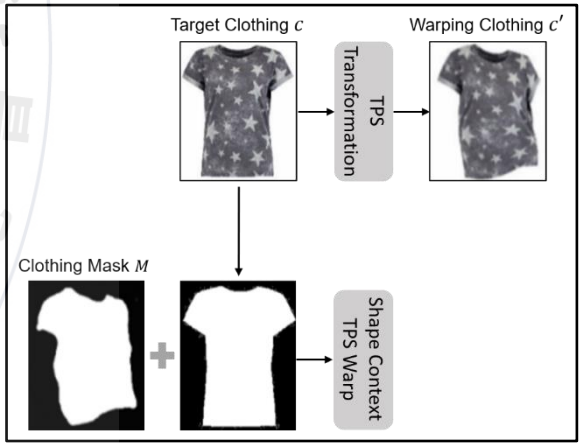
文獻探討

- 3D人體建模剪貼→需巨量運算、精密儀器→無法普及
- 基於圖像生成演算法→以模型扭曲、拼接圖片→VITON為典型模型
- 薄板樣條插值（Thin plate spline，簡稱TPS）、多任務學習

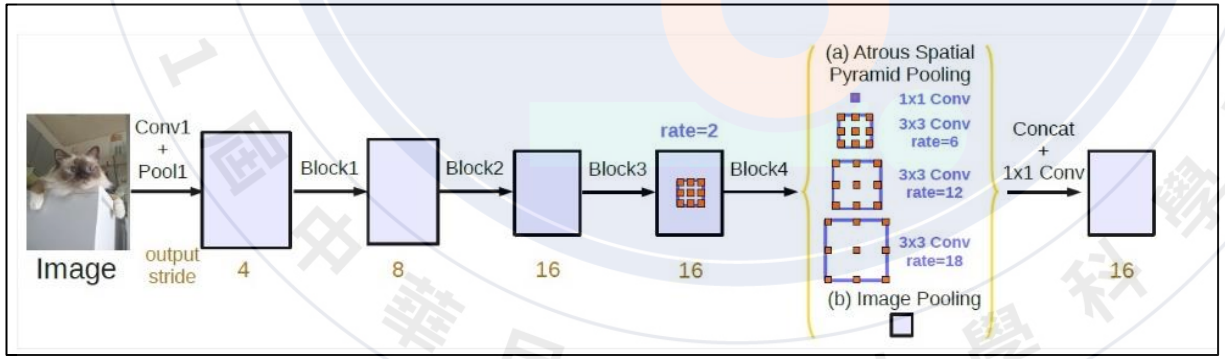


FCN模型架構

影像分割範例



TPS原理圖



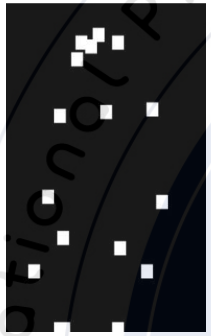
Deeplab v3 架構圖

研究方法

（一）建構高維特徵圖

input

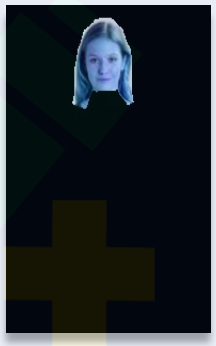
原始人物圖片



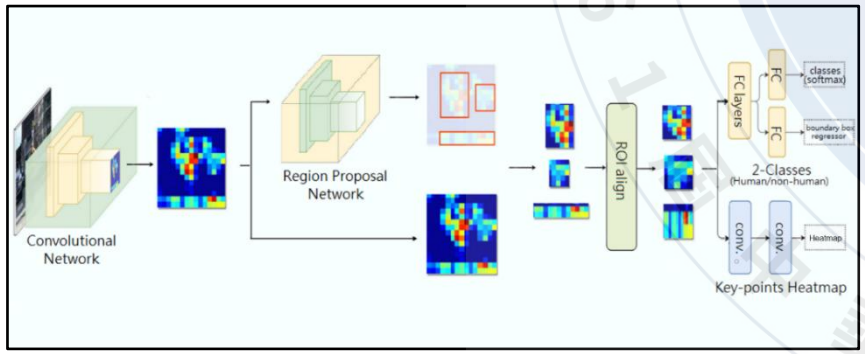
人體骨架節點

output

