

第二十三屆旺宏科學獎

成果報告書

參賽編號：SA23-682

隊伍名稱：以多任務學習 MTL 深度學習網路
建構識譜吹笛機器人

作品名稱：以多任務學習 MTL 深度學習網路
建構識譜吹笛機器人

參賽類別：電子電機

關鍵字：吹笛機器人、多任務學習、樂譜辨識

摘要

本研究目標在建構可以識譜及吹奏中國笛的吹笛機器人。中國笛演奏必須協調吹法及指法；藉由控制吹氣流速、吹嘴角度及六指按壓音孔的變化來控制音高以完美地吹奏樂曲，是一項複雜的演奏技術。機器人以模擬吹笛口型的吹嘴，搭配兩個風箱往復送氣到一個壓力調節風箱送氣，以微控制板控制六個機械手指來蓋放完成演奏，為在音尾可確實止氣，設計一個風門，利用風門開闔也可模仿吐音技巧讓笛聲明確發音。辨識樂譜方面收集樂譜樣本，樣本分成譜線、音符、節奏三套，透過多任務學習 MTL 的深度學習架構進行訓練，建構可以辨識五線四間及二、四、八分音符的樂譜辨識模型。經測試若樂譜在符合音域範圍內，可以完整的辨識，轉換成音符資料傳送給吹笛機器人吹奏。

壹、前言

一、研究動機

在看一段有關 2008 年早稻田大學所開發的吹笛機器人 Waseda Flutist Robot 影片，如圖 1，以快速的音階變化吹奏大黃蜂的長笛曲，引發研究開發吹笛機器人的構想。經請教學校懂得演奏中國笛的老師後，透過分析笛子演奏的技巧，吹笛者必須不斷控制吹氣速度達到正確音準。最後歸納需要使用的器官有口、胸、腹、指、舌、耳、眼，七大器官複雜的協調，操作笛子的一個吹孔及指孔，發出強弱不一的長、短、頓及連音，構成各種美妙的樂曲，具有極高的複雜性。為使機器人可以成為一個完整的吹奏家，在確實可以建構一部可吹氣、變換指法的吹笛機器人後，更進一步讓機器人識譜吹奏，成為一個完整的機器

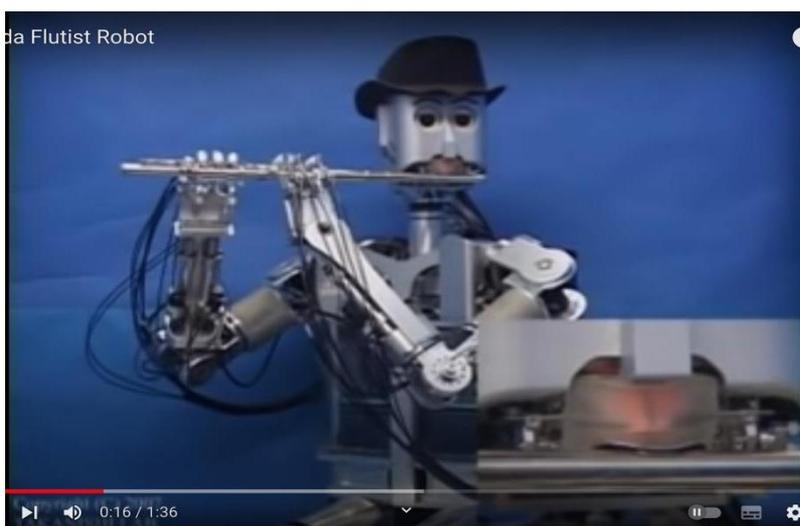


圖 1：吹笛機器人的 YouTube 影片
(資料來源：BotJunkie (2008)，Waseda Flutist Robot)

人。綜合上述，要完成一部演奏中國竹笛「吹笛機器人」，必須從吹笛技巧、樂理、電子電機控制各方面相關有所理解並可以應用外，識譜方面更需要影像處理、人工智慧方面的應用，是一項挑戰性極高的研究題材。

為了讓吹笛機器人具有識譜能力，參閱有關樂譜辨識的博碩士論文研究，多以影像處理及數學進行辨識音高及音長，近年多以神經網路技術建構辨識模型，基於樂譜辨識的多元性，本研究嘗試使用多任務學習的方法，建構樂譜的辨識模型。

表 1：與過去樂譜辨識相關之作品比較

(資料來源：廖威凱(2007)。基於類神經網路及樂理知識限制之印刷樂譜辨識系統。李豪韋(2016)。自動樂譜辨識。蔡自偉(2004)。印刷樂譜辨識系統。)

