

第二十一屆旺宏科學獎

成果報告書

參賽編號：SA21-305

作品名稱：動畫漫畫化：以漫畫格式呈現動畫

姓名：鄭名揚

關鍵字：分鏡分割、臉部偵測與辨識、漫畫生成

目錄

壹、 研究動機.....	2
貳、 研究目的.....	2
參、 研究過程與方法.....	2
肆、 實驗與研究結果.....	19
伍、 討論及應用.....	27
陸、 結論.....	29
柒、 參考資料及其他.....	30

表目錄

表 1 「動畫漫畫化」相關研究之探討與整理.....	5
表 2 本節的代號一覽表(按使用順序排序).....	6
表 3 使用內容感知法(預設 $T_c=30$)的節錄數據表.....	6
表 4 取用的動畫素材之部分資訊列表.....	19
表 5 判斷情形的混淆矩陣.....	20
表 6 素材 D 不同組合的語者辨識 DER 結果.....	23
表 7 本研究提出的方法與現存方法臉部辨識結果的評估指標.....	24
表 8 兩種機器學習方法的動畫角色辨識結果(表中數值為百分率).....	25
表 9 質化問題的部分結果.....	26

圖目錄

圖 1 熱門動畫電影的漫畫版書照	3
圖 2 動畫的層次結構示意圖.....	4
圖 3 人與電腦在動畫分割上的差異	4
圖 4 動畫漫畫化之流程圖	5
圖 5 「取得畫面」架構圖	5
圖 6 素材的第 91 影格與第 92 影格	7
圖 7 邊界像素判斷方式示意圖.....	8
圖 8 動畫劇照在經邊界偵測處理前後之比較圖	8
圖 9 「提取對話」架構圖	9
圖 10 語者辨識工具生成的波形暨分割例圖	9
圖 11 「辨識角色」架構圖.....	10
圖 12 「臉部偵測」架構圖.....	11
圖 13 左至右依序為 RGB 原圖、轉 HSV 圖、邊界圖與加寬變細後的邊界圖	11
圖 14 由左至右分別為原圖、全部的膚色區域、符合限制的膚色區域	12
圖 15 左上與左下分別為原圖與候選區域；右圖為輪廓分布(紅點)與擬合二次曲線(綠線).....	13
圖 16 臉候選區對應的擬合二次曲線(白線)與半橢圓(粉紅線)	13
圖 17 原圖、經填補後的臉候選區與其對應的擬合二次曲線(白線)與半橢圓(粉紅線).....	14
圖 18 臉部 y 座標差值數列示意圖	14
圖 19 逆序數對圖解.....	14
圖 20 由左至右依序為原圖、頭髮區域與平均髮色.....	15
圖 21 角色辨識過程示意圖.....	16
圖 22 資訊截取間的交互運作與漫畫的生成流程圖.....	17
圖 23 手動匹配對話與角色示意圖	17
圖 24 分鏡表節錄.....	18
圖 25 輸出的漫畫節錄	18
圖 26 由遠至近瞬間的視角切換	20
圖 27 內容感知法之閾值與各評估指標的關係(縱軸起始數值不一定為零)	21
圖 28 適應內容偵測法之閾值與各評估指標的關係(縱軸起始數值不為零)	21
圖 29 以 OpenCV 內件工具偵測臉部的結果舉例	23
圖 30 本研究提出的方法與現存方法各項評估指標的比較(縱軸起始數值不一定為零)	24
圖 31 漫畫成果的各项量化評測分數	26
圖 32 (Fukusato et al., 2016)的漫畫成果	28
圖 33 (Ishikawa et al., 2014)的漫畫成果	28

動畫漫畫化：以漫畫格式呈現動畫

Comicalization Of Animations: Render Animations In Comic Form

摘要

動畫公司會將暢銷的漫畫「動畫化」，那能不能反過來將現存的動畫「漫畫化」呢？本研究研發出一套自動化流程，能讓動畫以漫畫的格式呈現。生成的漫畫成果經問卷評測、與現存漫畫比較，確認它能有效傳遞動畫劇情，幫助讀者了解動畫內容。

實作上，先從輸入的動畫擷取畫面、對話與角色三種資訊，再統整這些資訊，最後生成漫畫。在「取得畫面」步驟中，進行分鏡分割(shot detection)後擷取每個分鏡的關鍵影格(key frame)作為漫畫畫面，其中在多數素材中分鏡分割的 F-measure 超過九成；在「提取對話」步驟中，從字幕提取角色台詞，並使用語者自動分段標記(Speaker diarization)進行語者辨識將台詞匹配至角色，經實驗 Diarization Error Rate 僅 21.7%；在「辨識角色」步驟中，使用自行研發的臉部偵測方法，以動畫角色的臉部附近特徵辨識角色，進而取得角色在畫面中的位置，實驗證實本研究提出的臉部偵測方法明顯優於現存方法，且使用隨機森林分類器(Random Forest Classifier)進行角色辨識的正確率 82.198185%；最後統整上述資訊生成考量角色間對話、具「對話框」的漫畫。

相較於先前的相關研究，本作品透過許多獨創的方法，針對動畫設計出一套漫畫風格的摘要方式。不僅如此，在自動生成的漫畫成果中更具有對話框 — 能夠完整呈現動畫中角色的對話，幫助讀者了解劇情。這是現有文獻生成的漫畫不具有的特色，乃一大創舉。藉由此成果，不但可以讓讀者快速了解動畫內容，甚至能減少製作影片漫畫版本的人力成本。

壹、研究動機

觀看「動畫」，是許多人休閒放鬆的方式之一。近年，因為新冠疫情的緣故，無法外出的限制更使這種「宅娛樂」愈加普遍，一幕幕的歡笑和悲傷，豐富了無數人的日常生活。

然而，對課業壓力大的學生或工作繁忙的上班族來說，騰出時間，悠閒地看動畫，實在是一件非常困難的事。為了節省寶貴時間，不少人會選擇可以自行掌控耗時的「漫畫」：或作為替代的娛樂方式，或用來粗略了解動畫的劇情、畫風等內容，以便判斷這部動畫合不合胃口、是否要投入時間觀看。

雖然現存的動畫大多是由漫畫改編的，但仍有許多值得一看的「原創動畫」缺乏漫畫的型式供人閱讀。因此，本研究的目標就是建立一道流程，自動將動畫轉為漫畫格式呈現，造福許多漫畫迷及因忙碌而無法觀看動畫的同好。

貳、研究目的

本研究欲自動將動畫轉為漫畫格式呈現，使人們能藉由生成的漫畫了解動畫內容。在實作上，需先擷取畫面、對話與角色三大項資訊，再將資訊統整，最後生成漫畫：

- 一、探討動畫的分鏡分割，並在每個分鏡中擷取關鍵影格作為漫畫畫面。
- 二、從字幕提取對話內容後與動畫角色匹配。
- 三、探討動畫角色的辨識方法以得知角色在畫面中的位置。
- 四、統整各項資訊生成考量角色間對話、具「對話框」之漫畫。

參、研究過程與方法

一、文獻探討

(一)漫畫化之定義與目的

為了更有效率地獲取影片資訊，人們設法在「刪除多餘的片段」和「保留代表性的段落」間，取得最佳平衡。從眾多影片摘要的方法中，發展出以漫畫格式的動畫摘要 (comic-style anime summarization)，其漫畫輸出稱為 film comics。在本研究中，定義這樣以漫畫格式摘要影片的動作為「漫畫化」。

(二)現存的漫畫化方法

傳統的漫畫化完全以人工方式擷取劇照與對話，圖 1 為以人工方式將動畫電影「漫畫化」的漫畫產物。這些漫畫版書籍可做為讀者觀影前的預覽或用於觀影後的回味，不同形式的媒材常帶給人不一樣的感受。



圖 1 熱門動畫電影的漫畫版書照

根據(Fukusato et al., 2016)，相較傳統的人工方法，自動的漫畫化研究能有效縮減傳統方法所耗的勞力與時間。以下為現存自動漫畫化的相關研究簡介。

1. Ishikawa 等人的研究(Ishikawa et al., 2014)
 - (1) 關鍵影格偵測：利用畫面中的邊界數量、膚色占比與聲音大小等特徵，在一個分鏡中選取關鍵影格。
 - (2) 推薦區域偵測(recommended area detection)：透過臉部偵測(face detection)來決定最後輸出的畫面區域。

2. Fukusato 等人的研究(Fukusato et al., 2016)
 - (1) 分鏡分割與關鍵影格擷取：透過兩影格間平均畫面的變化和(sum of absolute differences, SAD)來決定分鏡分割處；關鍵影格的擷取則是依照畫面的 HSV 直方圖與物體的移動來判斷。
 - (2) 分配漫畫格：根據關鍵影格的重要性，來決定漫畫格的大小與減少最終顯示的影格數。為避免刪除影格所造成結果的偏離(deflections)，故將輸入的動畫分成：起、承、轉、合四類，並制定不同的影格數下限。

(三)動畫的組成與分割

如同(Haq et al., 2019)提到的「層次結構」(hierarchical structure)，請見圖 2，動畫可以被切割為下列時間長度不一的單元，由時間短到長依序如下。

1. **影格(frame)**：一張張圖片，為動畫的基本單位。
2. **分鏡(shot)**：由具有相似影像特徵的連續影格組成，呈現一個視角或時間、空間上的連續運動，無法傳遞完整的意義。(Fukusato et al., 2016)
3. **幕(scene)**：由多個分鏡組成，能呈現部分劇情。
4. **動畫(animation)**：由數幕組成，能呈現完整劇情。

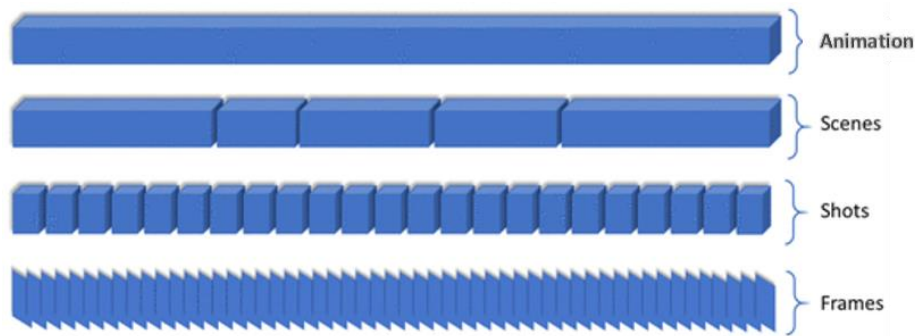


圖 2 動畫的層次結構示意圖

其中，「分鏡」的定義與漫畫中能呈現單一視角或動作的漫畫格不謀而合。因此，在動畫漫畫化的過程中，分鏡分割是不可或缺的一個環節。

而在動畫分割的流程上，則會因視角不同而有先後順序的差異，如下方簡介和圖 3 所示。正是因為這項差異，在使用電腦實作分割時須格外留意分割結果的合理性。

1. 從人的角度分割

由層級高至低進行。「動畫」依劇情或發生地點、出場角色分成數「幕」，每幕再由視角轉換、剪接方式分為數「分鏡」。

2. 電腦實作分割

由層級低至高進行。從「影格」取得數據並集成「分鏡」，再從分鏡提取特徵，進行「標記」(labeling)，歸類至不同的「幕」。(Haq et al., 2019)