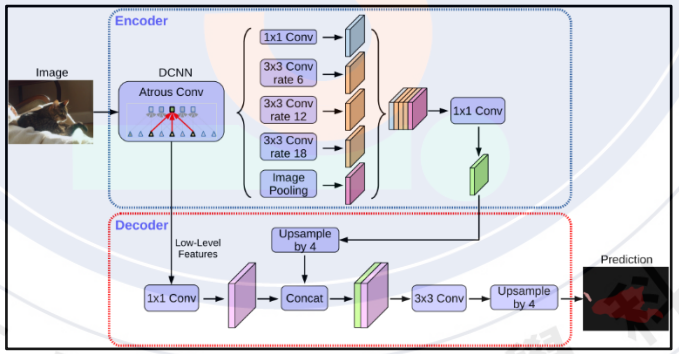
人體遮罩圖



頭部保留圖

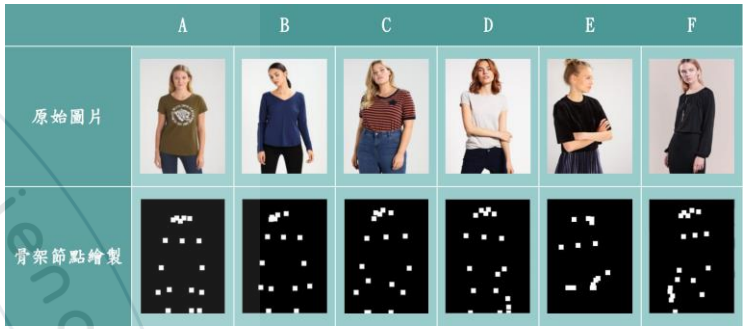


Keypoint RCNN 原理圖



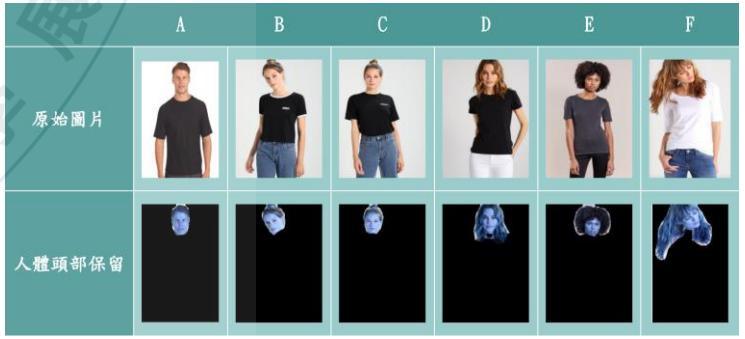
Deeplab v3+ 原理圖

人體骨架節點表格



人體遮罩圖表格

頭部保留圖表格



研究方法

（二）生成扭曲衣物

input

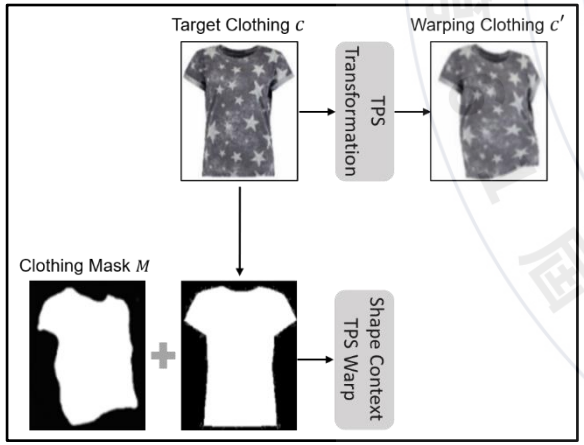
原始衣物

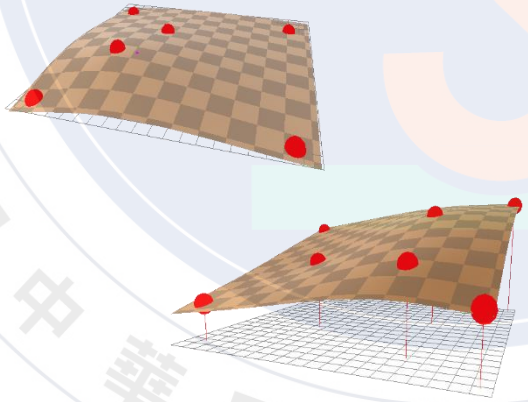
人體高維特徵圖

GMM 衣物扭曲參數

output

扭曲衣物





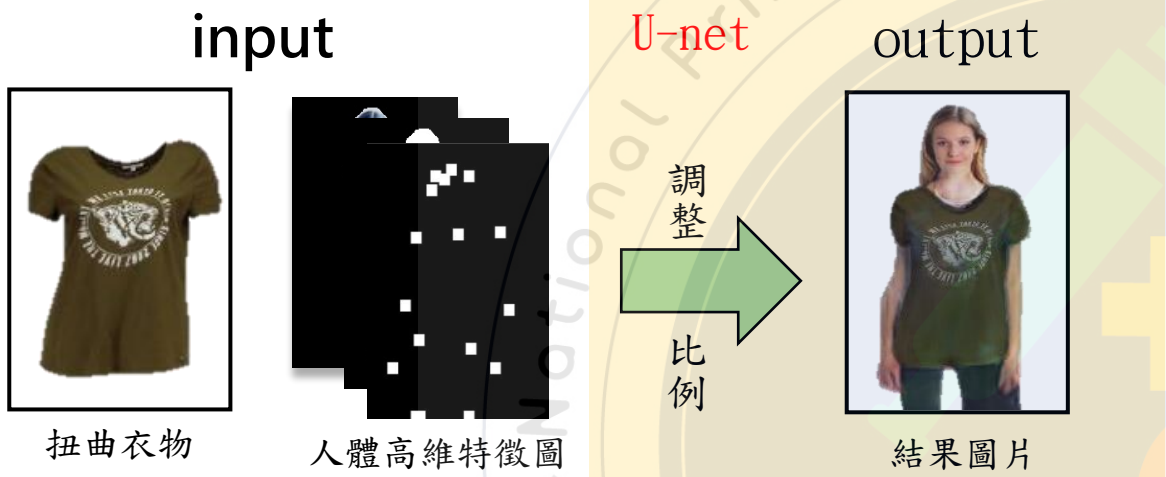
TPS 原理圖