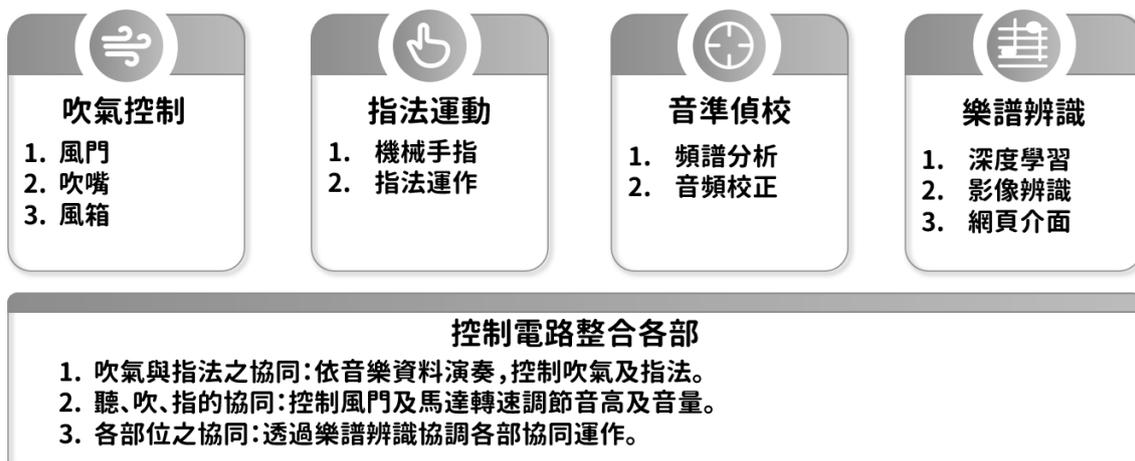
年代	作品名稱	簡介
2007	基於類神經網路及樂理知識限制之印刷樂譜辨識系統	透過結合神經網路與樂理知識自動辨識樂譜中的音樂符號
2016	自動樂譜辨識	透過樂譜前處理以及對於單一組五線譜的辨識演算
2024	印刷樂譜辨識系統	透過數學方式計算五線譜即個音符的細部組件推理音符並轉呈 MIDI 音檔

二、研究目的

吹奏時，必須以正確的「口」型放置在正確的吹孔位置，配合「胸」、「腹」控制出氣流速及流量，以發出準確的音高及音量，藉由手「指」操作六個音孔，配合氣流控制操作音階。配合「耳」的聽覺判別音色音準是否正確，再使用「口」、「胸」、「腹」搭配調整。「舌」可以操作吐音，凸顯樂句的頓、連，使用旋律結構更加豐富及明確。樂譜透過「眼」來識別音符、音高、速度、音調、強弱、表情技巧...等，執行上述各部門的協同動作，搭配竹笛的獨特設計的笛膜設計，發出清脆優美音色。

綜上各種技術的實現及理論的探究，歸納有吹、指、聽、眼四大部門進行研究，研究之目的如下：

表 2：整體研究架構圖 (資料來源：研究者自行製作)



三、文獻探討

(一) 竹笛的構造

中國笛主要的構造為吹孔及六個按音孔，可透過按壓音孔及調整吹氣的氣流及角度，來改變音量及音色。吹孔及按音孔間另外開一個膜孔、另外還有後出音孔及前出音孔各兩孔如圖 2。

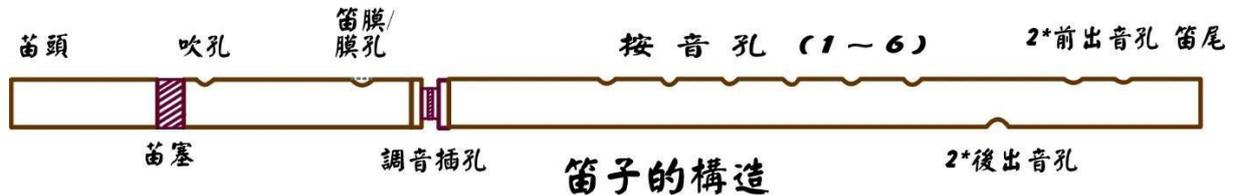


圖 2：竹笛的構造 (圖片來源：研究者自行繪製)