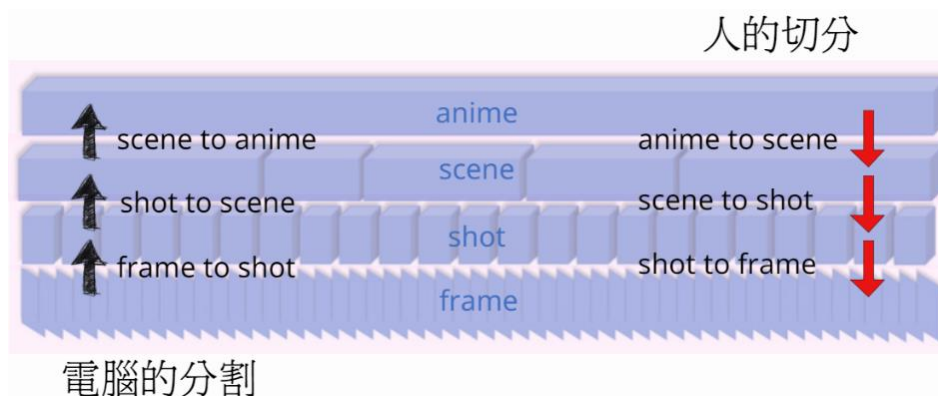


圖 3 人與電腦在動畫分割上的差異

二、研究流程

在了解現存的漫畫化方法後，不難發現這些方法中欠缺對於「對話」與「角色」的考量。本研究認為：未考量角色與對話將造成讀者在閱讀時無法充分了解劇情，而違背動畫漫畫化的初衷。因此，本研究欲針對該不足之處，提出新的構想加以改善。整理如表 1。

表 1 「動畫漫畫化」相關研究之探討與整理

研究	使用方法	探討
(Ishikawa et al., 2014)	<ol style="list-style-type: none"> 1. 選取關鍵影格 2. 決定輸出區域 	<ol style="list-style-type: none"> 1. 生成考慮人物位置的漫畫 2. 未處理對話
(Fukusato et al., 2016)	<ol style="list-style-type: none"> 1. 分鏡分割 2. 計算影格重要性 3. 分配漫畫格大小 	<ol style="list-style-type: none"> 1. 生成考慮排版大小的漫畫 2. 未掌握角色狀況 3. 未處理對話
本研究提出的構想	<ol style="list-style-type: none"> 1. 取得畫面(分割分鏡與擷取關鍵影格) 2. 提取對話(擷取對話並與角色匹配) 3. 辨識角色(識別動畫中的角色) 4. 資訊統整(連結畫面、對話與角色) 	<ol style="list-style-type: none"> 1. 生成考慮角色間對話的漫畫 2. 幫助讀者了解劇情

依照表 1 構想，制訂動畫漫畫化的實作流程大綱如圖 4。

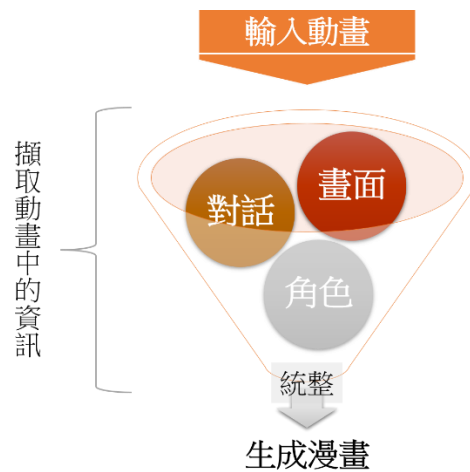


圖 4 動畫漫畫化之流程圖

三、取得畫面

探討動畫的分鏡分割，並在每個分鏡中擷取關鍵影格作為漫畫畫面。

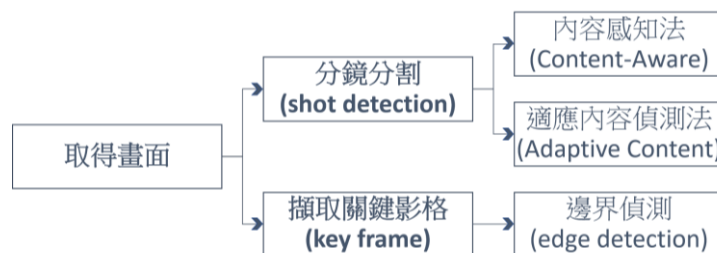


圖 5 「取得畫面」架構圖

(一)分鏡分割

在本研究中，使用分鏡分割工具 PySceneDetect 進行分鏡分割。工具提供了兩種分割演算法，分別為「內容感知法」與「適應內容偵測法」。為了方便理解公式中的代號，以整理本節各代號之意義。

表 2 本節的代號一覽表(按使用順序排序)

代號	英文名	意義
C	content value	兩影格之間的 HSV 平均差異
Tc	threshold of C	C 值的分割閾值
M	minimum scene length	累積影格數的分割下限
Fn	frame number	影格編號
A	rolling average	參考相鄰影格的 C 值之滾動平均值
W	windows width	在計算 A 時使用的前後相鄰影格數
Ta	threshold of A	A 值的分割閾值

1. 內容感知法(Content-Aware)

- (1) 將影格的色彩空間從 RGB 轉為 HSV。
- (2) 比較相鄰兩影格之間的平均差異 C，見(式 1)。
- (3) 若 C 大於設定的閾值(threshold) Tc 且繼上一次分割後累積的影格數超過下限 M，則在此處進行分割。

$$C_{\text{for every frame}} = \frac{\Delta H(\text{hue}) + \Delta L(\text{luminance}) + \Delta S(\text{saturation})}{3} \quad (\text{式 1})$$

舉個例子，表 3 為某動畫素材的部分實作結果，圖 6 為該動畫的兩個影格。表 3 中第三欄 content_val 為四~六欄前面提到的 C 值，而在第 92 影格數據中，可以發現 C 值大於設定的 Tc，因此在此切割出前後兩個分鏡，由此可知 Tc 值會影響分割的結果。

表 3 使用內容感知法(預設 Tc=30)的節錄數據表

Frame Number	Timecode	content_val	delta_hue	delta_lum	delta_sat
90	00:03.8	2.709210431	1.103461372	3.362630208	3.661539714
91	00:03.8	2.748634621	1.129882813	3.369113498	3.746907552
92	00:03.8	39.87332718	15.03917101	66.00073242	38.58007813
93	00:03.9	1.433575666	0.523844401	1.939561632	1.837320964
94	00:03.9	1.186107494	0.195122613	1.850775825	1.512424045



圖 6 素材的第 91 影格與第 92 影格

2. 適應內容偵測法(Adaptive Content)

類似「內容感知法」，但增加了參考相鄰影格變化的滾動平均值 A ，與計算 A 需使用的相鄰影格範圍 W^1 ，此方法有助於減少畫面快速運動的錯誤檢測，計算方式見(式 2)。根據原理，此方法中 T_c 、 T_a 、 W 三個參數同樣會影響分割結果。

$$A_{for\ every\ frame} = C_{Fn_{now}} \div \left[\left(\sum_{i=Fn_{now}-W}^{Fn_{now}+W} C_i \right) - C_{Fn_{now}} \right] \times (2W - 1) \quad (式\ 2)$$

從兩種方法的原理可知，方法中的各項參數將直接影響分割結果。為了驗證各項參數與分割結果之間的關係，並確定電腦自動分割結果的合理性，設計了有關參數調整的實驗，實驗結果請見 [P.19 肆、二、探討動畫的分鏡分割方法之實驗](#)。

(二)擷取關鍵影格

1. 關鍵影格定義

在每個分鏡中，選取一影格作為該分鏡的代表畫面，稱為「關鍵影格」。

2. 取用方式

參考(Fukusato et al., 2016)與(Ishikawa et al., 2014)，認為邊界數與畫面的豐富程度成正相關，故使用畫面中的「邊界數多寡」做為選取關鍵影格的判斷標準。

3. 邊界偵測

使用 OpenCV 內建邊界探測 Canny edge detector 作為演算法工具，同時使用灰階值的梯度(gradient)作為特徵檢測邊緣。先利用高斯濾波器(Gauss filter)來濾出邊緣的梯度值，再根據給定的兩個閾值 T_1 和 T_2 ，判斷是否為邊緣像素(pixel)。關於兩個閾值的運作規則可以用表示成函數 $CED(x)$ 如(式 3)，其中 x 為輸入像素的梯度值；圖 7 為規則之圖解，其中綠線為是邊界像素、紅線為非邊界像素。

¹ 若 $W=2$ ，計算 A 時則取用前兩個和後兩個影格的 C 值。

$$CAD(x) := \begin{cases} true & (x > T2) \\ false & (x < T1) \\ true \text{ (adjacent to edges)} & (T1 \leq x \leq T2) \\ false \text{ (otherwise)} & \end{cases} \quad (式 3)$$

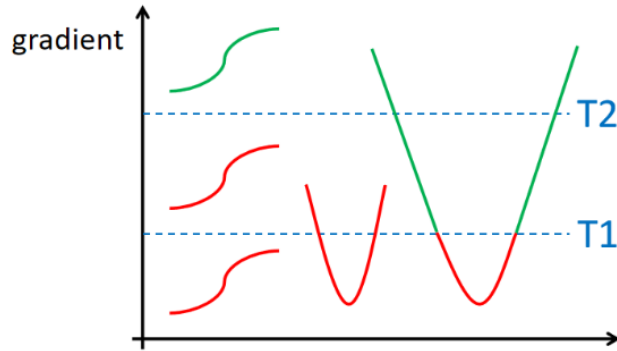


圖 7 邊界像素判斷方式示意圖



圖 8 動畫劇照在經邊界偵測處理前後之比較圖

4. 計算邊界數量

在取得邊界像素的集合後，使用 flood-fill 演算法取得邊界的條(區)數量；於分鏡中數量最高者，即認定為關鍵影格。

使用邊界數決定關鍵影格，經觀察後發現此方法能有效避免選取到模糊不清的影格。推測是因模糊畫面影格間的灰階梯度較低，在 Canny edge detector 中不易被認定為邊界。

