

國立臺北科技大學  
電機工程系  
一百零八學年度專題製作報告

專題名稱：

i-Piano 愛鋼琴

具樂譜分析與演奏評分機制之鋼琴練習系統

Piano practicing system with automatic sheet music  
recognition and performance scoring mechanism

指導教授：黃正民 副教授

專題組員：吳博翰 105310102

程滢臻 105350020

中華民國一零八年十二月十日

一百零八學年度專題製作

具樂譜分析與演奏評分機制

之鋼琴練習系統

研 究 組 員

吳 博 翰

105310102

程 澄 臻

105350020

組別：控制組

班別：電機四丙

指導老師：\_\_\_\_\_（簽章）

# 目錄

一、摘要	3-3
二、簡介	
(一) 研究動機	4-5
(二) 研究目標	6-7
(三) 系統開發工具	8-11
三、理論	
(一) 文獻回顧與探討	12-15
(二) 研究方法與步驟	16-20
(三) 系統架構圖	21
四、實驗	
(一) 工作進度與甘特圖	22
(二) 系統流程圖	23-24
(三) 系統測試	25-34
(三) 使用者介面	35-42
五、結論	
(一) 結論說明	43
(二) 未來展望	44
六、致謝	45
七、參考文獻	46-47

## 一、摘要

近年科技快速的進步，不斷地改變我們的生活模式，不管是醫療、娛樂、教育等等皆受其影響，尤其對於教育而言更是重大的改變。智慧型手機、平板的普及使學習更佳便利，而且學習方式也比以往更多元，透過機器輔助學習及增加趣味的模式備受矚目。舉例來說，在家自學鋼琴是件相對較為困難的學習，因此，如何方便且準確地獲取琴譜資訊，對彈奏部分進行精準的節奏、音準、指法判斷，並對其學習狀況進行紀錄及分析，以達到自主學習且同時提升學習效率之目的，是一個很有研究價值的問題。

此系統主要可分為「樂譜識別」、「彈奏判別」及「評分系統」三個部分，其一樂譜識別主要是透過行動裝置上的相機可將紙本樂譜資訊數位化，產生較容易閱讀與指導的影像和能立即理解樂譜的音檔，並同時儲存在雲端，當使用者有使用的需求時可以直接透過通訊軟體 Telegram 的指令獲取曾經匯入的樂譜及其所產生的音檔。其二彈奏判別的部分，此系統使用平板電腦連接 RGB 攝影機，即時取得使用者的彈奏資訊，並進行資料的分析，且將結果顯示、記錄存至雲端資料庫中，使用者亦可以透過 Telegram 的指令呼叫清單、查看紀錄。最後，評分系統則是建立學習者彈奏的評分機制及分析，提供使用者日後查看，並能更進一步將學習成果與同好分享或競賽，增進學習動機與樂趣。

## 二、簡介

### (一) 研究動機

樂譜是學所有樂器的基本，西方音樂之所以能夠普及並流傳，就是因為記譜系統的發展以及印刷術的普及，所以這個應用程式的開發就從樂譜出發，以樂譜為基礎，目標是以科技輔助音樂學習。

現今的科技在鋼琴或電子琴樂器的發展上，已出現平板電腦結合琴體，協助顯示電子樂譜以及上網瀏覽彈奏樂曲旋律和影片。

目前市面上主要與鋼琴、樂譜相關的產品有兩類，分別是試圖取代實體鋼琴的「虛擬鋼琴」以及以樂譜為主軸的「智慧琴譜」。其中，虛擬鋼琴又分成可以以雷射投影方式顯示的產品及手機平板螢幕顯示的 APP 產品。相關產品與本研究的功能及優缺點比較如下表 2.1 所示。

鋼琴	投影虛擬琴鍵	平板手機顯示	結合真實琴體	i - Piano
樂譜獲取方式	提供特定曲目或自行建立電子檔資訊			拍攝任意紙本樂譜， 自動辨識音符及指法
音訊辨識	有	有	有	有
指法辨識	無	無	無	有
雲端資料彙整平台	有	有	無	有
學習歷程分析	無	無	無	有

表 2.1 本專題與其他相關產品之功能、優缺點比較表

虛擬鋼琴是指使用非傳統鋼琴或電子鋼琴本身的鍵盤做彈奏，主要又分成用手機或平板裝置上的螢幕顯示的鋼琴鍵盤，以及用雷射投影的方式將鍵盤投影在

平面上的鋼琴鍵盤，其中雷射投影的鍵盤的感測方式是用紅外線感測偵測手指敲擊位置，設備上需要額外的雷射投影裝置及紅外線感測器。但以上兩種我們僅能由彈奏的琴鍵是否正確來判斷彈奏是否正確，無法得知指法是否也正確，而樂譜的提供也受限於 APP 的提供或需額外收費。智慧琴譜方面，雖然一樣是以實體的電子鋼琴呈現，但樂譜的來源一樣受限於 APP 的提供或需額外收費，與虛擬鋼琴一樣無法判別彈奏的指法是否正確。

總之，這些新科技的導入，仍未沒有帶來太多學習上的便利、技巧上的提升或是彈奏上的樂趣。因此，本專題建構了一個整合型智慧鋼琴練習輔助系統，藉由簡單的感測器裝設與手機或是平板的應用，就能夠輕鬆升級原本的琴體，並透過圖形化介面的設計，提供使用者可輕鬆簡單的完成各項操作。

## （二）研究目標

### 1. 解決紙本不方便的問題

為了使樂譜更方便攜帶，所以如果能將樂譜拍照，以數位化的方式儲存，對使用者而言會更為方便。但是因為圖片的扭曲、傾斜以及尺寸被縮小等問題都可能造成使用者閱讀上的不便，為此，將裝置所辨識到的樂譜經過傾斜校正處理後再以數位形式儲存，更能提升使用及學習上的便利性。

本研究使用手機或平板裝置的攝影機，使用者可以使用通訊軟體Telegram，將樂譜拍照後傳給TelegramBOT，接著TelegramBOT會回傳針對文本部分的傾斜進行校正後的樂譜並儲存在雲端資料庫，同時TelegramBOT也會回傳樂譜的MIDI音檔，可直接在Telegram上聆聽。

### 2. 了解自己的練習狀況

初學者在沒有老師指導時可能無法發現自己節奏不穩或彈錯，即使用錄音的方式去記錄練習的過程，也不容易將我們遇到的問題歸納，所以初學者在自主練習的情況下的確需要一個有效率且有系統的方式來記錄練習的過程。另外，如果能將每次練習的結果數據化，提供日後參考，可以明確知道自己的練習狀況及學習進度。

本研究直接使用電子鋼琴的音訊輸出，取得彈奏的音訊檔案，並將電子鋼琴所收到的音訊檔案與樂譜辨識後產生的音訊檔案比較，判斷使用者所彈奏的音是否正確，拍子是否在拍點上，音長是否足夠，以及當收到音訊時，RGB攝影機所辨識到的手指是否與樂譜上的提供的指法資訊一致，並綜合以上因素計分，最後將不符合標準的音以不一樣的顏色記錄在樂譜上，讓使用者能更清楚地了解彈奏問題所在。

### 3. 判斷並記錄指法

學者牛曼 (William Newman) 認為「一個好的指法可以成就一首樂曲」。學習一首新樂曲，一定要先弄好指法的問題。鋼琴指法是用鋼琴表現音樂的重要手段，音樂由節奏、旋律、和聲三大要素構成。節奏的準確、旋律的歌唱、和聲的和諧都與選擇適當的指法密切相關。

大部分的初學者很難在練習的過程中確認自己當下的指法正不正確，也很難在事後回想起當下的指法，所以需要一個完善的系統為使用者紀錄指法。

本研究直接以 RGB 攝影機接收影像，當系統收到音訊時會直接紀錄下當時的所使用的手指頭，可以讓使用者清楚知道自己使用的指法與樂譜上的是否一致。

### 4. 人性化使用介面

現今，APP 的發展越來越成熟，APP 開發的門檻越來越低，越來越多的 APP 上架。但近年來，使用者對於 App 的下載意願降低，企業推出 App 的成本相對變高，而可以透過使用通訊軟體直接與使用者互動的聊天機器人因為不需要額外下載，而且與我們日常使用的通訊軟體結合，就越來越受使用者歡迎，對企業而言，因為各大通訊軟體幾乎都有開放聊天機器人 API 串接，相較於 APP 的開發及推廣，聊天機器人的成本遠比使用 APP 低很多。其中，Telegram 是一款具開源特性的跨平台即時通訊軟體，官方亦提供非常完整的 API 給開發者使用。

本研究即採用 Telegram BOT 作為樂譜上傳、樂曲檢索的平台，除此之外，亦提供查詢學習紀錄、評分結果的功能。



### （三）系統開發工具

#### 1. 影像處理：OpenCV

OpenCV 的全稱是 Open Source Computer Vision Library，是一個開源的計算機視覺程式庫。1999 年由英特爾公司發起並參與開發，以 BSD 授權條款授權發行，可以在商業和研究領域中免費使用。

OpenCV 可用於開發實時的圖像處理、電腦視覺以及模式識別程式。該程式庫也可以使用英特爾公司的 IPP 進行加速處理。而 OpenCV 被廣泛使用於解決人機互動、圖像分割，圖像、臉部及動作的識別，除此之外，在運動物體的追蹤、機器人操作以及擴增實境的運用上，OpenCV 都很受開發者喜愛。

OpenCV 用 C++ 語言編寫，其主要介面也是 C++ 語言，但是依然保留了大量的 C 語言介面，同時也有大量的 Python、Java 和 MATLAB/OCTAVE 函式庫可供使用。雖然僅有短短 20 年的歷史，但由於其開源特性，網路上的資源亦非常龐大，有利於後進者能站在前人肩膀上進行更進一步的開發或應用。

#### 2. 深度學習套件：TensorFlow

TensorFlow 最初由 Google Brain 的團隊所開發，用於 Google 的研究和生產，是 Google 繼 DistBelief 之後的第二代機器學習系統，其特點是 TensorFlow 定義數學運算為 node、定義張量在數學運算中的流動為 edge，再利用其兩者組成的流程圖做為呈現，讓所有的使用者可以非常清楚當前所設計的神經網路架構。

TensorFlow 提供了一個 Python、C++、Haskell、Java、Go 和 Rust 的 API，第三方包可用於 C#、.NET Core、Julia、R 和 Scala。本專案我們選擇使用 Python 進行數字辨識讀的模型訓練。

TensorFlow 的底層核心引擎由 C++實現，通過 gRPC 實現網路互訪、分散式執行。雖然它的 Python/C++/Java AP 共用了大部分執行代碼，但是有關於反向傳播梯度計算的部分需要在不同語言單獨實現。

目前只有 Python API 較為豐富的實現了反向傳播部分。所以大多數人使用 Python 進行模型訓練，但是可以選擇使用其它語言進行線上推理。此外 Google 還推出專用的處理器 TPU，讓 TensorFlow 可以有更低的耗電量以及更高的效能。