扭曲衣物表格

- 長袖需在短距離中做大幅度的扭曲
  →在貼合人物上較短袖不精準
- 原始衣物扭曲圖與衣物遮罩圖互相影響

研究方法

(三)Try-on 模型



- 整體Try-on模型損失函數

$$\mathcal{L}_{TOM} = \lambda_{L1}\|I_0 - I_t\| + \lambda_{VGG}\mathcal{L}_{VGG}(\hat{I}, I) + \lambda_{mask}\|1 - M\|_1$$

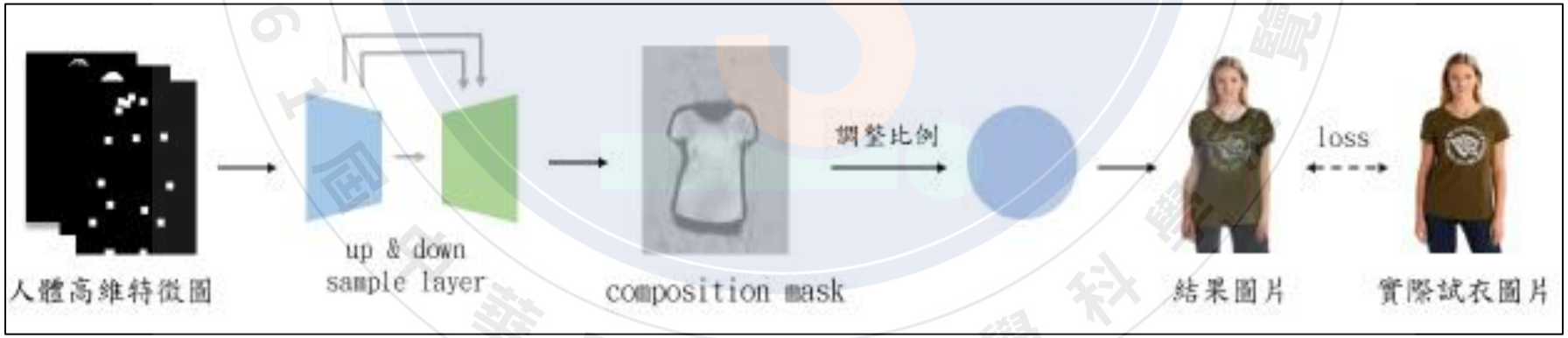
$\lambda_{L1}\|I_0 - I_t\|$：生成圖片和實際圖片誤差　$\lambda_{VGG}\mathcal{L}_{VGG}(\hat{I}, I)$：送入VGG網路取一層之L1誤差　$\lambda_{mask}\|1 - M\|_1$：限制遮罩不變全黑補正項