四、提取對話

除了擷取動畫中的對話內容，更將其與角色匹配，幫助了解「什麼角色在什麼時候講了什麼話」。

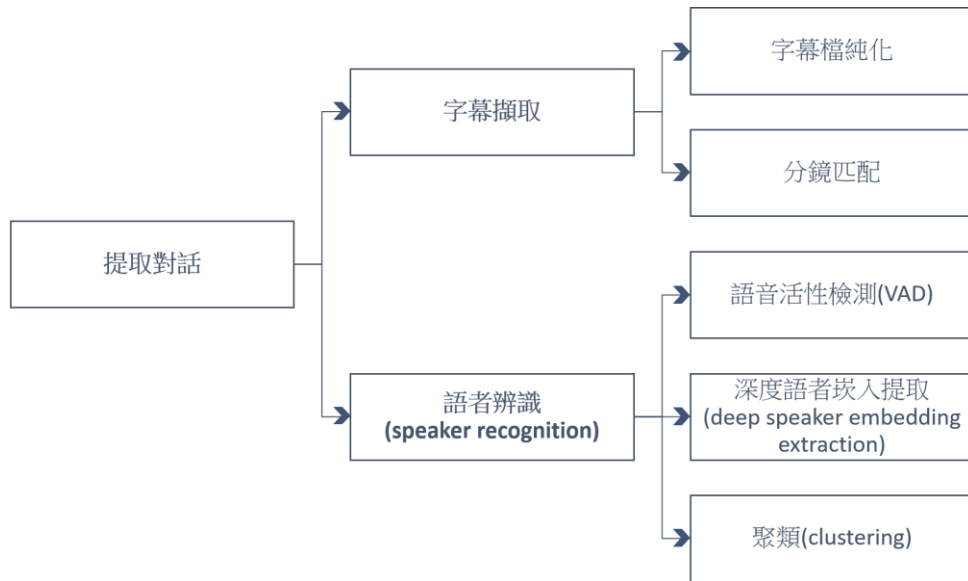


圖 9 「提取對話」架構圖

(一)字幕擷取

1. 字幕檔純化

讀取素材的字幕檔(.ass)，以程式進行純化取得：

- (1) 字幕開始與結束時間
- (2) 字幕內容

2. 分鏡匹配

使用 two pointers 演算法，將對話內容與各分鏡依時間碼進行匹配，每句台詞都能被分配到對應的分鏡。

(二)語者辨識

為得知對話中的各台詞是哪個角色說的，本研究以語者自動分段標記技術辨識語者。在取得動畫聲音的語者資料後，再與先前處理好的字幕和分鏡時間進行匹配。

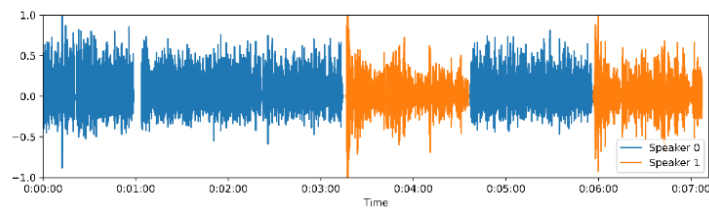


圖 10 語者辨識工具生成的波形暨分割例圖

在使用的工具 Cvqluu/simple_diarizer 中，語者辨識的步驟大致如下：

1. **語音活性檢測(Voice Activity Detection, VAD)**
用於判斷音訊中的語音段落。
2. **深度語者嵌入提取(deep speaker embedding extraction)**
取得用於語者辨識的相關特徵，工具中提供兩種提取模型供選擇：
 - (1) x-vectors
 - (2) Emphasized Channel Attention, Propagation and Aggregation in TDNN (ECAPA)
3. **聚類(clustering)**
將特徵聚類確定各語者的說話段落，工具中亦提供兩種聚類器供選擇：
 - (1) Agglomerative Hierarchical Clustering(AHC)
 - (2) Spectral Clustering(SC)

為了了解前述步驟 2 與步驟 3 中的兩種提取模型與兩種聚類器，何者在動畫對話的角色匹配有最佳效果，設計實驗針對四種組合進行評測。實驗結果請見 [P.22 肆、三、動畫的語者辨識實驗](#)。

五、辨識角色

辨識動畫中出現的主要角色並得知其在畫面中的位置，供生成漫畫時定位對話框(要把對話內容放到角色的旁邊)。相較會更換的服裝或差異較小的軀幹四肢，臉部附近特徵可說是從畫面辨識角色最重要的依據。因此在本研究中，擷取臉部附近特徵進行角色辨識。

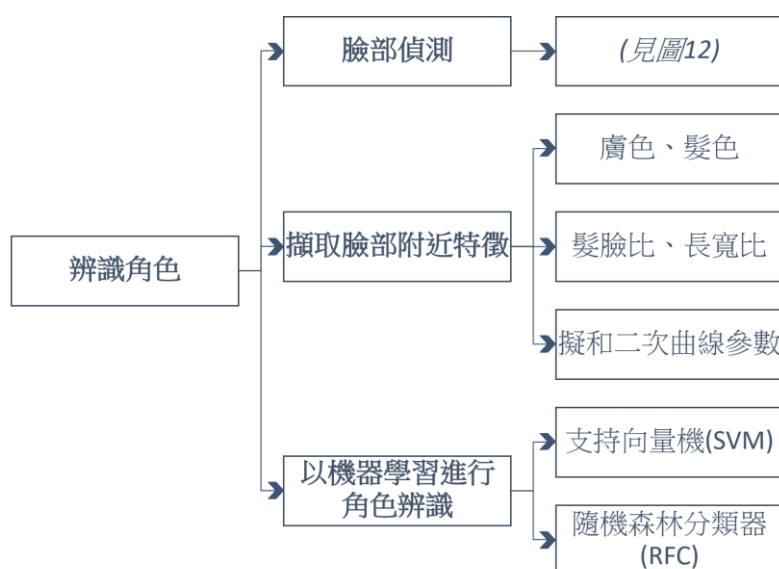


圖 11 「辨識角色」架構圖

(一)脸部偵測

在辨識脸部附近特徵前，必須先知道「臉在畫面中的哪裡？」，因此進行脸部偵測的探討。經探究，發現現存的脸部偵測方法大多只適用於真人的脸部，而在動畫臉部的偵

測上表現不佳，這是因為 2D 動畫的臉部，其五官位置、大小、顏色和陰影等方面，與 3D 真人的臉部有很大差異。(Takayama et al., 2012)

因為無法使用現有的臉部偵測工具來偵測動畫畫面中的臉部，因此必須自行設計一套方法，實現動畫角色的脸部偵測。

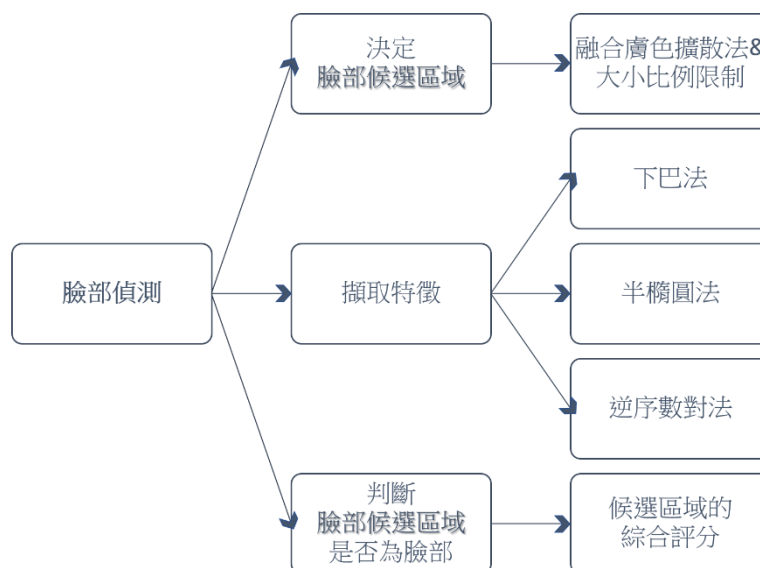


圖 12 「脸部偵測」架構圖

在本研究提出的方法中，除了改良(Takayama et al., 2012)中的方法：「膚色擴散法」&「下巴法」，更發想出新方法：「大小比例限制」、「半橢圓法」與「逆序數對法」。

首先找出畫面中的「脸部候選區域」，再使用三種評分方法：下巴法、半橢圓法與逆序數對法針對每個「脸部候選區域」計算各分數，最後將以綜合評分判斷「『脸部候選區域』是否是臉」。

1. 融合膚色擴散法&大小比例限制

根據(Takayama et al., 2012)，使用轉 HSV 的原圖與邊界圖作為輸入圖片。為了讓邊界能更清楚的切割不同的區域，在使用 flood-fill 演算法擴散前，會先加寬邊界線條再變細(如圖 13)，這是為了確保各個邊界能確實將不同區域分隔。

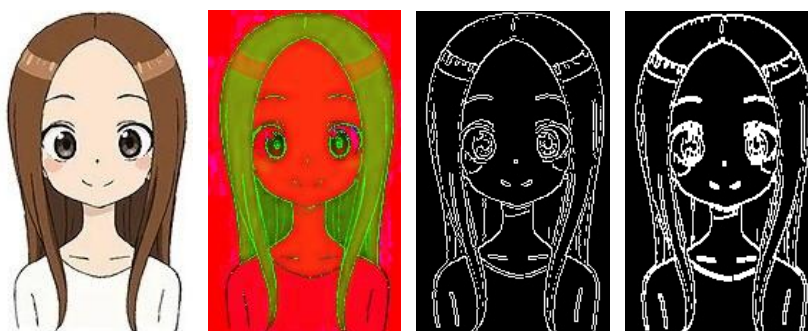


圖 13 左至右依序為 RGB 原圖、轉 HSV 圖、邊界圖與加寬變細後的邊界圖

接著決定「臉部候選區域」：

- (1) 依據膚色與邊界，使用 flood-fill 演算法判斷出數個「膚色區域」，可能包括臉部、脖子、手臂或其他符合膚色數值的區域。
- (2) 為了提升偵測效率，設置「大小比例限制」刪除一些過小、長寬比極端的區域，如圖 14 的中圖裡的零碎色塊。
- (3) 其餘符合限制的膚色區域將作為「臉部候選區域」，如圖 14 的右圖，在圖中共有五塊臉部候選區域。

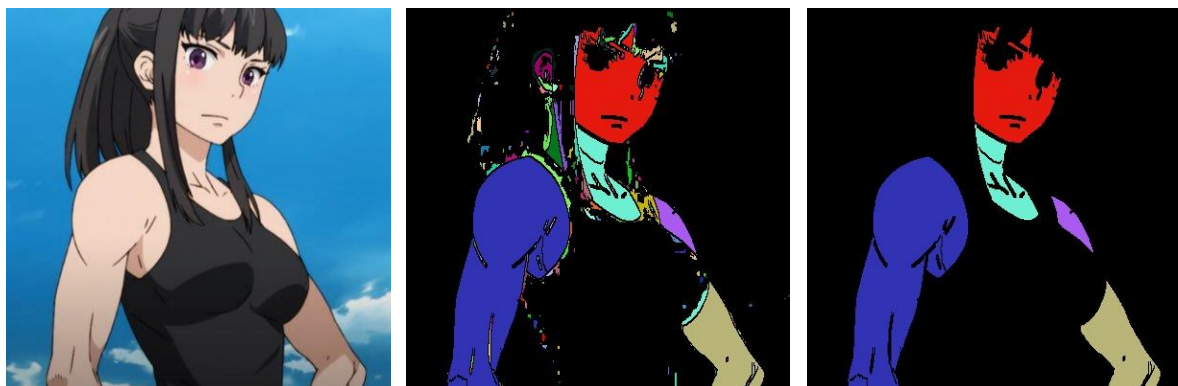


圖 14 由左至右分別為原圖、全部的膚色區域、符合限制的膚色區域

2. 下巴法

經觀察，角色的下巴不易被頭髮或其他部位遮擋，因此通常清晰可見，因此「下巴」可做為臉部偵測的重要特徵。

假設下巴輪廓趨近於一條二次曲線，依下列步驟計算分數 R^2 ：

- (1) 提取各候選區域的下方輪廓²。
- (2) 以最小平方法將其擬合(fit)成二次曲線。
- (3) 假設某輪廓是下巴，該輪廓趨近於擬合的二次曲線。
- (4) 計算 R^2 作為相似程度如(式 4)，其中 n 為輪廓點總數、 \bar{y} 為所有 y_i 的算數平均值、 S 為輪廓對擬合二次曲線的殘差平方和。 S 的計算方式見(式 5)，其中 $f(x)$ 為擬合二次函數。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{S} \quad (\text{式 4})$$