### 3. 開發語言：Python

Python 是一種廣泛使用的直譯式、進階編程、通用型的物件導向高階程式語言，其函式、模組、數字、字串都是物件，並且完全支援繼承、重載、衍生、多重繼承，有益於增強原始碼的複用性。強調程式碼的簡潔、明確，因此相比其他高階語言能夠用更少的程式碼完成功能，方便使用並可在大多數的系統中運行，以減少開發及維護成本的觀念進行發展。

此外，Python 提供了豐富的 API 和工具，也具有腳本語言中最豐富和強大的函式庫，足以支持許多日常應用，像是 Web 程式、作業系統、科學計算、GUI 開發、遊戲、網頁爬蟲甚至機器學習等 Python 都經常被拿來應用。而本專案我們透過 Python 來整合 OpenCV 和 TensorFlow 這些影像分析及深度學習之間的功能。

### 4. 應用程式開發：TelegramBOT

Telegram 是跨平台的即時通訊軟體，其用戶端是自由及開放原始碼軟體，但伺服器是專有軟體。使用者可以相互交換加密，傳送相片、影片等所有類型檔案。官方提供手機版 (Android、iOS、Windows Phone)、電腦版 (Windows、macOS、Linux) 和網頁版等多種平台用戶端；同時官方開放應用程式介面，因此擁有許

多第三方的用戶端可供選擇。

為了支援各種不同的系統平台，官方提供多個用戶端版本，並允許各地好手開發自己的版本。它的伺服器系統分散並且分布在全球各地，以地理位置自動分配伺服器，加快全球回應速度。

Telegram 官方有提供非常完整的 API 給開發者使用，Telegram 的 API 分為兩種：其中一種是 Telegram API，可以提供開發者做一個自己的 Telegram Client；另一種則是 Bot API，可以讓開發者控制一個 Telegram Bot。

## 5. 數位音樂產生：MIDI

數位音樂介面（Musical Instrument Digital Interface，簡稱為 MIDI），是一個工業標準的電子通訊協定，在 1980 年代初期，由 MIDI 產業協會制定出第一版的標準協定，現今被廣泛應用於電子樂器、電腦與編曲機上。其記錄之資訊並非如 MP3、WAV 等的音訊頻率，而是傳送像是音調、音樂強度、節奏和音量等參數的控制訊號。

由於並非儲存頻率，MIDI 的優點是通常檔案大小都不大，就像是給電腦讀的電子樂譜，僅靠著必要的音樂資訊，就能由電腦演奏出美妙的音符。本研究預計將紙本樂譜數位化之後的資訊，以 MIDI 檔案格式儲存，以利於之後進行播放音樂以及比對音訊內容的作業。

## 6. 雲端儲存

### (1) imgur

imgur 從 2009 年開始提供免費圖片空間的服務，整體介面與操作方式相當簡單，提供直接連結，可外部連結圖片。同時支援多種常見圖片格式，例如

JPEG, GIF, PNG, TIFF, BMP, PDF, XFC (GIMP)。圖片在上傳後會保存最少一個月，假如一個月內都無人瀏覽就會自動刪除。

## (2) Dropbox

Dropbox 是 Dropbox 公司的線上儲存服務，通過雲端儲存實現網際網路上的檔案同步，用戶可以儲存並共用檔案和資料夾。

Dropbox 提供免費和收費服務，在不同作業系統下有客戶端軟體，並且有網頁客戶端。用戶可以通過 Dropbox 桌面應用軟體，把檔案放入指定資料夾，然後檔案就會被同步到雲端，以及該用戶其他裝有 Dropbox 桌面應用軟體的其他電腦中。Dropbox 資料夾中的檔案隨後就可以與其他 Dropbox 用戶分享，或通過網頁來取得。用戶也可以通過網頁瀏覽器來手工上傳檔案。

此外，Dropbox 支援恢復歷史紀錄，即使檔案被刪，也可以從任何一個同步電腦中得以恢復。用戶通過 Dropbox 的版本控制，可以知道他們共同作業檔案的歷史紀錄，即使多人參與編輯、再發布檔案，就不會因為並行而遺失先前的紀錄。

### 三、理論

#### (一) 文獻回顧與探討

##### 1. 音樂記譜法

樂譜是一種以印刷或手寫製作，用符號來記錄音樂的方法。不同的文化和地區發展了不同的記譜方法，除了現今常見音樂的五線譜和簡譜外，在傳統音樂中還有工尺譜，以及古琴特有的減字譜。

音樂記譜法除了會記錄樂曲中的音準、節奏外，也會對力度、裝飾音、樂句段落等進行標記。本研究將以影像辨識分析五線譜中的音符記號。節奏標記以音符本身為主，需辨別是否為實心、有無附點、有無桿子、有無尾巴，不同型態的音符有各自不同的節奏長度。











音符名稱	圖示	休止符	圖示
全音符		全休止符	
二分音符		二分休止符	
附點二分音符		附點二分休止符	
四分音符		四分休止符	
八分音符		八分休止符	

表 3, 1.1 音符類別與休止符類別

音高的標記則以音符在五線譜上的位置來辨別，高音譜記號的音高如下圖

3.1.1-1 所示。



圖 3.1.1-1 高音譜記號的音高

同時，也會因為不同譜號而有不同的音高定義，如下圖 3.1.1-2 所示，分別是高音譜記號、低音譜記號，及中音譜記號於不同位置所代表的音高。

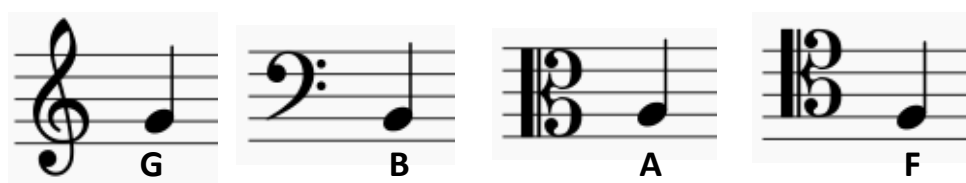


圖 3.1.1-2 在同一位置，不同譜號所代表的音高

另外，大部分初學者的樂譜都會附上指法。樂譜上的指法往往不是作曲者所寫的，而是編譯者根據自己的喜好、習慣將指法填上去的，同一首曲子在不同樂譜中可能會有許多不一樣的寫法。所以當學習者尚未具備自己判斷指法的能力時，通常會根據譜上標注的指法彈奏。左右手的部分都分別以 1 表示大拇指，2 表示食指，3 表示中指，4 表示無名指，5 表示小指，可參照下圖 3.1.1-3。



來源：[http://www.how01.com/post\\_IQMOaR8wpwRdz.html](http://www.how01.com/post_IQMOaR8wpwRdz.html)

圖 3.1.1-3 右手指法示意圖

## 2. 機器學習

機器學習演算法是從過往的資料和經驗中學習並分析，找到其運行規則，並利用規律對未知資料進行預測的演算法。

機器學習又可分為兩大類，分別是監督學習（Supervised Learning）與非監督學習（Unsupervised Learning）。其差別在於監督學習在輸入數據時會加上人工標注，並利用迴歸分析使程式得以做出預測；而非監督學習是讓演算法從大量的數據串流中，找出模式並將這些數據做出分類。機器學習已廣泛應用於資料探勘、電腦視覺、自然語言處理、生物特徵辨識、搜尋引擎、醫學診斷、檢測信用卡欺詐、證券市場分析、DNA 序列測序、語音和手寫辨識、戰略遊戲和機器人等領域。

## 3. 聊天機器人

聊天機器人是經由對話或文字進行交談的電腦程式，可用於實用的目的，如客戶服務或資訊獲取等，一般為大眾所熟知的虛擬助理（如 Google 智能助理）即屬於聊天機器人的範疇，聊天機器人可以與許多公司開發的應用程式、網站以及即時消息平台連接，達到更高的使用目的。有些聊天機器人會搭載自然語言處理系統，但大多簡單的系統只會擷取輸入的關鍵字，再從資料庫中找尋最合適的應答句。

目前聊天機器人廣泛運用於即時通訊平台，例如 Facebook Messenger、WeChat、LINE、Telegram 等，以娛樂、零售行銷、以及客服為目的。此外，即時通訊平台提供易於整合的 webhook，使得第三方開發商易於可通用於不同通訊平台之聊天機器人。這些軟體機器人以客服的身份出現或是成為團體聊天的一員。

#### 4. 手指辨識

手指辨識一直是影像處理中的一大學問。傳統影像處理方式通常是將手部顏色與背景分離後，取得影像中的輪廓，再從中找到凸包與凹包來確認手指位置。然而此種影像處理方式容易受到環境光、影子、遮擋等問題影響。因此本專題中使用的手指辨識方式為卡內基梅隆大學(CMU)於 2017 年提出的模型處理方式 - 《Hand Keypoint Detection in Single Images using Multiview Bootstrapping》。

CMU 的團隊擁有由 31 個高清攝影機所組成的多視圖系統。他們首先從一小組帶有標籤的手型圖開始訓練，利用卷積神經網路取得手型關節點的粗略估計，再將這些圖像通過檢測器來獲得更多的關節點預測。接下來，利用三角剖分已獲得關節點的 3D 位置，這些 3D 位置將重新再投影回 2D 圖像，以優化原先檢測器的 2D 預測。經過幾次的迭代便能夠大幅改善檢測器的預測準確率，最終優化完成的檢測器即為我們專題中所使用的手指辨識模型。



## (二) 研究方法及步驟

### 1. 樂譜辨識

#### (1) 樂譜數位化

##### a. 尋找文本部分

用 OpenCV 將圖片模糊化，可以讓文本內的線條輪廓變少，再做邊緣檢測，這時檢測到的邊緣大多是文本的外框，再利用 `findCounter()` 函式找到輪廓，此時找到的輪廓也會包括文本內的輪廓，所以找到文本輪廓後做輪廓逼近，即可以找到最符合文本的輪廓。

##### b. 文本轉正

使用 OpenCV 的透視變換函式將文本校正，同時將邊緣去除。OpenCV 中用來實現透視變換的函式有兩種，其中 `getPerspectiveTransform()` 是利用座標點計算矩陣，再利用矩陣完成轉換；另外 `findHomography()` 則是直接利用透視平面完成轉換。本研究採用的透視轉換函式是 `getPerspectiveTransform()`。