- 結果圖片相似度計算

$$SSIM(x,y) = [1(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma}$$

$1(x,y)$：亮度比較　$c(x,y)$：對比度比較　$s(x,y)$：結構比較

SSIM具有對稱性，$SSIM(x,y) = SSIM(y,x)$



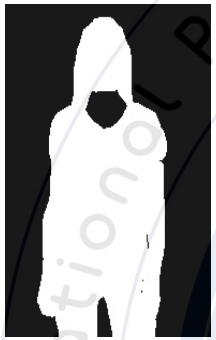
TOM流程圖

研究方法

(四)去除雜訊、模型整合GUI

結果圖片

套入



人體遮罩圖

原始人物圖片

扣除

人體遮罩圖

合併

GUI介面

研究結果

(一)實際試衣結果

- 長袖與短袖進行互換。
- 轉頭/側身/背對等不標準姿勢試衣。
- 淺色衣物容易發生缺漏的現象。
- 複雜圖案容易發生模糊的現象。

研究結果

(二)探討原始衣物圖案對於模擬試衣的影響

	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						
相似度	0.7435926	0.6852192	0.7388035	0.6441410	0.7143156	0.67436737
去除雜訊圖片						
去雜訊相似度	0.7824041	0.7222175	0.7795442	0.6824085	0.7698996	0.7308249

- 衣服圖案皆為純色之相似度大於複雜圖案者
- 推論GMM難以對細節材質進行控制，易過度扭曲或產生缺漏

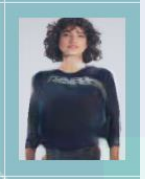
(三)探討原始衣物色彩對比度對於模擬試衣的影響

	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						
相似度	0.7371262	0.6970588	0.7280142	0.6808669	0.7143156	0.6977324
去除雜訊圖片						
去雜訊相似度	0.7762095	0.7381903	0.7674871	0.7223954	0.7698996	0.7257591

- 原始衣物與背景色彩對比度越高，相似度越大
- 推論GMM難以圈選出正確的衣物位置，導致後續進行試衣模擬時，易過度扭曲或產生缺漏

研究結果

(四)探討原始衣物款式對於模擬試衣的影響

	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						
相似度	0.7399091	0.7338381	0.7198432	0.7148836	0.7143156	0.6091847
去除雜訊圖片						
去雜訊相似度	0.7791063	0.7748552	0.7589462	0.7556799	0.7698996	0.6336080

- 原始衣物款式為短袖者之相似度大於長袖者
- 推論因長袖衣物在GMM中需在短距離內進行大幅度的扭曲，易發生過度扭曲或產生缺漏

(五)探討原始圖片人物拍攝姿勢對於模擬試衣的影響

	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						
相似度	0.7352293	0.7299168	0.7126221	0.6934602	0.7143156	0.6845921
去除雜訊圖片						
去雜訊相似度	0.7750565	0.7682734	0.7533926	0.7338854	0.7698996	0.726942

- 標準拍攝姿勢(手貼大腿且面相前方)之相似度大於不標準者
- 推論因資料集中缺乏非正面角度拍攝的影像，衣服不易隨角度轉換、正常覆蓋人體

結論

- 本研究是以深度學習模型，模擬出穿著目標衣物的人物圖片。並於實作測試中找出此模型可能會有的侷限，進行成因分析與討論。
- 本研究參考了CP-VTON，分解出每個步驟的輸出輸入整合成完整程式。其中透過與市售3D建模方式不同的TPS進行扭曲衣物，此方法對於硬體設備的需求較低，較能符合一般大眾的使用情況。
- 根據我們的研究可以對於模型細節進行調整以提高相似度，又綜合以上所得到的分析結果與討論，我們可得知原始圖片人物拍攝姿勢標準及原始衣物款式為短袖、圖案單純、與背景色彩對比度較高者對於試衣結果成效較佳。
- 未來我們希望可以在研究中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。

參考文獻

[1]本文作者(民國110年2月) 基於深度學習之服裝試衣系統 2021國際科學展覽會

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

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