$$S = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (\text{式 5})$$

² 對於每個在圖寬內的整數 X ，找出 Y 最大且落在區域上的點 (X, Y) ，下方輪廓即為合法點之集合。

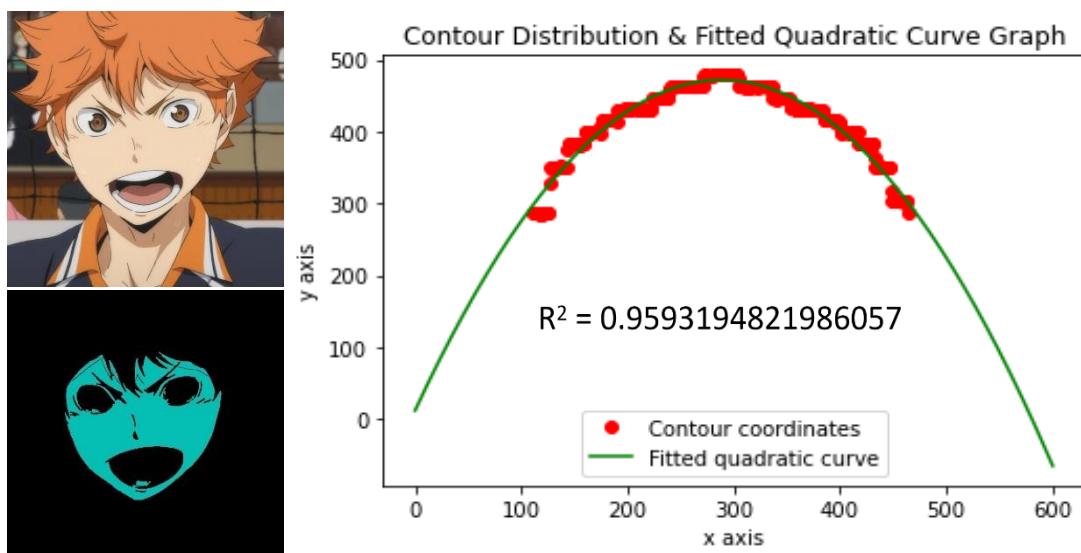


圖 15 左上與左下分別為原圖與候選區域；右圖為輪廓分布(紅點)與擬合二次曲線(綠線)

3. 半橢圓法

在進行前述「下巴法」的實作時，發現有些下方平直、類似矩形的「非臉」候選區域獲得意料之外的高分，細查後發現該類平直輪廓的擬合二次曲線為一條貼緊下方的水平直線，因此得到較高的 R^2 。

為降低上述錯誤狀況帶來的影響，設計了「半橢圓法」。從擬合二次曲線獲得啟發，發現角色的臉部大多趨近於一個半橢圓：由下巴與側臉構成曲線，直線(長軸)則為最上方瀏海遮蓋處。

定義橢圓如下：

- (1) 一頂點：前章節擬合二次曲線的頂點。
- (2) 半短軸：頂點和過候選區域最上方的水平線 $y=y_{\min}$ 之垂直線段。
- (3) 長軸：在 $y=y_{\min}$ 上被擬合二次曲線所截之線段。

假設候選區占半橢圓面積越高越可能是個「臉」，計算候選區域在半橢圓中的占比率 F ，作為「綜合評分」的分數之一。結果請參考圖 16，圖中數值即為 F 。



圖 16 臉候選區對應的擬合二次曲線(白線)與半橢圓(粉紅線)

測試後發現，大多數非臉的矩形候選區域在此方法獲得較低的 F 值，能有效排除前述在「下巴法」獲得特別高分數的非臉候選區域。

另外，為了提高鑑別度方便比較，在計算占比率 F 前嘗試先利用 flood-fill 演算法填補候選區域中的空缺。實測後發現此改動能有效提升臉部的 F 值。

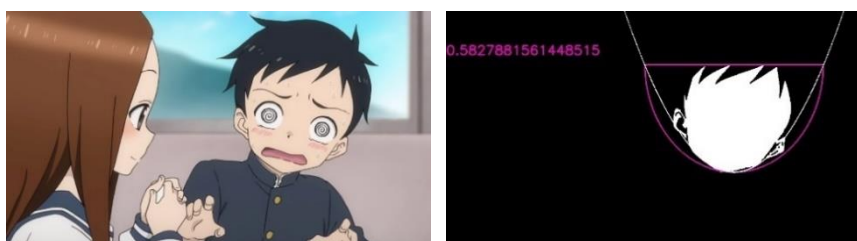


圖 17 原圖、經填補後的臉候選區與其對應的擬合二次曲線(白線)與半橢圓(粉紅線)

4. 逆序數對法

經觀察發現，臉部的縱切長度(y 座標差值)基本上呈現中間長左右短的趨勢，能近似成以縱切長度最長者為中央，向左右延伸的非嚴格遞減數列。例如圖 18，在臉部區域簡單作數條代表 y 差值的黃線，對齊整理後如右圖，可以觀察到中間長左右短的趨勢。

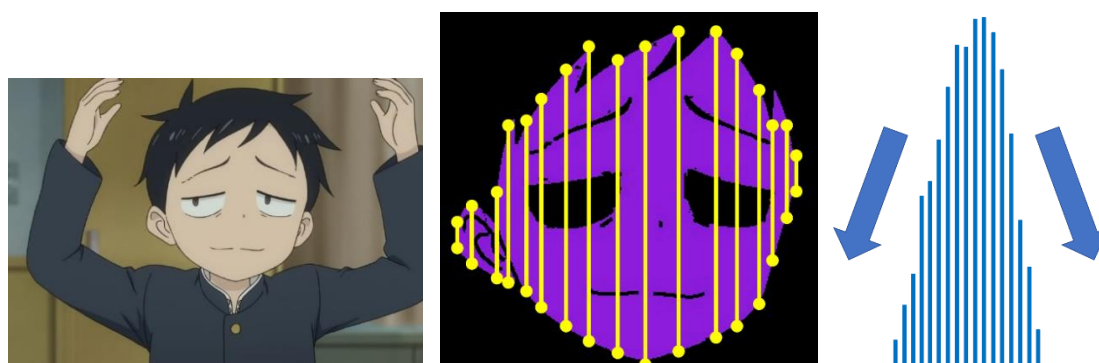


圖 18 臉部 y 座標差值數列示意圖

這樣的現象讓人聯想到程式設計的經典題目「逆序數對」。「逆序數對」是指在一個有 n 個數的數陣列 $a[n]$ 中，存在整數 i 與 j 且 $0 \leq i < j < n$ ，使 $a[j] < a[i]$ ，則稱 $a[i]$ 和 $a[j]$ 為一組逆序數對。

以圖 19 為例，以各長條的高度為數列，可以發現編號(3,4)、(3,5)、(3,7)、(4,5)、(6,7)之長條組的高度異於左低右高的「順序」，故這五個數對組為逆序數對。

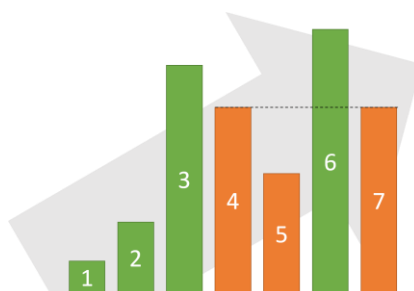


圖 19 逆序數對圖解

依照下列步驟計算數列中的非逆序數對占比 I 作為判斷是否是臉部的標準：

- (1) 對於左右兩數列，定義標準數列為一非嚴格遞增的數列。
- (2) 對於特定點的 y 座標差值，以候選區域最左、最右的 y 座標差值為首項，區域中最大 y 座標差值為末項，分別對兩數列以複雜度為 $O(n \log n)$ 的 **divide and conquer** 演算法計算逆序數對數量。
- (3) 計算非逆序數對在所有數對中的占比 I ，計算方式如(式 6)，其中 n 為候選區域的寬度， k 為逆序數對數量。

$$I = 1 - \frac{k}{C_2^n} \quad (\text{式 6})$$

5. 候選區域的綜合評分

將「下巴法」取得的 R^2 、「半橢圓法」取得的 F 與「逆序數對法」取得的 I 加總作為候選區域的綜合評分，每個「臉部候選區域」都會有一個對應的綜合評分。定義一候選區域符合限制條件若且唯若該區域是臉部。

為了驗證提出的新方法在動畫角色臉部偵測上的優勢，進行實驗了解辨識效果並與現存真人臉部偵測方法比較，實驗結果見 [P.23 肆、四、探討動畫人物的臉部偵測實驗](#)。

(二)擷取臉部附近特徵

對於每個偵測出的臉部區域，提取「膚色」、「髮色」、「髮臉比」、「長寬比」和「擬和二次曲線參數」作為角色辨識的特徵。

1. 定義頭髮區域

- (1) 取用臉部上方輪廓之上的數個像素作為頭髮樣本。
- (2) 將樣本以 **k-means clustering** 分群分類，其中 k 值取用係依據不同 k 值之組內平方和(Within Cluster Sum of Squares, WCSS)的差值。
- (3) 求取最大分群的平均色彩值，此色彩值即為頭髮的「樣本顏色」。
- (4) 導入邊界偵測圖，使用 **flood-fill** 演算法擴散色彩與樣本顏色相近的像素。
- (5) 擴散後的區域即是頭髮區域，請參考圖 20。



圖 20 由左至右依序為原圖、頭髮區域與平均髮色

2. 膚色、髮色

平均所有臉部區域、頭髮區域的像素色彩做為臉部特徵。為了增加特徵的種類，色彩空間使用了 RGB、HSV 與 CIELab。

3. 髮臉比³、長寬比

分別計算頭髮、臉的區域面積與臉的長寬，相除得髮臉比、長寬比。

值得一提的是，使用「髮臉比」而不使用「臉髮比」做為特徵的原因：不同角色頭髮面積的差異大於臉部，因此將頭髮面積當作分子、臉面積當作分母。

4. 擬合二次曲線參數

臉型是角色的重要特徵之一，而臉部下方邊緣的擬合曲線參數能代表臉型。使用同臉部偵測中下巴法的計算方式，取得三個參數。

(三)以機器學習進行角色辨識

在取得所需特徵後，本研究使用兩種常見的機器學習模組：支持向量機(support vector machine, SVM)與隨機森林分類器(Random Forest Classifier, RFC)進行角色辨識。依循常規的機器學習流程進行如下：

1. 訓練(training)

將已標記的臉部作為訓練資料(training data)導入機器學習模組，進行訓練以建立角色資料庫。

2. 測試(testing)

接著將未知的臉部特徵作為測試資料(testing data)導入機器學習模組，進而預測所屬的分群類別，即辨識之角色結果。

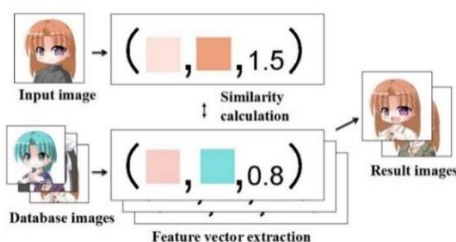


圖 21 角色辨識過程示意圖

為了比較兩種機器學習模型在動畫角色辨識上的效果，進行了相關實驗。為了使實驗結果更具說服力，乃採用交叉驗證(cross-validation)取用訓練與測試資料，之後再將結果平均作為最終結果。比較結果請見 [P.25 肆、五、探討使用不同機器學習方法的動畫角色辨識結果](#)。

六、資訊統整&成果輸出

(一)資訊截取間的交互關係

在先前的章節，說明了如何從輸入的動畫中提取三種資訊：畫面、對話與角色。事實上，這些資訊的擷取並非獨立進行的，而是需要倚靠彼此的資訊，才能得到完整的細節。各資訊構成的資料庫間存在數條交互關係，詳如圖 22。

³ (Takayama et al., 2012)中是提取「髮量」(hair quantity)作為特徵，但本研究認為髮量可能因畫面遠近而有差異，因此使用頭髮與臉部面積的比例做為特徵。