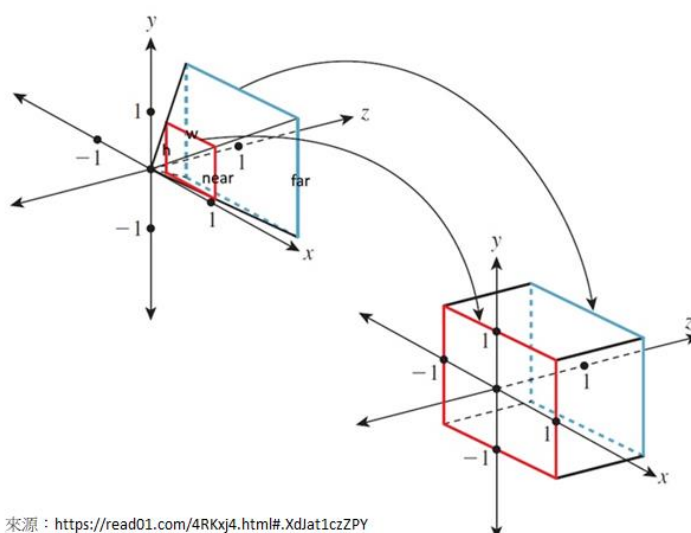


圖 3.2.1-1 透視轉換示意圖

### c. 二值化

如果要將圖片做最合適的二值化處理，首先必須先找到最適合的閾值。所謂最適合的影像所使用的閾值即為能使目標物體與非目標物體能有最明顯差異的閾值大小。

OpenCV 在二值化在設定閾值上主要有兩種作法，分別是固定閾值和自適應閾值。其中，自適應閾值又分成以算術平均數計算的自適應閾值（cv2.ADAPTIVE\_THRESH\_MEAN\_C），以及以高斯加權平均法計算的自適應閾值（cv2.ADAPTIVE\_THRESH\_GAUSSIAN\_C）。本研究經多次實驗後選擇採用高斯加權平均法計算的自適應閾值，使用在 101\*101 範圍內的平均像素和作為樂譜二值化時所採用的閾值。

## (2) 找五線譜及音符位置

### a. 模板匹配

圖像匹配是指圖像之間的比較，得到不同圖像之間的相似度，在機器識別的過程中把不同傳感器或同一傳感器在不同時間、不同成像條件下對同一景物獲得的兩幅或多幅圖像在空間上對準，或根據已知模式到另一幅圖中尋找對應的模式。匹配方法大體分為基於灰度和基於特徵兩類，其中基於特徵的配準方法研究較多。

基於灰度的模板匹配方法原理簡單且在光照良好的條件下可以得到比較滿意的匹配結果，故在本方案中根據輸入的模板圖像從採集圖像中搜尋可進一步匹配的目標可採用此方法；基於特徵的匹配方法要先提取各個圖像中的特徵再完成特徵間的匹配，通過匹配的特徵建立圖像間的映射關係求出匹配圖像。一般而言提取特徵點相對容易，能對圖像間的分辨率、旋轉、平移、光照變化等保持不變，故在利用模板匹配得到抓取的圖像後進一步進行特徵匹配。

為了利用模板匹配從源圖像中得到匹配區域，從源圖像選取該區域作為進行匹配的模板。模板從源圖像左上角開始每次以一個像素點為單位進行移動，每到達一個位置，就會計算模板矩陣和源圖像當前位置矩陣匹配的相似程度。

模板滑動與源圖像匹配過程中，將模板和當前模板覆蓋區域的矩陣的計算結果存儲在矩陣 R 中，矩陣 R 中每一個位置的座標都包含了匹配矩陣的計算結果。

OpenCV 提供了六種匹配度演算法，分別是差值平方和匹配法（CV\_TM\_SQDIFF）、標準化差值平方和匹配法（CV\_TM\_SQDIFF\_NORMED）、標準相關匹配法（CV\_TM\_CCORR\_NORMED）、相關匹配法（CV\_TM\_CCORR 及 CV\_TM\_CCOEFF）和標準相關匹配法（CV\_TM\_CCOEFF\_NORMED）。

本研究所採用的演算法是標準相關匹配法，是 OpenCV 所支援的演算法中最複雜的一種相似度演算法，計算出的相關係數被限制在了-1 到 1 之間：1 表示完全相同，-1 表示兩幅影像的亮度正好相反，0 則表示兩幅影像之間沒有線性關係如此可以保證影像和模板分別改變光照亮不影響計算結果。

#### b. 水平投影法與垂直投影法

在電腦視覺領域常使用垂直投影和水平投影可以對目標物進行精確投影，以便於後期的分割。

主要可依投影的方向分為水平投影及垂直投影。水平投影是指二維圖像在 y 軸上的投影，垂直投影則是二維圖像在 x 軸上的投影。

### (3) 數字辨識

在指法數字部分，我們採用機器學習的方式訓練模型來進行辨識。

Tensorflow 是由 Google 開發的機器學習套件，我們使用其中的工具來進行模型訓練。首先利用 Tensorflow 提供的 MNIST 數字資料庫進行訓練，由於鋼琴指法為數字 1~5，因此此模型僅以 1~5 的數字做為訓練資料。而在模型的選擇上則是使用 softmax 回歸模型架構出我們的神經網路，softmax 模型會將輸入的資訊，透過乘上權重與加上偏差，再經由 softmax 函式轉換成積率的方式，得到每一個類別所代表的可能性並作為輸出，進而決定輸入的資訊是屬於何種類別。

## 2. 雲端儲存

本專題之系統架構主要有兩大部分，分別是「使用者手機介面」與「彈奏評分系統」，而使用者手機介面包含了「樂譜辨識」與「聊天機器人」兩功能。在多重系統運作下，彼此資訊與處理結果必須互相交流使用。為了確保資料的一致性，我們利用雲端工具 imgur 和 Dropbox 進行儲存。

系統中主要處理的資料有數位樂譜圖片、樂曲清單、音訊 midi 檔、音符 bytes 資訊檔及使用者分數清單。

### (1) 數位樂譜圖片

使用 imgur 網站作為圖片庫的雲端儲存位置。使用者將圖片傳送給聊天機器人，首先會先將樂譜進行透視轉換，輸出為一個可供閱覽的琴譜。此琴譜會透過 imgur 提供的 api 上傳至其伺服器，並取得一個可訪問的連結 url。

### (2) 樂曲清單

樂曲清單儲存樂曲名稱與樂譜位置。在取得數位樂譜的連結與使用者輸入的樂曲名稱後，程式會下載 dropbox 雲端上的樂曲清單，將新增的樂譜資料更新

至檔案，再將更新後的樂曲清單上傳至 dropbox 雲端。

### (3) 音訊 midi 檔

樂譜辨識完成後所產生的 midi 檔案會儲存在 dropbox 雲端，使用者可以透過"樂曲查詢"功能來存儲音訊檔案。另外，在使用彈奏評分系統時，該系統會使用此 midi 檔案作為其中評分機制的一環。

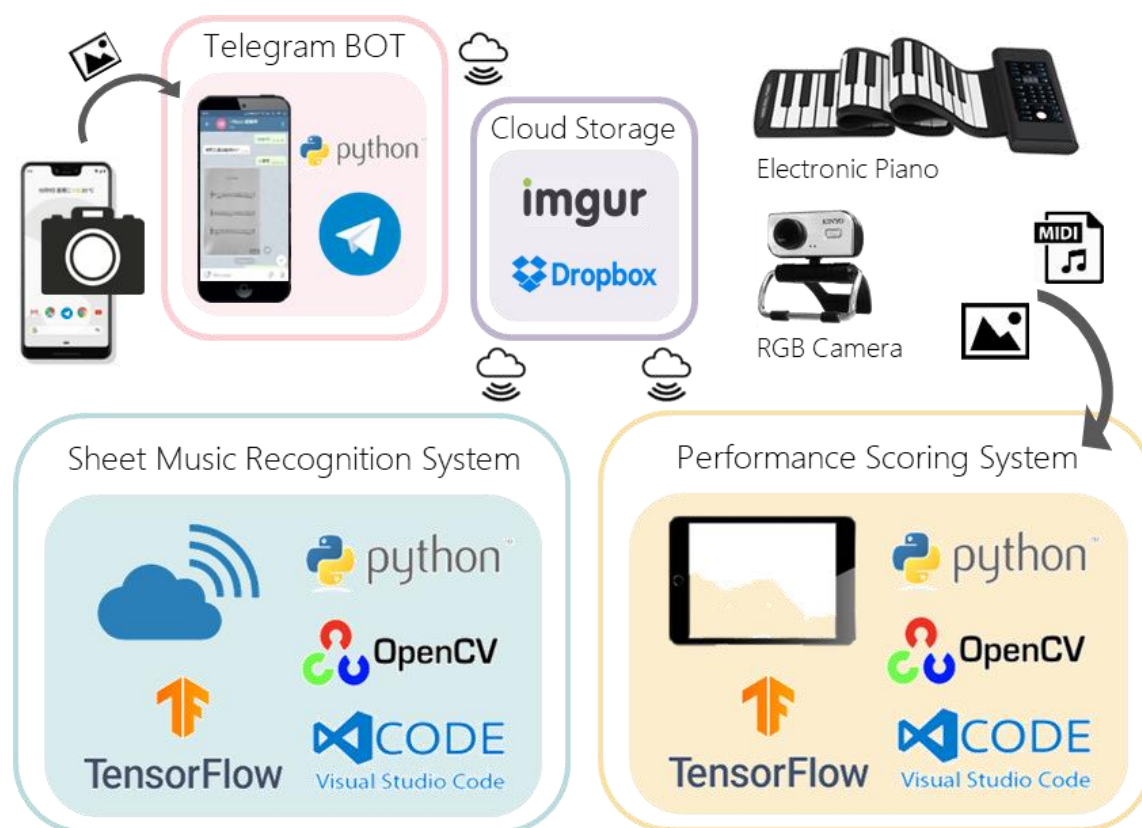
### (4) 音符 bytes 資訊檔

同樣在樂譜辨識完成後產生的音符資訊檔，其主要存取的資訊為音符位置、指法等非音訊相關的資訊。同樣儲存於 dropbox 雲端，在使用彈奏評分系統時，作為評分標準使用。