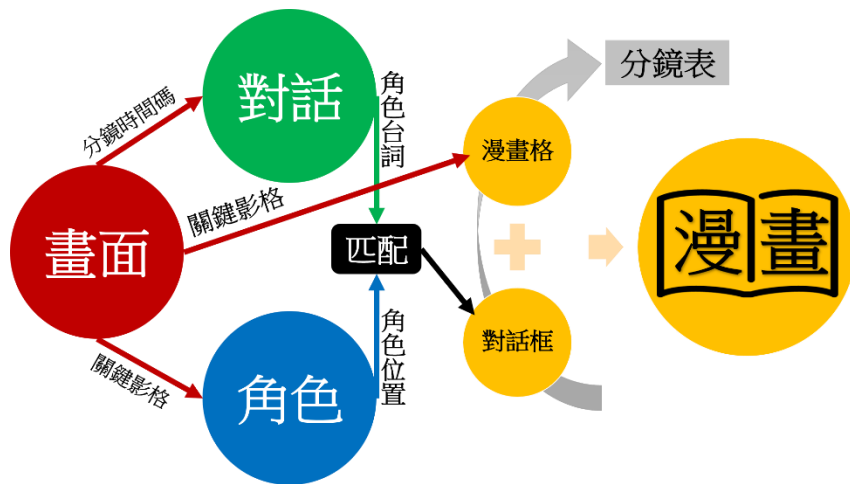


圖 22 資訊截取間的交互運作與漫畫的生成流程圖

在「取得畫面」時，取得了各個分鏡與每個分鏡中的關鍵影格，分鏡的時間碼會用於將對話匹配至分鏡，關鍵影格則會作為最終的漫畫中的畫面以及角色辨識的對象。因為畫面資料庫中的資訊會被用於其他資訊擷取步驟，因此「取得畫面」必須優先進行。

(二)生成對話框的前置作業

先前提過，在「取得畫面」取得的關鍵影格將直接作為每個漫畫格的畫面。相比起來，生成對話框困難許多：在動畫中，缺乏對話與角色的連結資訊，因此必須手動匹配「提取對話」的角色台詞與「辨識角色」的角色位置。這也成為整個漫畫化過程唯一需要人工參與之處。

以圖 23 說明，在先前的步驟中，已經自動取得了數堆的對話與角色資料，但不知道哪一堆台詞(ABC...)與哪一堆位置(123...)是對應到同一個動畫角色。因此需要依賴手動匹配確定台詞與位置的歸屬。

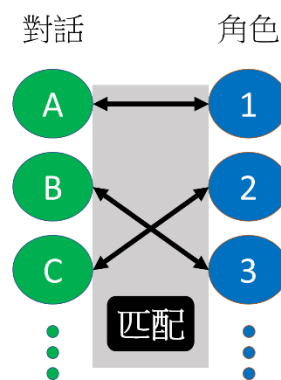


圖 23 手動匹配對話與角色示意圖

(三)以分鏡表(storyboard)整理資訊

根據國家電影及視聽文化中心的介紹，「分鏡表」是在影片實際拍攝或動畫繪製之前，以圖格的方式，來說明預備拍攝影像畫面的構圖，此具有統合信息、規劃製作與節省時間等重要功能，分鏡表是動畫製作前期不可或缺的重要工具之一。

為了清楚整理取得的各分鏡資料，將資料以分鏡表整理並使用 PHP 輸出：

1. 畫面(左側)：分鏡編號、關鍵影格
2. 細節資訊(右側)：持續時間、出現角色、角色對話內容

 <p>36</p>	<p>時間：00:00:51.009 - 00:00:55.764</p> <p>出現角色：女主角</p> <p>對話：女主角：敦君敦君 今天天氣真好呢</p>
 <p>37</p>	<p>時間：00:00:55.764 - 00:00:59.476</p> <p>出現角色：男主角</p> <p>對話：男主角：你好噁心</p>
 <p>38</p>	<p>時間：00:00:59.476 - 00:00:59.476</p> <p>出現角色：女主角 男主角</p>

圖 24 分鏡表節錄

(四)生成漫畫

生成「動畫的漫畫版本」為本研究的總目標。以四格漫畫的方式輸出每個動畫分鏡的關鍵影格，並將與角色匹配完成的對話內容置於角色的臉部下方作為對話框。最後以 PHP 輸出。

圖 25 為部分的漫畫輸出。在圖中，以粉藍色呈現男主角的台詞、粉紅色呈現女主角的台詞，對話框位置皆設於該角色臉部下方。值得一提的是，左上方的分鏡因為畫面中人物臉部太小，導致在臉部偵測的步驟未成功偵測，自然也無法進行角色辨識。對於此類分鏡，統一將台詞至於畫面左上方。



圖 25 輸出的漫畫節錄

肆、實驗與研究結果

一、選擇實驗素材 — 動畫

挑選動畫作為本研究的實驗素材。

(一) 動畫限制

取用具有字幕的動畫，其中字幕必須儲存於分離的字幕檔，而非安插在影片中。有這項限制的原因有二：

1. 方便取得台詞

在「擷取對話」的步驟，將直接從字幕擷取對話內容。

2. 避免影響邊界數的判讀

使用嵌入在影片內的字幕，會導致在「取得畫面」計算邊界數時，額外計算字幕字體的邊界。此乃不樂見之狀況，故取用具有分離的字幕檔的動畫。

(二) 使用聲明

在本研究中，所有使用的素材影片皆是從網路上的開放平台下載，僅作本次研究使用，絕無他用，並於研究結束時全數刪除。所有使用行為根據中華民國《著作權法》合理使用，若有未盡之處敬請各方指教。

(三) 取用的動畫素材

表 4 取用的動畫素材之部分資訊列表

編碼	動畫名稱	製作公司	使用集數
A	鬼滅之刃 第一季《竈門炭治郎 立志篇》	ufotable	10
B	擅長捉弄人的高木同學 S1	新銳動畫	5
C	排球少年！！ 第一季	Production I.G	5
D	敦君與女朋友	夢太公司	5

二、探討動畫的分鏡分割方法之實驗

為了驗證分鏡分割方法中各項參數與分割結果之間的關係，並確定電腦自動分割的合理性，所以設計了有關參數調整的實驗。

(一) 定義效能評估方式

根據(Haq et al., 2019)，評估變化準確性(change accuracy)的常用指標為精密度 P(Precision)、召回度 R(Recall)和 F 度量 F(F-measure/F1-score)。它們的計算方式請參照表 5 的混淆矩陣(confusion matrix)與(式 7)的計算，其中 FP 即為一型錯誤(α 錯誤)、FN 即為二型錯誤(β 錯誤)。

表 5 判斷情形的混淆矩陣

總數據		系統判斷	
		預測符合	預測不符合
人工判斷	實際符合	正確正例 (true positive, TP)	錯誤負例 (false negative, FN)
	實際不符合	錯誤正例 (false positive, FP)	正確負例 (true negative, TN)

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \quad F = \frac{2 \times P \times R}{P + R} \quad (\text{式 7})$$

(二) 制定分割基準

在本實驗中，分割方法為「系統判斷」、分割出的分鏡結果為「預測符合」之數據，各評估指標的意義如下：

1. **精密度**：為 TP 除以工具判斷的分鏡數，用來檢視分割的錯誤程度。
2. **召回度**：為 TP 除以人工分割的分鏡數，用來檢視分割的缺漏程度。
3. **F 度量**：為精密度和召回度的調和平均數(harmonic mean)，用於綜合評估。

根據(Haq et al., 2019)，研究在準備基準上，常以研究者的手動分割作為基準，即基本事實(ground truth)，分鏡分割的判斷標準如下。

1. 畫面在瞬間呈現時間推移
2. 畫面在瞬間出現視角切換(如圖 26)

當動畫之「影像」符合上述二條件之一時，則進行人工分割(手動標記時間碼)。



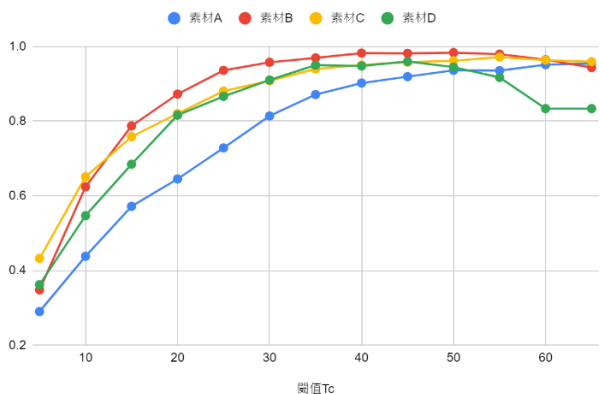
圖 26 由遠至近瞬間的視角切換

(三) 閾值參數實驗結果

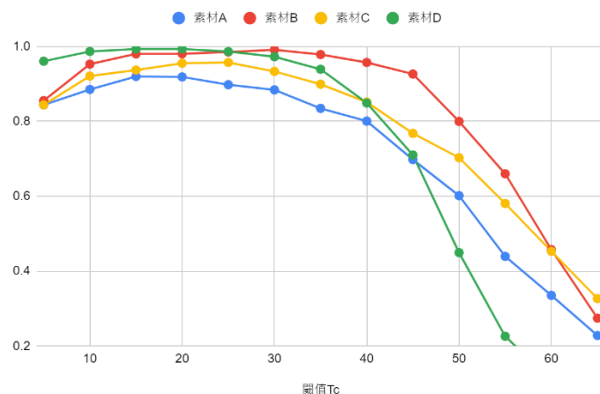
針對「內容感知法」與「適應內容偵測法」，探討內容感知法的 C 閾值 Tc 及適應內容偵測法的 A 閾值 Ta 與分鏡分割的關係。

1. **內容感知法**：Tc 數值={5,10,15,20,25,30,35,40,45,50,55,60,65}
2. **適應內容偵測法**：Ta 數值={2,3,4,5,6,7,8,9,10}

[內容感知法]各素材精密度與閾值Tc的關係折線圖



[內容感知法]各素材召回度與閾值Tc的關係折線圖



[內容感知法]各素材F度量與閾值Tc的關係折線圖

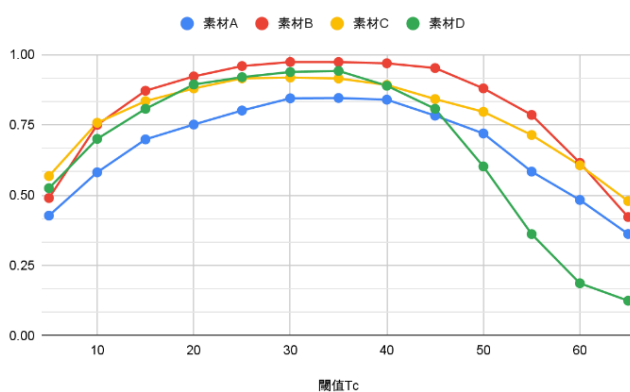
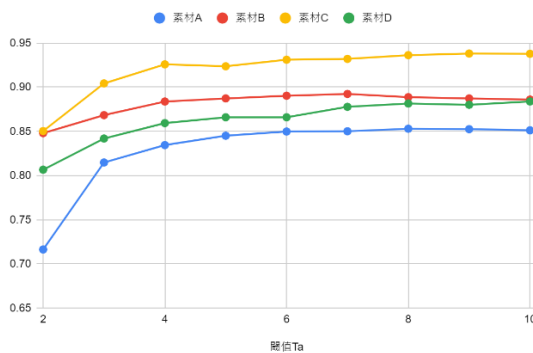
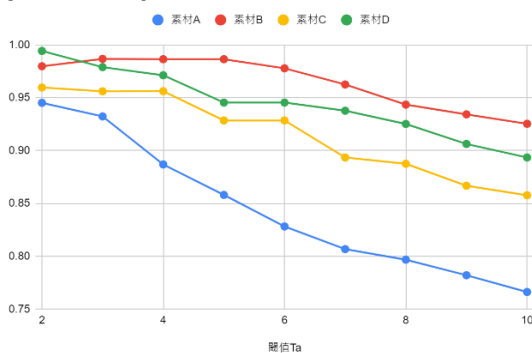


圖 27 內容感知法之閾值與各評估指標的關係(縱軸起始數值不一定為零)

[適應內容偵測法]各素材精密度與閾值Ta的關係折線圖



[適應內容偵測法]各素材召回度與閾值Ta的關係折線圖



[適應內容偵測法]各素材F度量與閾值Ta的關係折線圖

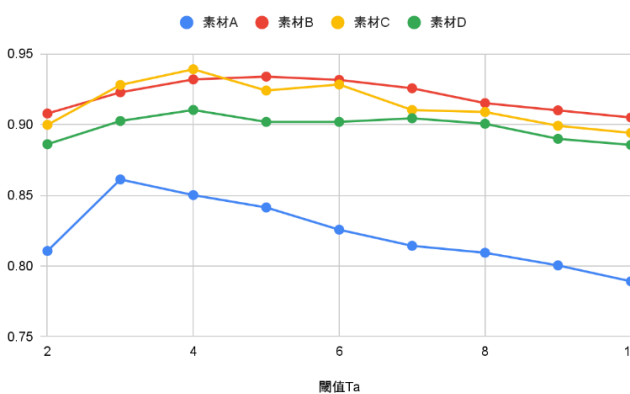


圖 28 適應內容偵測法之閾值與各評估指標的關係(縱軸起始數值不為零)

除了兩種方法的主要閾值實驗外，也進行了適應內容偵測法的參數 W 與 C 閾值 Tc 之實驗，但未發現明顯關聯，請參閱 [P.31 附錄壹](#)。

(四)實驗討論

1. 內容感知法

- (1) 隨著閾值 Tc 增加，精密度大致呈遞增趨勢，相反的召回度成遞減趨勢。推測這是因為閾值愈高愈難達到分割條件，造成分割出的分鏡數減少。
- (2) F 度量高峰出現在 Tc 於區間 30~35 時，與工具預設的 Tc=30 差異不大。

2. 適應內容偵測法

- (1) 精密度與召回度與閾值的關係類似「內容感知法」，但差異較小：精密度與閾值呈正相關，召回度則呈負相關。
- (2) 在 F 度量方面，除了素材 A 在 Ta=3 時有明顯高峰，其他素材不論 Ta 大小皆獲得九成上下的良好分割結果。
- (3) 素材 A 的分割結果皆略低於其他素材，推測是因為在素材 A 中較多畫面切換快速的動作、打鬥畫面，進而降低工具自動分割的正確率。

綜合兩個分割方法的實驗結果，可以發現只要使用適當的參數，都能有八成以上的分鏡分割表現。對於本研究而言，分鏡分割僅是眾多步驟之一，因此認為這樣的分割效果已經十分足夠。

另外，異於其他素材，素材 A 在適應內容偵測法的表現優於內容感知法，推測是因為在處理快速運動畫面上，適應內容偵測法有較好的分割能力：此演算法的精髓為比較影格與前後影格的關係。考量上述結果，在後續的實驗中，分鏡分割步驟統一使用適應內容偵測法(Ta=3, Tc=25, W=2, M=15)進行分鏡分割。

三、動畫的語者辨識實驗

為了了解語者辨識中的的兩種提取模型與兩種聚類器，何者在動畫對話的角色匹配有最佳效果，設計實驗針對四種組合進行評測。

(一)評量方式

語者自動分段標記常用 *Diarization Error Rate (DER)* 評分，計算方式如(式 8)。

$$DER = \frac{False\ Alarm + Miss + Confusion}{Reference\ Length} \quad (式\ 8)$$

在(式 8)中的各個術語定義如下，其單位皆為秒：

1. **Reference Length**：人工制定的基準的總長度，為基本事實。
2. **False Alarm**：工具認定是說話片段、但事實上不是的片段長度。
3. **Miss**：工具遺漏的說話段落長度。
4. **Confusion**：系統錯誤辨識語者的片段長度。

(二)實驗結果

表 6 素材 D 不同組合的語者辨識 DER 結果

		extraction model	
		x-vector	ECAPA
cluster	AHC	80%	27.9%
	SC	22.3%	21.7%

從表 6 可以得知 ECAPA 搭配 SC 有最好的辨識效果，DER 僅 21.7%。因此在後續的流程中，皆採用該組合進行語者辨識。

四、探討動畫人物的臉部偵測實驗

評測提出的臉部偵測方法之偵測效果，並與現存方法進行比較，以便驗證提出的新方法在動畫角色臉部偵測上的優勢。

(一)評量方式

使用與本章二、(一)相同的評估指標，以輸出的臉部結果為「預測正確」之數據、人工判定的臉部作為參考基準，指標的意義如下：

1. **精密度**能評估偵測臉部的錯誤比例。
2. **召回度**用來形容系統在畫面中找出正確臉部的能力。
3. **F 度量**用於綜合評估。

(二)與現存方法比較

本研究使用相同的實驗素材，採用針對真人臉部的 OpenCV 內建的臉部偵測器「cascade 分類器」當作比較。分類器使用 OpenCV 中已訓練的臉部偵測模型 haarcascade_frontalface_default.xml。



圖 29 以 OpenCV 內件工具偵測臉部的結果舉例

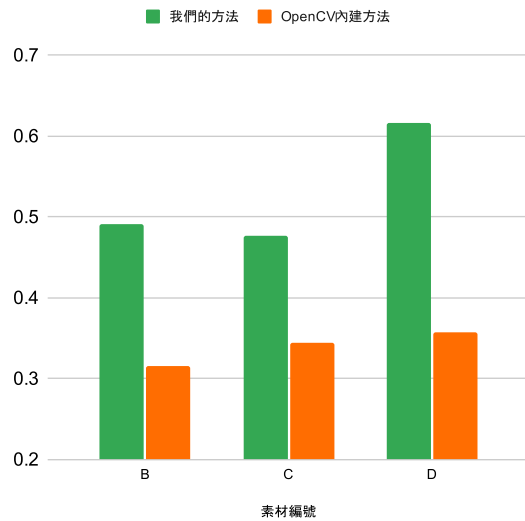
(三)實驗結果

從表 7 和圖 30 可見，在實驗結果中，對於所有實驗素材，本研究提出的方法皆優於現存方法，這說明本研究提出的方法在動畫角色臉部偵測上的優勢。

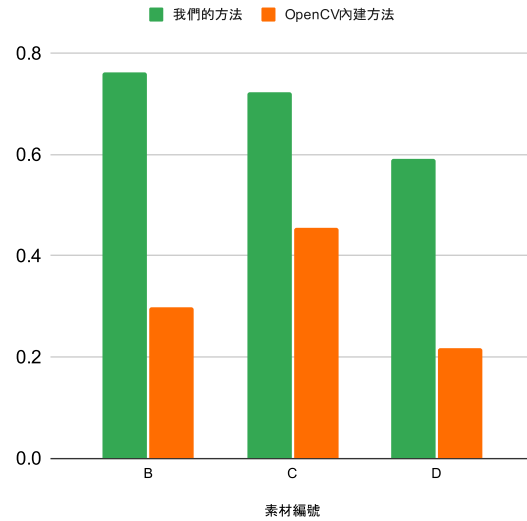
表 7 本研究提出的方法與現存方法臉部辨識結果的評估指標

素材編碼	B		C		D	
本研究的方法	P	0.490085434	P	0.477001908	P	0.615942029
	R	0.763274263	R	0.722491902	R	0.590301003
	F	0.596907016	F	0.574625751	F	0.602848991
OpenCV 內建方法	P	0.315444015	P	0.344109313	P	0.357142857
	R	0.29709608	R	0.454049519	R	0.217391304
	F	0.305995252	F	0.391507709	F	0.27027027

本研究提出的方法與現存方法的精確度比較



本研究提出的方法與現存方法的召回度比較



本研究提出的方法與現存方法的F度量比較

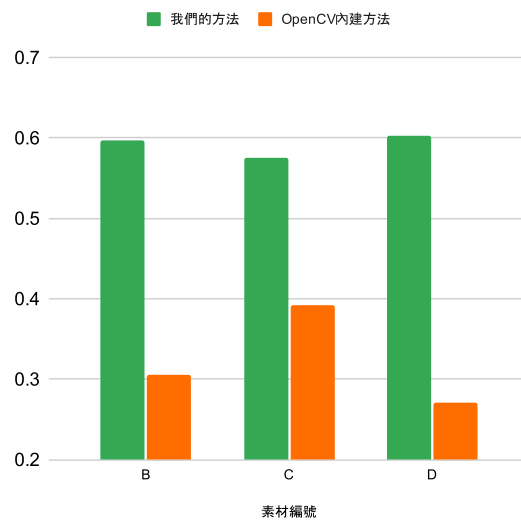


圖 30 本研究提出的方法與現存方法各項評估指標的比較(縱軸起始數值不一定為零)

五、探討使用不同機器學習方法的動畫角色辨識結果

使用準備好的臉部特徵集，先以人工標記角色標籤(label)，接著使用 Scikit-learn 裡的 RFC 和 SVM 來建模進行預測，最後與標籤比較計算得分。

(一)評量方式

以交叉驗證的方式設計實驗，並使用 Scikit-learn 內建函式 score 進行評分，回傳值計算方式如(式 9)：

$$score = \frac{TP}{testing\ data\ length} \quad (式\ 9)$$

(二)實驗結果

表 8 兩種機器學習方法的動畫角色辨識結果(表中數值為百分率)

機器學習工具	SVM		RFC	
實驗編號/素材編碼	B	D	B	D
1	90.6977	36.8421	85.7142	94.7368
2	76.7442	36.8421	91.8367	68.421
3	76.7442	47.3684	77.551	100
4	86.0465	42.1052	89.7959	68.421
5	74.4186	52.6315	97.9591	94.7368
6	76.7442	42.1052	93.8775	94.7368
7	74.4186	52.6315	93.8775	89.4736
8	58.1395	42.1052	87.7551	73.6842
9	67.4419	26.3157	95.9183	68.421
10	81.3953	36.8421	91.8367	84.2105
average	76.27907	41.5789	90.6122	73.78417
	58.928985		82.198185	

由表 8 可知，在動畫角色辨識上，使用 RFC 的效果優於 SVM，推測是因為 SVM 有部分特徵值造成了反效果，而 RFC 則可以找出明顯的分類判斷式，像是髮色等明顯的區分條件。

六、評量漫畫成果

設計問卷，請受試者針對漫畫成果進行評測。

(一)問卷設計

問卷分為三個部分：

1. **第一部分：基本資料** (預計耗時：小於 1 分鐘)
蒐集受試者的性別、年齡。
2. **第二部分：觀看原始動畫與漫畫成果** (預計耗時：3 分鐘)
請受試者觀看原始動畫和漫畫成果。
3. **第三部分：評測** (預計耗時：1 分鐘)

(1) 量化問題(分五級距)