### (5) 使用者分數清單

每次的評分結果會儲存在 dropbox 雲端，使用者分數清單儲存彈奏評分系統之結果，除了可以回傳每次的分數紀錄，也可以回傳評分的紀錄圖。

### (三) 系統架構圖



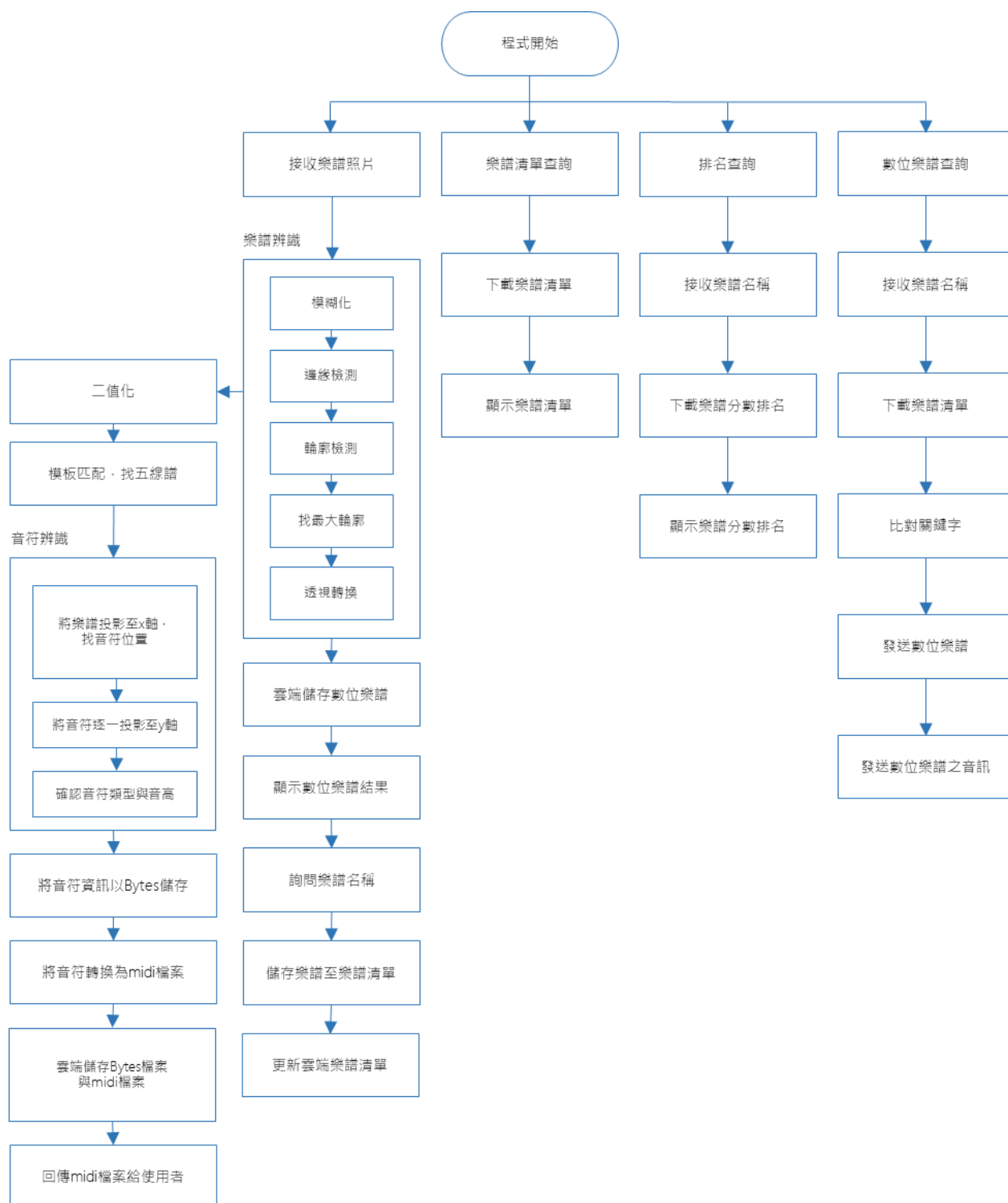
## 四、實驗

### (一) 工作進度與甘特圖

階段	預期目標	三月	四月	五月	六月	七月	八月	九月	十月
前期研究	定義範圍 / 蒐集資料								
	需求調查 / 競品分析								
	確立題目及系統架構								
第一階段	文本分析								
	音符辨識								
	音高判斷								
	數字辨識								
第二階段	TelegramBOT 流程設計								
	TelegramBOT 開發								
第三階段	彈奏評分系統介面								
	練習模式								
	評分功能								
	指法辨識								
第四階段	雲端儲存								
	系統整合								
結論	成品測試								
	系統優化								
	報告撰寫								

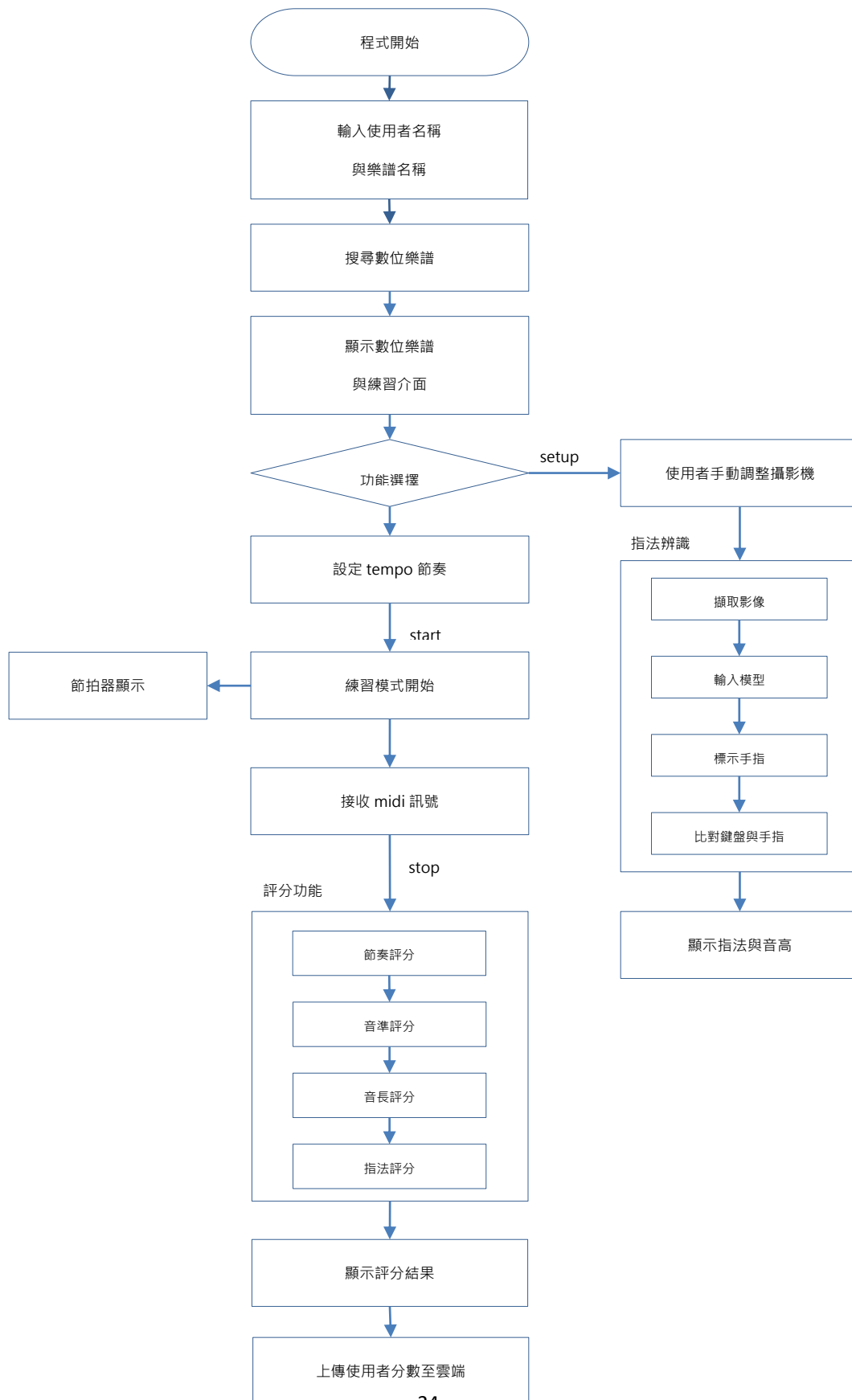
## (二) 系統流程圖

### 1. 樂譜辨識





## 2. 彈奏評分系統



### (三) 系統測試

#### 1. 樂譜辨識

##### (1) 紙本樂譜數位化

首先，以 Telegram Bot 傳入樂譜圖片（需保留四個完整的邊長），如下圖 4.3.1-1 所示。

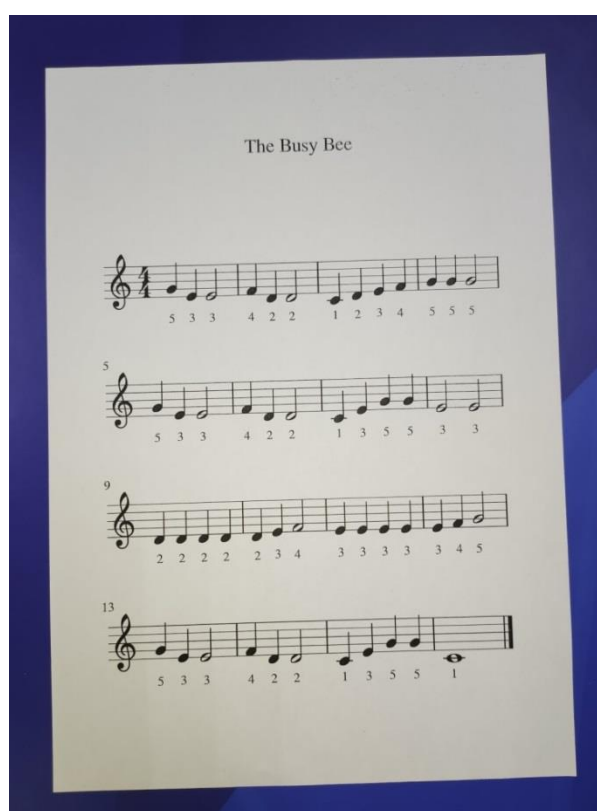


圖 4.3.1-1 拍攝的樂譜原圖

系統會將傳入的樂譜圖片經模糊化處理，盡量將樂譜內的內容模糊處理，讓整張圖片的邊緣最少化，以便下個階段尋找樂譜外框。

使用 OpenCV 的函式尋找圖像裡的輪廓，並將輪廓依大小排序，將輪廓最大的物體是為樂譜外框。

尋找樂譜外框的四個頂點，並判斷四個頂點的相對位置，並將四個頂點分別與右上、左上、右下、左下配對，計算轉換矩陣，使用 OpenCV 函式進行透視變換轉換。

最後 Telegram Bot 會回傳轉換後的樂譜，如下圖 4.3.1-2。

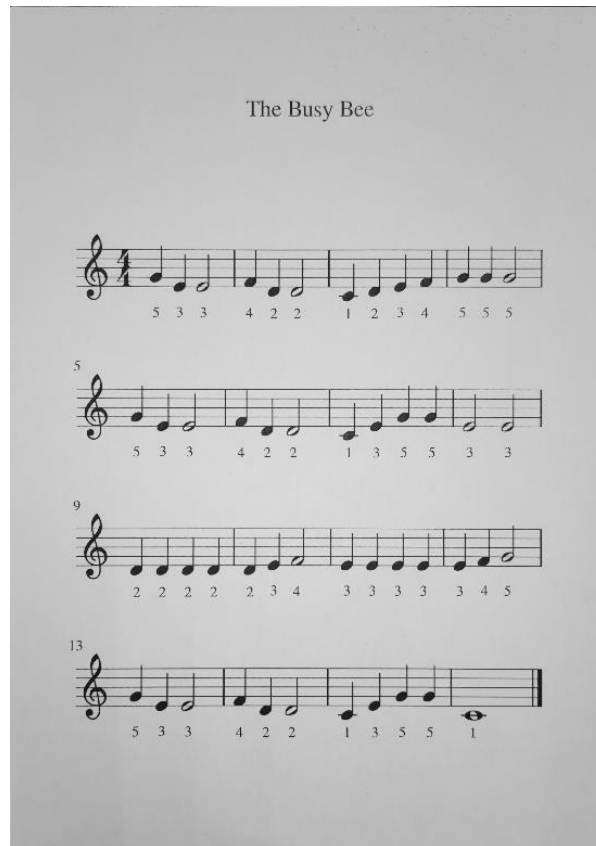


圖 4.3.1-2 傾斜校正後的樂譜

## (2) 音符辨識

### a. 模板匹配

OpenCV 有六種模板匹配的計算方法，本研究採用的是標準相關匹配法，是 OpenCV 所支援的演算法中最複雜的一種相似度演算法，計算出的相關係數被限制在了-1 到 1 之間：1 表示完全相同，-1 表示兩幅影像的亮度正好

相反，0 則表示兩幅影像之間沒有線性關係如此可以保證影像和模板分別改變光照亮不影響計算結果。

本研究在尋找五線譜的部分使用了模板匹配，使用到的模板如下圖

4.3.1-3。



圖 4.3.1-3 本專題中，五線譜進行模板匹配時所使用的模板

因為每份樂譜的規格不一樣，而模板也不一定會剛好符合每一份樂譜，所以本研究針對模板的比例進行縮放，尋找符合要求的模板。

本研究所使用的模板有三個，每個模板都會經過三次內插法（cv2.INTER\_CUBIC），將模板大小範圍設定成由 40%到 80%對整份經過傾斜校正後的樂譜進行比對，比對的相似度若大於 70%，會將結果存下來。

最後，將匹配到的模板畫出來，可以看到沒有音符的部分都會被模板所匹配到，如下圖 4.3.1-4。

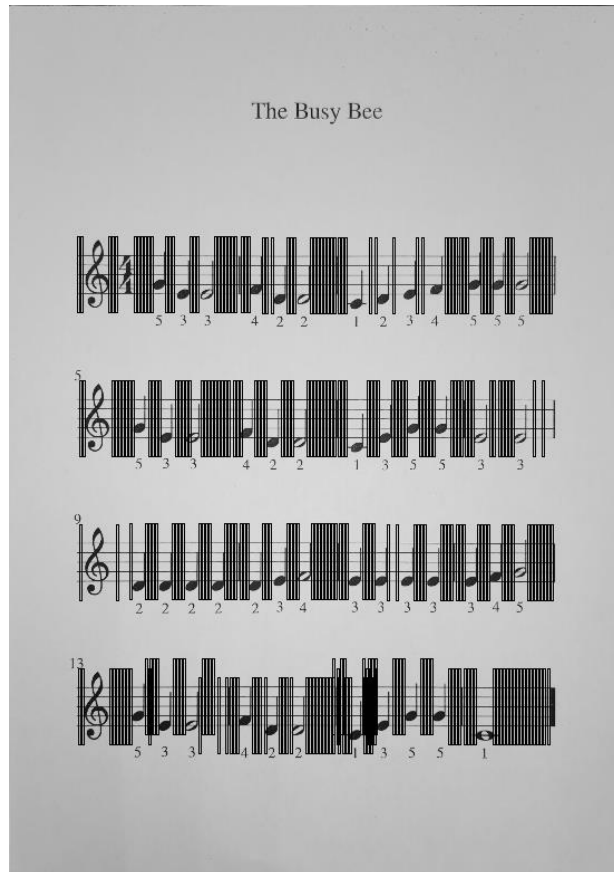


圖 4.3.1-4 模板匹配的結果圖

由模板匹配的結果發現，由於每一行最右邊都是沒有音符，接近空白的狀態，所以選擇使用每一橫排的最右邊所框到的模板，視為五線譜的大概位置。

#### b. 投影

取得五線譜約略的位置後，使用投影法讓所針對的目標的位置更加顯，以便取得目標的準確座標，以利後續音高的判斷。

首先我們選擇水平投影，使用每一橫排的最右邊所框到的模板的範圍，將其投影到到 y 軸，取得五線譜的投影，如下圖 4.3.1-5。

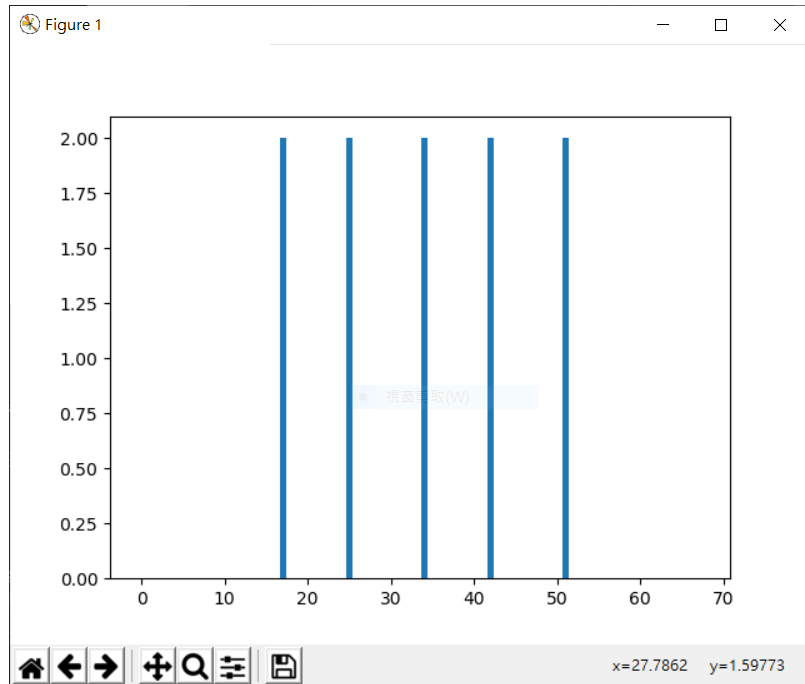


圖 4.3.1-5 模板匹配所取到的模板投影圖

我們將所選取到的模板的  $y$  座標視為五線譜大概的位置，接著將整行譜進行垂直投影到  $x$  軸上，如下圖 4.3.1-6 所示。

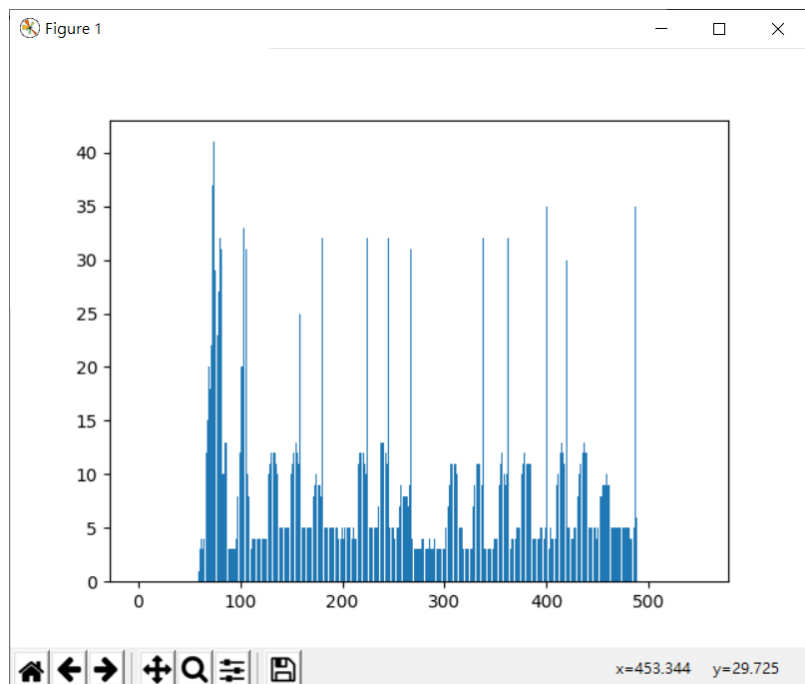


圖 4.3.1-7 整行譜垂直投影到  $x$  軸的投影圖

觀察上圖 4.3.1-7 可以發現，累計像素值在 4 到 6 之間的可能是五條線的累加，所以可以由此推斷累計像素值在 4 到 6 之間的位置是沒有音符的，而累計像素值超過 20 的則可能是小節線或是符桿的部分，其餘的部分則可能是四分音符、二分音符或全音符。

所以如果要取得音符的部分，必須把五線譜的部分、小節線及符桿扣掉，剩下的即為四分音符、二分音符或全音符。可以由觀察像素點分布的位置，判斷這些投影圖的音高，如下圖 4.3.1-8、圖 4.3.1-9、圖 4.3.1-10 與圖 4.3.1-11，分別是中央 C、中央 D、中央 F 及中央 G。

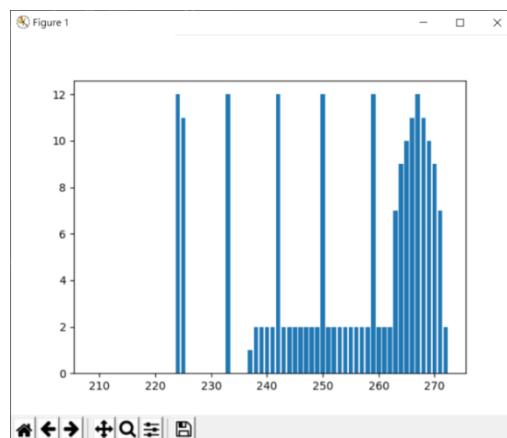


圖 4.3.1-8 中央 C 投影圖

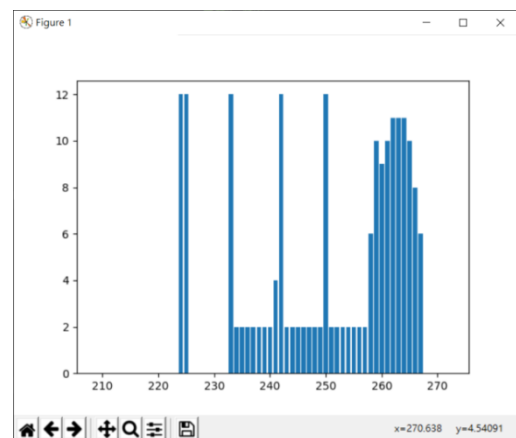


圖 4.3.1-9 中央 D 投影圖

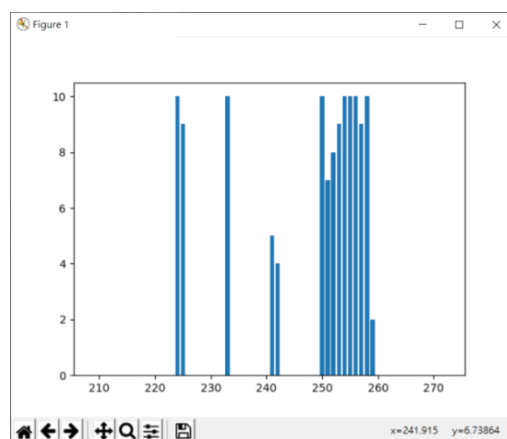


圖 4.3.1-10 中央 F 投影圖

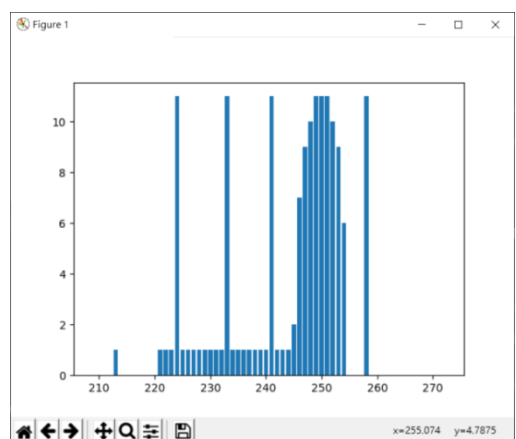


圖 4.3.1-11 中央 G 投影圖

如下圖 4.3.1-12 和圖 4.3.1-13，可以由像素點分布的位置觀察到這兩張圖都是中央 E 的投影，另外，可由像素點的多寡判斷圖 4.3.1-12 為四分音符，圖 4.3.1-13 為二分音符。