整體的表現、畫面的還原程度、對話的還原程度、對話與角色的匹配程度

(2) 質化問題

對漫畫成果的評論

(二)回收結果

在報告書定稿前，共收到 21 位有效受試者的回應，性別以男性居多，年齡主要分佈在 13 至 18 歲。問卷第三部分的評測結果請見圖 31 與表 9。

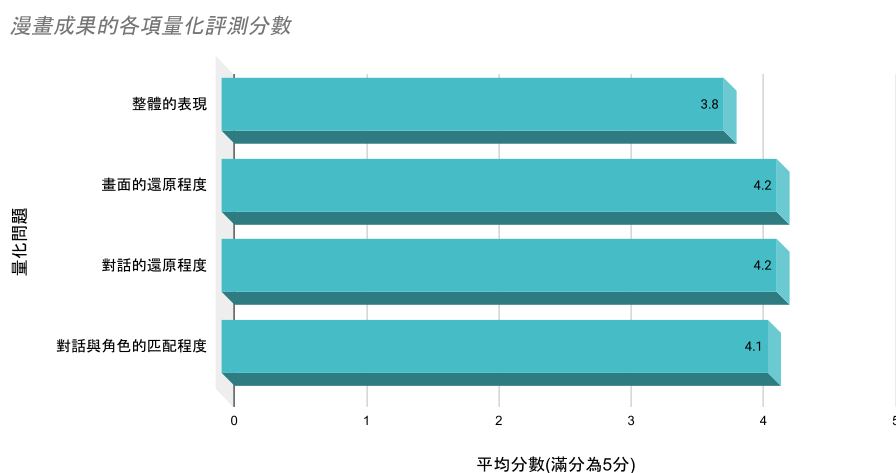


圖 31 漫畫成果的各项量化評測分數

表 9 質化問題的部分結果

非常好畫面，動作非常好👍
我認為美工部分有美中不足的地方 但發想和處理部分我給很高的評價
就我的了解，較細緻的漫畫會考慮到每一欄的大小，甚至會搭配跨頁的畫法表達張力，或許這能作為後續精進的一種方式。

從圖 31 和表 9 可以得知，結合畫面、對話與角色生成的漫畫成果在問卷中的量化與質化問題皆獲得正向回饋。

伍、討論及應用

一、漫畫化過程中的困難與解決方法

(一)常見的語者辨識錯誤狀況

觀察辨識錯誤的對話，歸納出一些會導致辨識錯誤的原因：

1. 將片頭、片尾曲納入辨識

當歌曲中的女聲加入辨識時，會使女性角色的整體音高特徵高於正常說話值。如此一來，當女性角色說話聲調略低時，就容易分辨錯誤。將片頭曲、片尾曲靜音後再進行辨識能解決上述問題。

2. 角色的尖叫聲

角色的尖叫聲同樣會影響整體音高特徵造成分辨錯誤。在進行辨識前，預先處理動畫聲音，將背景雜音移除。而角色的尖叫聲恰好會被視作雜音而剔除，即可解決此問題。

(二)修正膚色數值範圍

初次測試本研究的臉部偵測方法時，評估結果不甚理想，觀察後發現，不同動畫的膚色差異性會影響結果。在原先的偵測方式中，膚色數值參考(Takayama et al., 2012)，設定成 $HSV=[(0,40],[0,255],[0,255])$ 。但經測試後發現，此限制範圍過於寬鬆，使得許多顏色相近的背景色塊被當作臉部候選區域選入。作為改善方式，本研究決定依照不同的素材微調膚色數值的取用範圍。針對 HSV 中的 H 設限調整膚色數值範圍，確實能有效提升偵測效果。

(三)OpenCV 中照片傳遞(讀取)問題

在探討臉部偵測的方法時，曾觀察到圖片中有許多不合常理、未知來源的有色像素。後來發現這可能是 OpenCV 在傳遞或讀取照片時，產生肉眼無法發現的「雜訊」像素，這會使在判斷臉部候選區域的邊緣時發生問題。

藉由統一改用陣列傳遞圖片資訊，取代先輸出再讀取的圖片傳遞方法，成功解決此問題。

二、與現存的漫畫化方法比較成果

相較於現有文獻的漫畫化方法，本研究在許多步驟進行了有效的優化，特別是在角色辨識中的臉部偵測上：增加特徵擷取的項目，讓臉部的判斷不僅僅依靠單一特徵⁴。

除此之外，本研究的漫畫成果更具有對話框—能夠完整呈現動畫中角色的對話，幫助讀者了解劇情，解決現有文獻生成的漫畫無法清楚傳遞劇情的問題，乃一大創舉。現有文獻生成的漫畫如圖 32 及圖 32，其研究生成的漫畫成果皆未包含角色間的對話。

⁴ 在(Takayama et al., 2012)中，當下巴法的 $R^2 < 0.85$ 即判斷該候選區域非臉部。

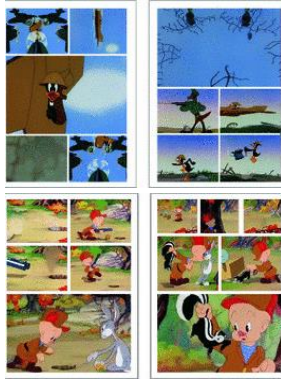


圖 32 (Fukusato et al., 2016)的漫畫成果



圖 33 (Ishikawa et al., 2014)的漫畫成果

三、比較漫畫成果與原作漫畫

在研究動機曾提過，本研究的初衷是為了讓大眾也能觀看「原創動畫」的漫畫版本。但為了檢驗生成的漫畫成果是否能有效傳遞劇情，本研究選用的動畫素材皆為「漫改動畫」，有原作漫畫可供比較。比較結果置於 [P.32 附錄貳](#)，請參照。經研究者比較，本研究自動生成的漫畫與原作漫畫高度相似，且沒有劇情上的缺漏，能夠讓讀者充分了解動畫內容。

四、本作品在實務上的具體應用

(一)生成的漫畫成果

本研究之成果能作為快速得知動畫內容的途徑之一，透過具有對話框的漫畫成果，幫助那些有興趣但沒時間的動畫同好能快速了解動畫劇情與畫風。

自動漫畫化的技術更能節省人工漫畫化的勞力與時間。漫畫編輯者可以直接根據本研究之成果，經修改、微調後出版，省去人工選取劇照與擷取台詞的耗時與勞力。

(二)擷取的資訊與分鏡表

在本研究中透過各式發想的方法成功擷取動畫中的重要資訊：畫面、對話與角色，這些資訊能提供使用者檢索動畫的分鏡片段。本研究以分鏡表的形式呈現資訊更讓這些資訊一目瞭然。

五、釋出程式碼供大眾使用

為了使所有人皆能使用本研究按需求將動畫漫畫化，公開程式碼於 <https://replit.com/@SamuelYang1323/Comicalization-Of-Animations#readme>。

為何選擇釋出程式碼而非架設網頁？在研究的過程中，發現隨著輸入的動畫的不同，許多閾值、參數的最適數值都會有所改變，如：膚色數值的範圍、臉部偵測的閾值等。動畫素材的畫風與分鏡切換的節奏乃不可抗力的外變因，而不同動畫的最適參數仰賴人工微調。藉由公開程式碼，能讓使用者依輸入的動畫自行調整參數，以取得最好的漫畫生成效果。

陸、結論

透過分鏡分割、語者自動分段標記及機器學習等技術，結合新發想的臉部偵測方法，成功將動畫以漫畫格式呈現，且輸出之漫畫成果具有對話框，能解決現存漫畫化方法無法傳遞劇情的問題。漫畫成果在問卷調查中獲得受試者們的正面回饋，也透過與原作漫畫比較確認還原劇情的能力。

進行了相關實驗以驗證效果各個自動化步驟的效果，得到以下結論：

- 一、在動畫的分鏡分割方法中，「適應內容偵測法」在閾值 $T_a=3$, $T_c=25$ 、參數 $W=2$, $M=15$ 時，對於所有實驗素材皆能取得九成以上的的 F 度量。說明在「取得畫面」步驟，使用的分鏡分割工具能有效分割動畫分鏡。
- 二、以語者自動分段標記技術進行語者辨識，提取模型及聚類器的組合以 ECAPA+SC 有最佳的動畫對話內容辨識效果，DER 僅 21.7%。說明在「提取對話」步驟，使用的語者辨識工具能有效以聲音匹配角色台詞。
- 三、提出動畫角色的脸部偵測方法明顯優於現存的真人脸部偵測方法，偵測結果平均 F 度量達 0.591，遠高於現存方法的 0.323。說明在「辨識角色」步驟，能有效找出畫面中人物的位置。
- 四、使用動畫角色的脸部附近特徵進行角色辨識，RFC 優於 SVM，平均 score=82.198185%。說明在「辨識角色」步驟，能有效得知不同角色在畫面中的位置。

相較於先前的動畫化研究，本研究針對動畫設計出一套能考量角色間對話的漫畫風格摘要方式。在生成的漫畫成果以對話框完整呈現動畫中角色的對話，幫助讀者了解劇情，開創了全新的漫畫化形式，也回應最初的研究動機 — 讓讀者藉此漫畫了解動畫內容。

本研究在應用面上，除了可以作為人們了解動畫內容的新途徑，亦可減少人工編輯漫畫版本所耗之心力，幫助相關產業發展。

柒、參考資料及其他

一、參考文獻

- [1] Fukusato, T., Hirai, T., Kawamura, S., & Morishima, S. (2016, January). Computational Cartoonist: A Comic-Style Video Summarization System for Anime Films. *22nd International Conference on Multimedia Modeling 2016*. ResearchGate. 10.1007/978-3-319-27671-7_4
- [2] Haq, I. U., Muhammad, K., Hussain, T., Kwon, S., Sodanil, M., Baik, S. W., & Lee, M. Y. (2019, June 25). Movie scene segmentation using object detection and set theory. *SAGE journals*. Retrieved from <https://journals.sagepub.com/doi/full/10.1177/1550147719845277>
- [3] Ishikawa, T., Okazaki, K., Kakimoto, M., & Nishita, T. (2014, August 10-14). *A Video Summarization Technique of Animation Products According to Film Comic Format*. Retrieved from http://nishitalab.org/user/UEI/publication/Ishikawa_SIG2014.pdf
- [4] Takayama, K., Johan, H., & Nishita, T. (2012, November 21-24). Face Detection And Face Recognition Of Cartoon Characters Using Feature Extraction. *Proceedings of the IEEE Image Electronics and Visual Computing Workshop 2012*. Retrieved from <https://www.semanticscholar.org/paper/FACE-DETECTION-AND-FACE-RECOGNITION-OF-CARTOON-Takayama-Johan/080ecd6c10fb637490e5dd7d1b592a18418f33c7>

二、使用工具

- [1] Cvqluu/simple_diarizer (https://github.com/cvqluu/simple_diarizer)
- [2] Open Source Computer Vision Library, OpenCV (<https://opencv.org/>)
- [3] PySceneDetect (<http://scenedetect.com/en/latest/>)