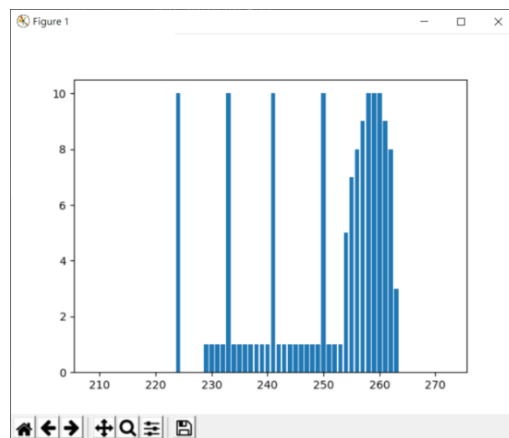


圖 4.3.1-12 四分音符投影圖

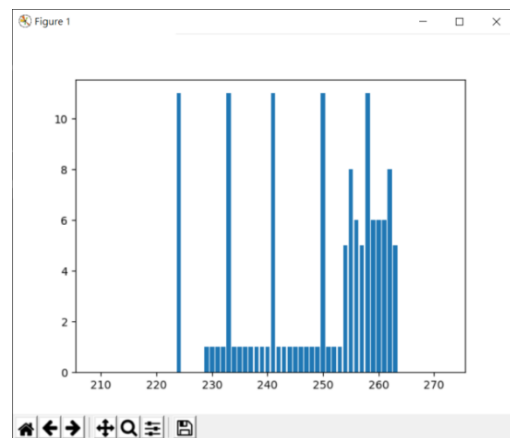


圖 4.3.1-13 二分音符投影圖

所以使用計算面積比例的方式，將所有讀到的音符分成實心及空心，實心的為四分音符，空心的為二分音符或四分音符。

以 OpenCV 的 drawRectangle() 函式將讀取到的音符框起來，以利後續使用這些物件的座標判斷音高，讀取到的音符示意圖如下圖 4.3.1-14，



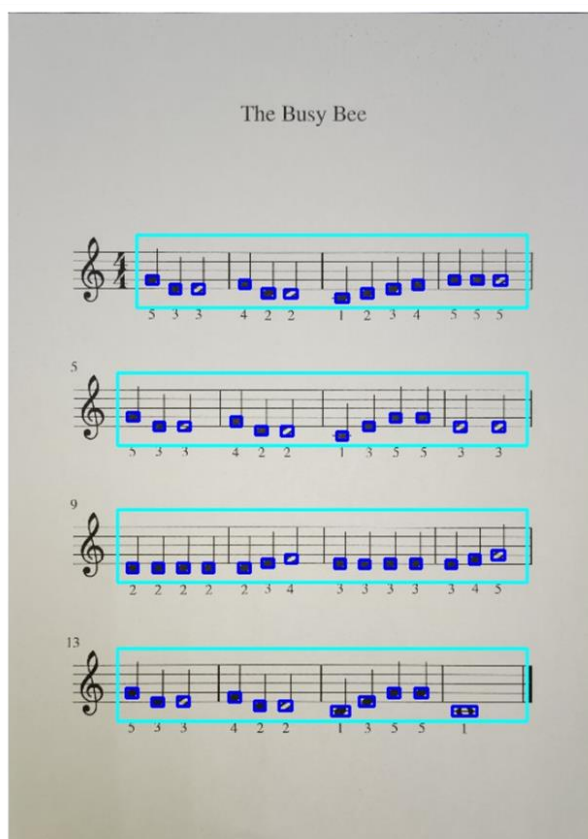


圖 4.3.1-14 讀取到的五線譜及音符

另外，空心的部分再另外確認有無符桿，判斷是全音符或二分音符。

接著以音符與五線譜的相對位置判斷音符的音高。

並將判斷到的音符資訊以 byte 的資訊儲存，如下圖，並產生 midi 檔供使用者於 Telegram 上播放。



圖 4.3.1-15 音符 bytes 資訊檔

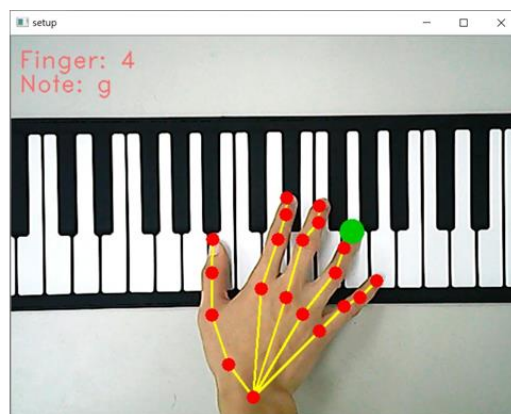
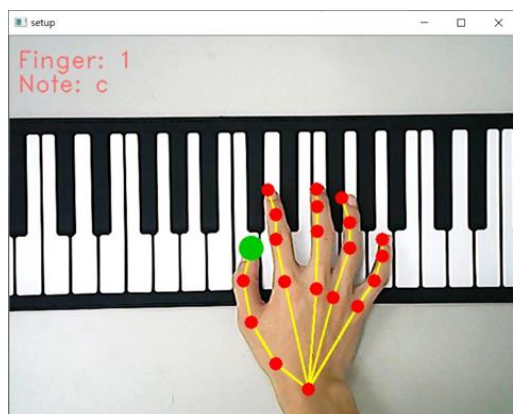
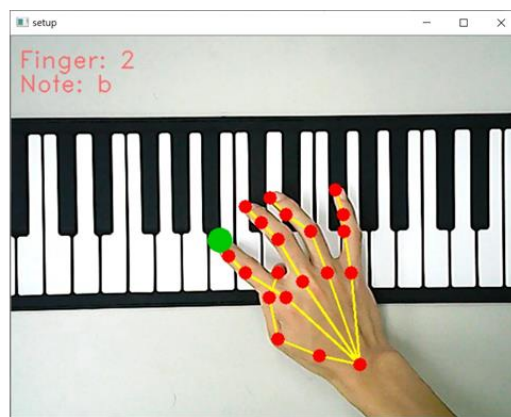
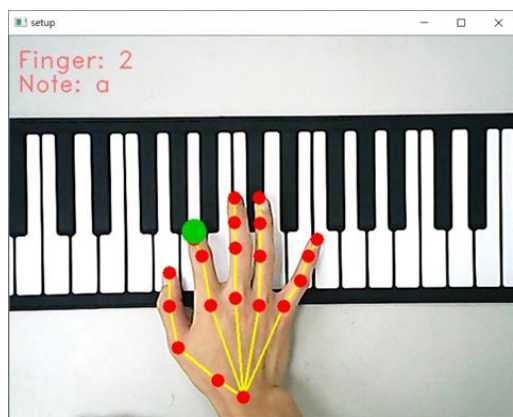
### (3) 數字指法辨識

樂譜上的指法總是會對到該音符之位置，因此取得音符位置後，我們同時也能取得數字指法的位置，將其擷取之後，輸入至以機器學習訓練完成的模型進行辨識，最終得到指法結果。



## 2. 手指指法辨識

彈奏評分系統配置彩色攝影機與 midi 訊號輸入，使用者按下琴鍵時，系統會接收到 midi 訊號的輸入，與此同時，攝影機會擷取彈奏當下的影像。影像輸入系統後，首先會針對手指進行辨識，我們使用卡內基梅隆大學（CMU）於 2017 年提出的模型處理方式《Hand Keypoint Detection in Single Images using Multiview Bootstrapping》。接下來會辨識各個琴鍵位置，將 midi 訊號與手指辨識結果位置進行比對，方能取得使用者所使用的指法與音準資訊。



### (三) 使用者介面

#### 1. Telegram 介面

##### (1) 樂譜辨識

第一階段先透過 Telegram 將圖片匯入，如下圖 4.3.3-1 所示，系統經過雲端處理後會回傳校正後的樂譜，接著使用者必須輸入曲名。接著 Telegram Bot 會回傳樂譜經過辨識後自動產生的音檔，如下圖 4.3.3-2 所示，並連同歌曲名稱、樂譜及音檔一併儲存在雲端，以利後續查詢及其他功能的使用。

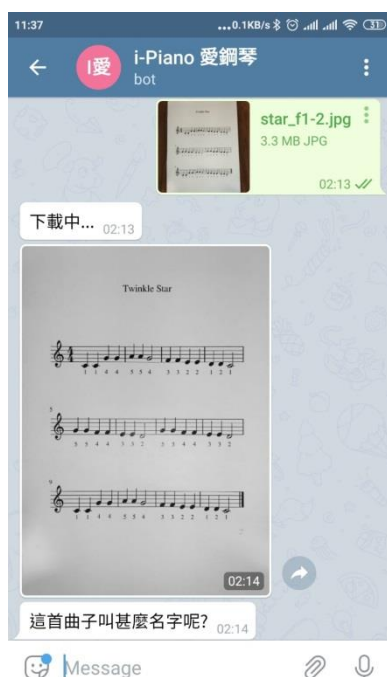


圖 4.3.3-1 圖片上傳

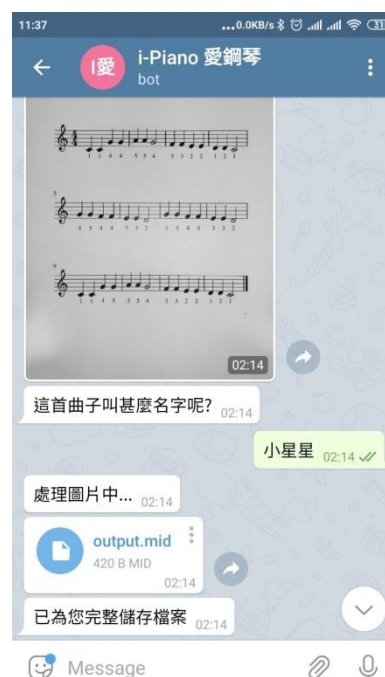


圖 4.3.3-2 圖片處理並回傳

##### (2) 清單查詢

建立雲端資料後我們便可以透過 Telegram Bot 查詢自己曾經上傳的樂譜清單，以便我們快速了解資料庫中目前有哪些樂譜，如下圖 4.3.3-3 所示。



圖 4.3.3-3 清單查詢

### (3) 樂曲查詢

使用者可以透過樂曲查詢的功能查詢清單內的曲目，如下圖 4.3.3-4，接著 Telegram Bot 會回傳跟此曲目相關的資訊，包括樂譜及音檔，如下圖 4.3.3-5 所示。



圖 4.3.3-4 樂曲查詢



圖 4.3.3-5 樂曲查詢結果

#### (4) 分數查詢

使用者除了可以透過查詢清單內的曲目，還可以查詢關於這些曲目的分數紀錄，如下圖 4.3.3-6，接著 Telegram Bot 會回傳使用者自己的歷史分數紀錄以及其他用戶的排名，如下圖 4.3.3-7 所示。



圖 4.3.3-6 分數查詢



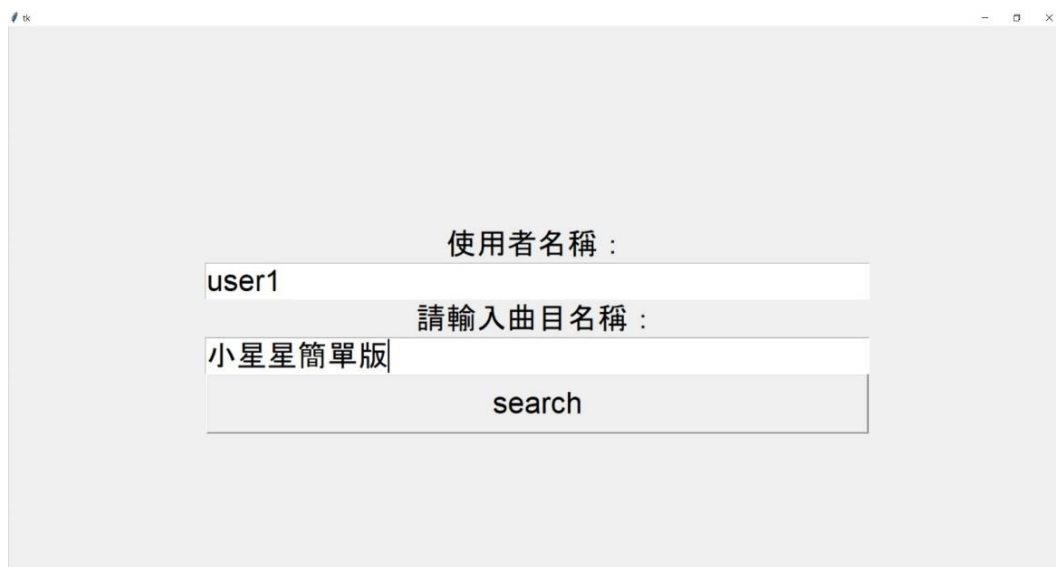
圖 4.3.3-7 分數查詢結果

## 2. 彈奏評分介面

彈奏評分介面需連結 RGB 攝影機及電子鋼琴的音訊輸出，並於電腦介面上操作。

### (1) 設定使用者資訊及曲名

使用者必須輸入用戶名稱及曲目，如圖 4.3.3-8，系統才能開始運作。



使用者名稱 :

user1

請輸入曲目名稱 :

小星星簡單版

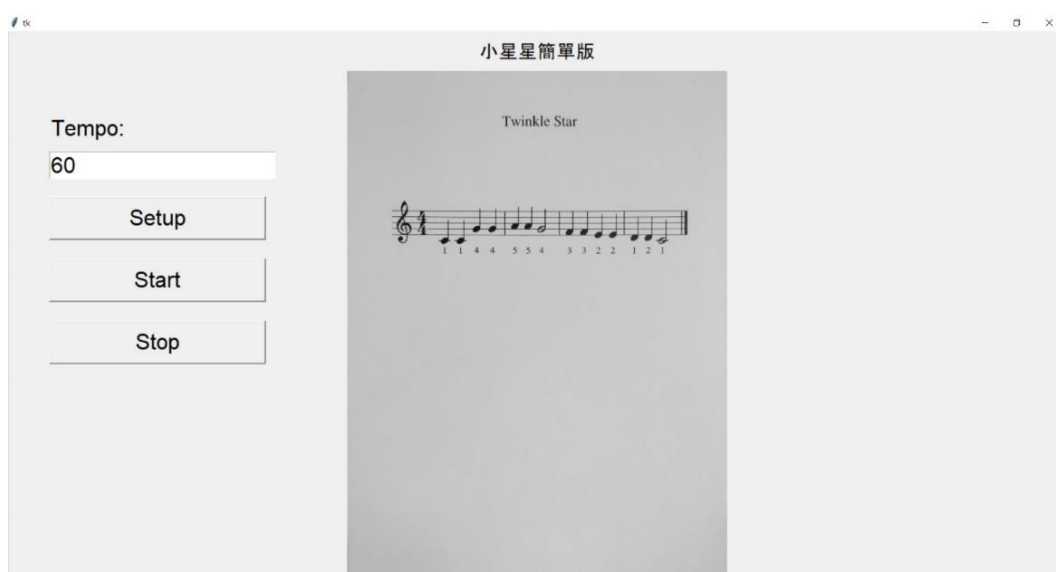
search

圖 4.3.3-8 使用者名稱、曲目輸入介面

## (2) 設定節奏

輸入使用者名稱及曲目後，系統會自動搜尋樂譜清單裡的樂譜，如下圖 4.3.3-9 所示，接著可以在左邊的欄位中設定速度。

速度的部份，現代習慣以每分鐘多少拍作為量度單位，下圖 4.3.3-9 的 60 即表示一分鐘 60 拍。



小星星簡單版

Tempo:

60

Setup

Start

Stop

Twinkle Star

1 1 4 4 5 5 4 3 3 2 2 1 2 1

圖 4.3.3-9 顯示樂譜、設定節奏

### (3) 彈奏指法辨識

設定節奏後，介面上會出現節拍器，使用者必須跟著節奏練習。系統會在每次收到 MIDI 訊號輸入時判斷當時使用的指法以及是否符合當時被按下的音，如下圖 4.3.3-10 所示。

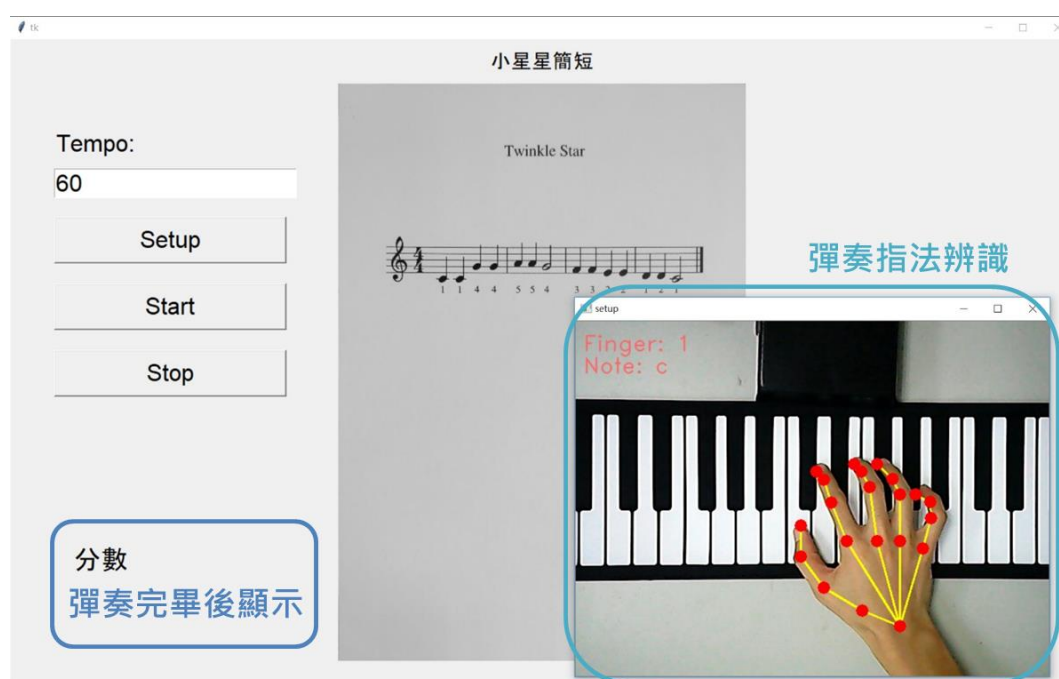


圖 4.3.3-10 彈奏指法辨識

### (4) 評分結果

彈奏結束後，分數將會出現在左下角，如上圖 4.3.3-10 所示。

評分機制如下表 4.3.3 所示，評分的项目包括音準、指法的正確與否以及拍子及音長的準確度。

計分的部份除了音準和指法是以正確或錯誤來判斷之外，其他的計分設計都有彈性空間。



評分機制	標準	得分	標準	得分
音準	正確	1	錯誤	0
指法	正確	1	錯誤	0
拍子	在 $\pm 25\%$ 內	1	在 $\pm 25 \sim \pm 5\%$ 間	0.5
音長	不到 0.5%	0.5	0.5 ~ 1 %	1

表 4.3.3 評分機制說明

每個音符的滿分是四分，若得分為 3.5 分或是 4 分就不會在譜上出現記號，若是 3 分就會以最淺的顏色表示，2 分或是 2.5 分會以稍微深一點的珊瑚色表示，2 分以下則以更深的紅色表示。以同一色系的深淺色來標記可以更直觀地讓使用者理解自己需要加強的部份。