三、圖片來源與版權歸屬

請參照 [P.35 附錄參](#)。

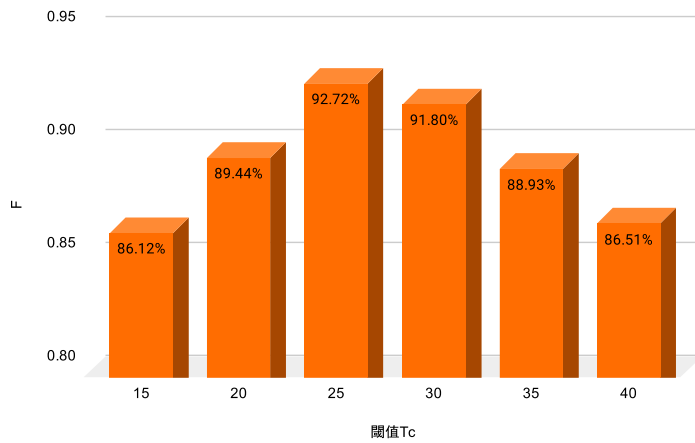
附錄

壹、其他實驗結果

使用素材 A 進行分鏡分割的參數實驗

[適應內容偵測法]F度量與閾值Tc的關係長條圖

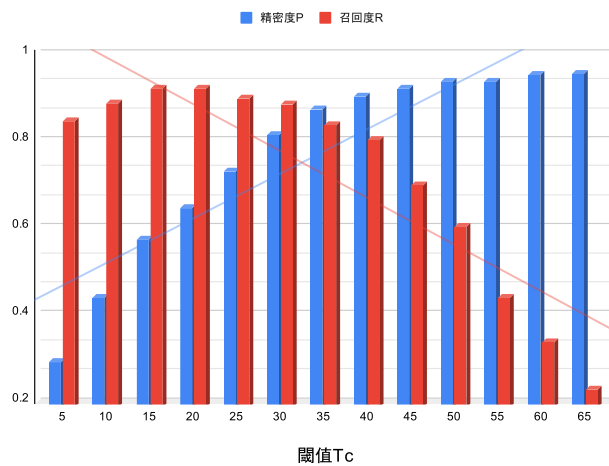
縱軸:F, 橫軸:Tc



[適應內容偵測法]操縱變因為W的評估指標表

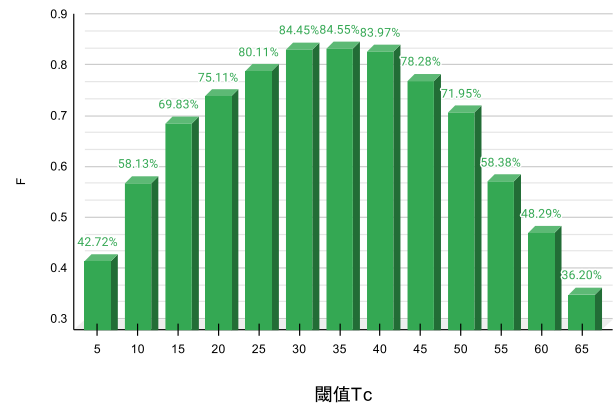
W	P	R	F
1	0.7007755	0.922652	0.791852
2	0.8147268	0.9322213	0.8611585
3	0.7903324	0.9223503	0.8472875
4	0.8226465	0.9194895	0.862942
5	0.788861	0.924146	0.8466359
6	0.8231277	0.921625	0.8648555
7	0.7964864	0.926527	0.8524966
8	0.8046229	0.9292334	0.8579296
9	0.8037959	0.9365667	0.8612038
10	0.7954635	0.9328325	0.8544416

[內容感知法]精密度、召回度與閾值Tc的關係長條圖

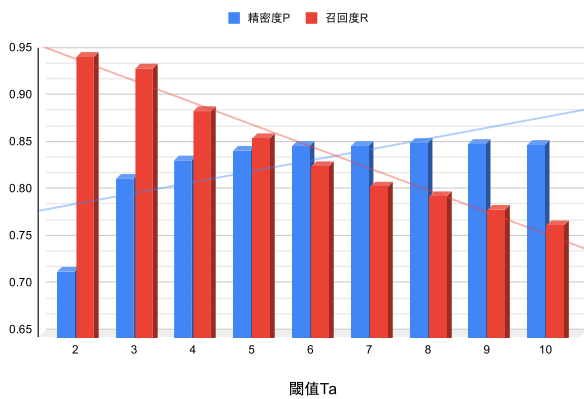


[內容感知法測試]F度量與閾值Tc的關係長條圖

縱軸:F, 橫軸:Tc

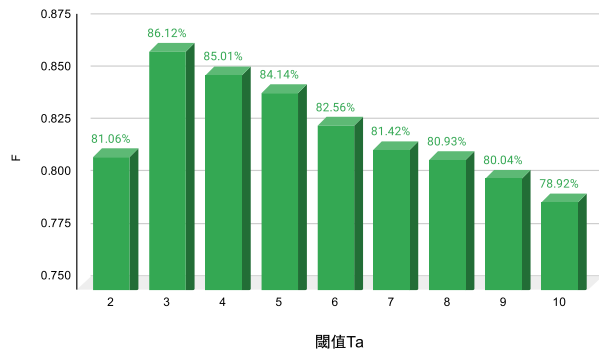


[適應內容偵測法]精密度、召回度與閾值Ta的關係長條圖



[適應內容偵測法]F度量與閾值Ta的關係長條圖

縱軸:F, 橫軸:Ta



貳、漫畫成果與原作漫畫之比較

以下為素材 B 的原作漫畫與本研究的漫畫成果之比較。

原作漫畫	自動生成的漫畫成果
 <p>我到底... 在幹什麼啊...</p>	 <p>我究竟在幹什麼呢</p>
 <p>竟然... 在窗簾後面...</p> <p>啊哈哈 沒事吧?</p> <p>哇... 唔啊!!</p>	 <p>竟然藏在窗簾後面...</p> <p>沒想到吧</p>

可惡啊……
表情還這麼開心



可惡 一副得意洋洋的表情



吶西片

可惡……
高木同學……

学生会小組

你不覺得
這個世界只有
我們兩個嗎？

在安靜的教室裡
兩人獨處



14

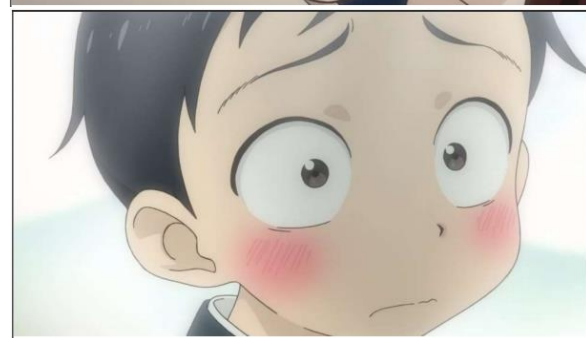


可惡的高木同學
女主角：喂 西片



男主角：你不覺得 兩個人獨處在安靜的教室裡
女主角：會有一種整個世界只有我們兩個人的感覺嗎





原作漫畫來源：<https://m.manhuagui.com/comic/14735/147012.html#p=1>

漫畫成果原圖來源：動畫素材截圖。版權歸屬：からかい上手の高木さん ©シンエイ動画株式会社

參、圖片來源與版權歸屬

圖 1 熱門動畫電影的漫畫版書照	3
圖片來源：左 https://reurl.cc/k7zGnn ；右 https://reurl.cc/Opyxmx	
圖 2 動畫的層次結構示意圖	4
研究者編繪。素材來源：(Haq et al., 2019)	
圖 3 人與電腦在動畫分割上的差異	4
研究者編繪。素材來源：(Haq et al., 2019)	
圖 4 動畫漫畫化之流程圖	5
研究者編繪。	
圖 5 「取得畫面」架構圖	5
研究者編繪。	
圖 6 素材的第 91 影格與第 92 影格	7
圖片來源：動畫素材截圖。版權歸屬：鬼滅の刃 © ufotable, Inc.	
圖 7 邊界像素判斷方式示意圖	8
研究者編繪。	
圖 8 動畫劇照在經邊界偵測處理前後之比較圖	8
程式製圖。原圖來源：動畫素材截圖。版權歸屬：炎炎ノ消防隊 © 株式會社 david production	
圖 9 「提取對話」架構圖	9
研究者編繪。	
圖 10 語者辨識工具生成的波形暨分割例圖	9
圖片來源： https://github.com/cvqluu/simple_diarizer	
圖 11 「辨識角色」架構圖	10
研究者編繪。	
圖 12 「臉部偵測」架構圖	11
研究者編繪。	
圖 13 左至右依序為 RGB 原圖、轉 HSV 圖、邊界圖與加寬變細後的邊界圖	11
程式製圖。原圖來源： https://reurl.cc/e6rVeb 。版權歸屬：からかい上手の高木さん ©シンエイ動画株式会社	
圖 14 由左至右分別為原圖、全部的膚色區域、符合限制的膚色區域	12
程式製圖。原圖來源： https://reurl.cc/l914rq 。版權歸屬：炎炎ノ消防隊 ©株式會社 david production	
圖 15 左上與左下分別為原圖與候選區域；右圖為輪廓分布(紅點)與擬合二次曲線(綠線)	13
程式製圖。原圖來源： https://reurl.cc/AOOrlp 。版權歸屬：ハイキュー!! ©株式會社 Production I.G	
圖 16 臉候選區對應的擬合二次曲線(白線)與半橢圓(粉紅線)	13
程式製圖。原圖來源： https://reurl.cc/AOOrlp 。版權歸屬：ハイキュー!! ©株式會社 Production I.G	
圖 17 原圖、經填補後的臉候選區與其對應的擬合二次曲線(白線)與半橢圓(粉紅線)	14
程式製圖。原圖來源：動畫素材截圖。版權歸屬：からかい上手の高木さん ©シンエイ動画株式会社	
圖 18 臉部 y 座標差值數列示意圖	14
研究者編繪。	

圖 19 逆序數對圖解	14
研究者編繪。	
圖 20 由左至右依序為原圖、頭髮區域與平均髮色	15
程式製圖。原圖來源： https://reurl.cc/AOOr1p 。版權歸屬：ハイキュー!! ©株式會社 Production I.G	
圖 21 角色辨識過程示意圖	16
圖片來源：(Takayama et al., 2012)	
圖 22 資訊截取間的交互運作與漫畫的生成流程圖	17
研究者編繪。	
圖 23 手動匹配對話與角色示意圖	17
研究者編繪。	
圖 24 分鏡表節錄	18
程式製圖。素材來源：動畫素材截圖。版權歸屬：あつくんとカノジョ©株式会社ゆめ太カンパニー	
圖 25 輸出的漫畫節錄	18
程式製圖。素材來源：動畫素材截圖。版權歸屬：からかい上手の高木さん ©シンエイ動画株式会社	
圖 26 由遠至近瞬間的視角切換	20
圖片來源：動畫素材截圖。版權歸屬：鬼滅の刃 © ufotable,Inc.	
圖 27 內容感知法之閾值與各評估指標的關係(縱軸起始數值不一定為零)	21
研究者編繪。	
圖 28 適應內容偵測法之閾值與各評估指標的關係(縱軸起始數值不為零)	21
研究者編繪。	
圖 29 以 OpenCV 內件工具偵測臉部的結果舉例	23
程式製圖。原圖來源：動畫素材截圖。版權歸屬：ハイキュー!! ©株式會社 Production I.G	
圖 30 本研究提出的方法與現存方法各項評估指標的比較(縱軸起始數值不一定為零)	24
研究者編繪。	
圖 31 漫畫成果的各项量化評測分數	26
研究者編繪。	
圖 32 (Fukasato et al., 2016)的漫畫成果	28
圖片來源：(Fukasato et al., 2016)。版權說明：上圖“Daffy The Commando”(1943, public domain)；下圖“A Wild Hare”(1941, public domain)	
圖 33 (Ishikawa et al., 2014)的漫畫成果	28
圖片來源：(Ishikawa et al., 2014)。版權歸屬：“Eden of The East the movie I - The King of Eden -”, © Asmik Ace, Inc.	