實際彈奏的結果如下圖 4.3.3-11 所示。

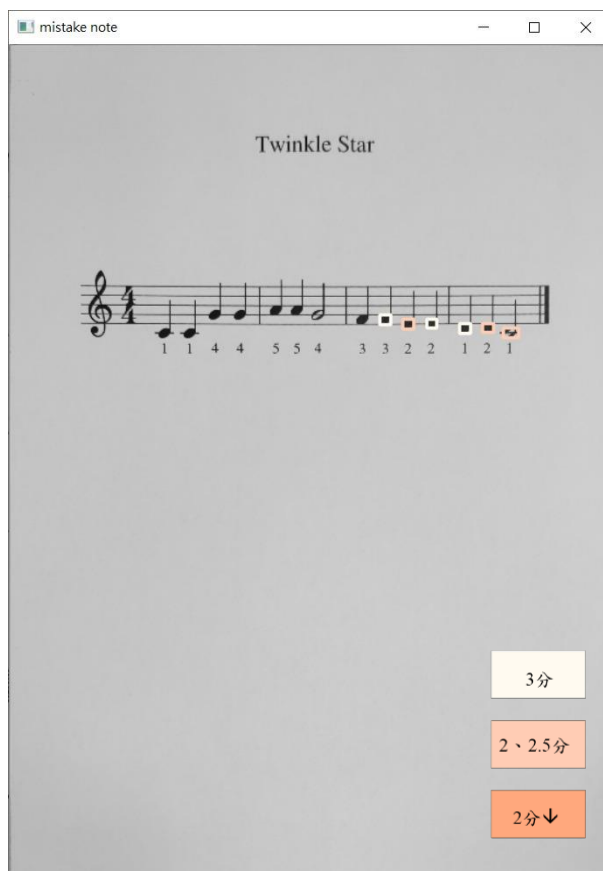


圖 4.3.3-11 評分結果

另外，系統同樣會將評分的数据圖像化，如下圖 4.3.3-12，可以分別看到使用者該曲目在音準、指法、拍子及音長所得到的分數。

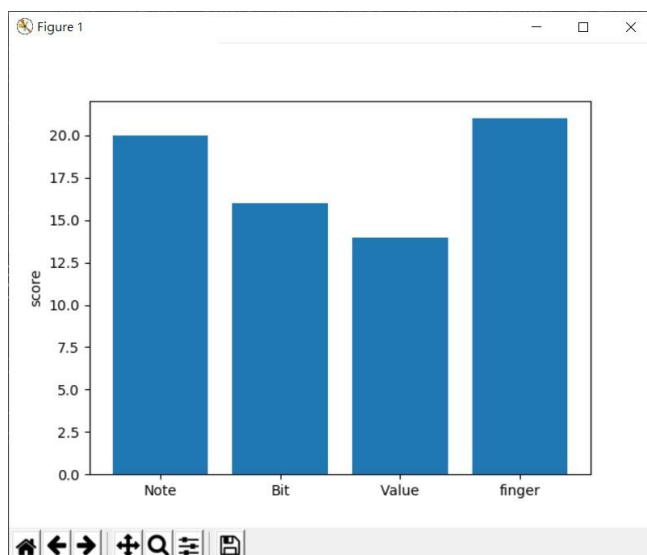


圖 4.3.3-12 得分統計

### 3. 實驗數據

#### (1) 樂譜辨識

樂譜辨識測試	行數	音符種類			音準	指法
		四分音符	二分音符	全音符		
樣本1	1/1	12/12	2/2	0/0	14/14	14/14
樣本2	3/3	36/36	6/6	0/0	42/42	41/42
樣本3	4/4	38/38	9/11	1/1	47/50	48/50
樣本4	3/4	30/44	3/6	2/2	38/52	52/52
樣本5	4/4	43/44	6/6	2/2	51/52	51/52
樣本6	4/4	44/44	8/8	1/1	53/53	52/53
樣本7	4/4	41/42	6/7	2/2	45/51	50/51
樣本8	4/4	43/44	6/8	1/1	50/53	52/53
樣本9	3/3	31/32	4/4	2/2	37/38	38/38
樣本10	3/3	32/34	4/5	1/1	36/40	40/40
樣本11	4/4	36/36	7/8	2/3	45/47	47/47
樣本12	4/4	48/48	7/8	0/0	54/56	55/56
樣本13	4/4	47/48	5/6	1/1	55/56	55/56
樣本14	3/4	36/48	2/6	1/1	39/56	54/56
樣本15	4/4	46/48	7/8	0/0	53/56	55/56
準確率(%)	97.83	94.15	82.83	94.12	92.04	98.19

#### (2) 指法辨識

彈奏指法識別	樣本1	樣本2	樣本3	樣本4	樣本5	樣本6	樣本7
實驗1	14/14	42/42	49/50	51/52	52/52	53/53	50/51
實驗2	14/14	42/42	50/50	50/52	51/52	52/53	50/51
準確率(%)	100	100	99	97.11	99.03	99.06	98.04
彈奏指法識別	樣本8	樣本9	樣本10	樣本11	樣本12	樣本13	準 確 率 (%)
實驗1	52/53	37/38	40/40	47/47	56/56	55/56	
實驗2	53/53	38/38	40/40	46/47	55/56	56/56	
準確率(%)	99.06	98.68	100	98.94	99.12	99.12	

## 五、結論

### （一）結論說明

近年來各項影像追蹤技術的發展，使得機器可以取得比起以往更多的資訊，可以應用於更多元的領域上。在學習教育方面，利用機器來輔助學習的相關產品也不斷的增加。而我們設計的「具樂譜分析與演奏評分機制之鋼琴練習系統」提供便利獲取樂譜資訊、MIDI 檔的方式，並可記錄、評分使用者的練習情況，且透過平台以圖形化呈現的方式提供使用者可直覺的察看，使其更加了解自身的學習現況，可針對學習力較弱的部分加強練習。

本專題所建構的練習系統適用於各年齡層之初階、中階的鋼琴學習者，提供單人的學習及練習模式以及一個健全的使用者資料的彙整平台，使用者除了可針對同一樂曲察看其他使用者的學習狀況，也可以針對一段時間或同一首曲子了解自己的練習狀況，更可與同儕、家人間以競賽的方式彼此互相激勵學習，提升學習效率。

除此之外，使用者不需攜帶厚重的紙本樂譜，僅需將樂譜拍下並透過此系統轉換，即可得到經過校正後的數位樂譜，並產生可於任意環境下重複播放及使用的 MIDI 檔。而若想更進一步學習或訓練時，則可架設攝影機，搭配彈奏判別系統，判斷音準、指法是否正確，並將練習資料進行紀錄及評分。

本系統亦提供人性化的操作介面，可以用通訊軟體 Telegram 將樂譜上傳至資料庫，並透過 Telegram Bot 查詢樂譜資訊及分數資料等，同時也可以在此介面查詢樂譜並播放樂譜所產生的音檔；另外的彈奏評分介面則提供了使用者在彈奏後查看整首樂曲的評分，更可以了解每個音符所得到的分數及每個評分項目所獲得的得分資訊，讓使用者可更直觀且容易地找出弱點的部分，並針對該部分進行加強練習。

## （二）未來展望

### 1. 樂譜跨頁問題

本系統使用了影像處理技術將樂譜資訊以數位方式儲存，並將樂譜資訊與使用者所彈奏產生 MIDI 檔案比對，達到分析與紀錄使用者彈奏狀況的目的。未來可即時將 MIDI 檔案與樂譜資訊與樂譜資訊比對，使用者當下即可知道錯誤，並且可以透過音符追蹤的方式完成自動翻頁的功能。

### 2. 提高樂譜辨識率，增加音樂術語辨識及查詢

目前本系統可以辨識的部份是四分音符、二分音符、全音符及阿拉伯數字，八分音符及升降記號的部份尚未齊全，希望後續能用更進階的影像處理技術，處理更複雜的樂譜。除此之外，樂譜上的音樂術語大多由英文字母組成，若能建立資料庫，並辨識樂譜上的音樂術語，讓使用者不需經過查詢即能在樂譜上顯示意思，可以更直觀地提醒使用者，達到提醒使用者注意表情符號的目的。

### 3. 使用者介面改良

本專題所使用的介面分別有兩個，分別是 Telegram Bot 與彈奏評分介面，若能整合成同一介面，應該可以提供使用者更佳的使用者體驗。

## 六、致謝

為期一年的專題研究順利結束，感謝一路上幫助我們的教授、學長姐以及同學，若沒有這些人的協助，就沒有今天的研究成果。

首先感謝我們的專題教授 黃正民教授，在研究初期，和我們討論研究方向。對於沒有先備知識的我們，仍然耐心地從基礎知識開始教導。每週一次的開會督促專題的研究進度，可以感受到教授對學生的重視。黃正民教授提出的建議總是能指引我們正確的方向，在專題遇到瓶頸與困難時，幫助我們突破困境。另外，也很感謝實驗室的何信鴻學長與王佩琪學姊，除了技術指導外，也在研究報告撰寫上給予很多幫助。

最後，感謝北科大電機工程系提供我們如此完善的環境與制度，讓我們能夠磨練實作能力與研究方法，不論未來選擇進入職場，亦或是繼續進修於研究所，相信對我們都有很大的幫助。

## 七、參考文獻

[1] 音樂記譜法

<https://zh.wikipedia.org/wiki/%E9%9F%B3%E6%A8%82%E8%A8%98%E8%AD%9C%E6%B3%95>

[2] 樂器演奏中音符與節奏的辨識。薛正本（2000）。國立成功大學數學系。

<https://hdl.handle.net/11296/hh563q>

[3] 樂譜自動辨識之電腦視覺系統設計。林文靖（2001）。國立中正大學電機工程研究所。<https://hdl.handle.net/11296/t4fph5>

[4] 應用影像處理之自動樂譜辨識系統。彭鐘賢（2014）。中原大學資訊工程研究所。<https://hdl.handle.net/11296/a5asc8>

[5] OpenCV 相關資料

<https://opencv.org/about.html>

[http://monkeycoding.com/?page\\_id=12](http://monkeycoding.com/?page_id=12)

<https://zh.wikipedia.org/wiki/OpenCV>

<https://blog.csdn.net/on2way/article/details/46812121>

[6] TensorFlow 相關資料

<https://www.tensorflow.org/>

[7] Python 相關資料

<https://www.python.org/downloads/>

<https://morvanzhou.github.io/tutorials/python-basic/basic/>

[8] MIDI 相關資料

<https://zh.wikipedia.org/wiki/MIDI>

[9] 聊天機器人相關資料

<https://zh.wikipedia.org/wiki/%E8%81%8A%E5%A4%A9%E6%A9%9F%E5%99%A8%E4%BA%BA>

<https://searchdomino.techtarget.com/definition/IM-bot>

<https://gizmodo.com/google-assistant-is-a-mega-chatbot-that-wants-to-be-abs-1777351140>

<https://www.smartm.com.tw/article/34383130cea3>

<https://zi.media/@sharingtcincubatorcom/post/3FPRFc>

[10] Telegeam 相關資料

<https://zh.wikipedia.org/wiki/Telegram>

<https://telegram.org/blog/6-years>

[11] imgur 相關資料

<https://zh.wikipedia.org/wiki/Imgur>

<https://www.alexa.com/siteinfo/imgur.com>

[12] Dropbox 相關資料

<https://zh.wikipedia.org/wiki/Dropbox>

<https://help.dropbox.com/accounts-billing/security/physical-location-data-storage>

[11] 圖像幾何變換相關資料

<https://blog.csdn.net/on2way/article/details/46801063>

[https://blog.csdn.net/xiaowei\\_cqu/article/details/26478135](https://blog.csdn.net/xiaowei_cqu/article/details/26478135)

[12] 模板匹配相關資料

<https://www.itread01.com/content/1544283140.html>

[13] 投影法相關資料

<https://www.bilibili.com/read/cv1231639/>

<https://zhuanlan.zhihu.com/p/75351673>

[14] 圖像幾何變換相關資料

<https://blog.csdn.net/on2way/article/details/46801063>