(二) 竹笛的吹奏技巧

1. 吹法：

- (1) 發音技巧：笛子使用嘴唇自然的閉合，上下嘴唇以念注音符號口的感覺，露出一個扁長的孔隙，稱之為風門，將風門靠近吹孔的外緣輕輕送氣便可以發出聲音。

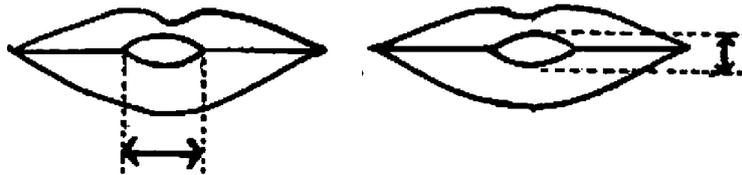


圖 3：上下嘴唇間形成的風門的形狀及大小

(資料來源：陳中申(1985)。玉笛春風)

- (2) 送氣技巧：吹奏笛子重視長音，一般來說一個長音至少要持續 8 秒以上，才可以應付大部分的曲子。為了容納更多的氣體容量，吹笛者會練習腹式呼吸法，也就是將氣體往橫隔膜的方向下存放，再藉由控制橫隔膜自然向上推送，將氣緩緩送出。
- (3) 高低音控制：六孔竹笛一般都可以吹出兩個八度的音，技巧上低音吹氣流速緩，高音流速快。流速的控制可以用氣壓強弱或變化口腔空間來控制。
- (4) 音量的控制：吐氣流速可以控制音高，音量則可以控制風門大小來控制聲音大小，也就是說，流量大音量就大。

2. 指法

- (1) 音階：竹笛六個音孔如果以按放全孔吹奏，可以滿足大部份演奏曲的演奏，音階如下表所示。笛子的調性會以全按四五六孔的音來定調，不管是 C、D 或 G 調笛，主音也就是唱名為 Do 的音，都是按四五六孔。詳細的指法如表 4。

表 3：竹笛的簡易音階表(資料來源：研究者自行繪製)

唱名	簡譜	音名			按音孔					
		C 調笛	D 調笛	G 調笛	六	五	四	三	二	一
Do	1	C	D	G	●	●	●			
Re	2	D	E	A	●	●				
Me	3	E	F#	B	●					
Fa	4	F	G	C		●	●			
Sol	5	G	A	D	●	●	●	●	●	●
La	6	A	B	E	●	●	●	●	●	
Ti	7	B	C#	F#	●	●	●	●		

- (2) 轉調：六孔竹笛無法完全表現 12 平均律的 7 個全音及 5 個半音，導致轉調上的困難。為了表現不同的聲調，會製作各種不同聲調的笛子，也就是透過換笛子來轉調。比如說 D 調轉 G 調，可以從 D 調笛換成 G 調笛。看似麻煩，但因為轉調指法不變，所以也有方便之處。

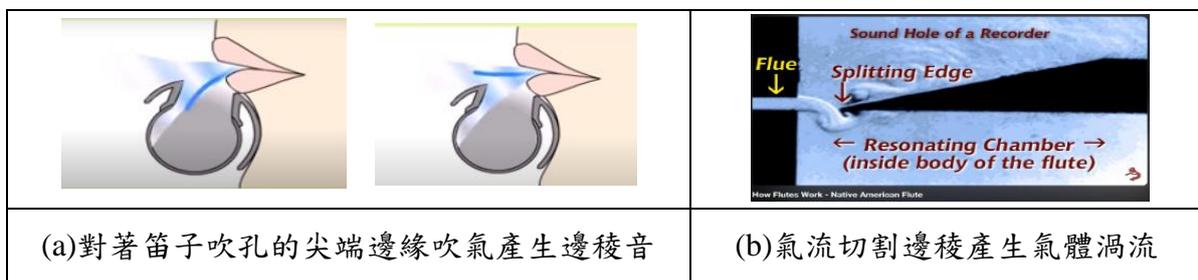
(三) 竹笛的物理原理

1. 邊稜音 (Edge Tone) 原理：笛、蕭都是以邊稜音發音的樂器，當吹奏者貼近吹孔，並對準吹孔另外一邊尖銳邊緣吹氣，氣流會在尖錐上下形成上下兩個氣體渦流 (vortex)，形成邊稜音。邊稜音聲音細小而音高，雖並非笛子主要的音色，但卻因此觸發笛子空氣柱共振產生聲音

表 4：笛子的邊稜音原理

(資料來源(a)：Yamaha Japan(2013)，【フルート】フルートの発音原理。)

(資料來源(b)：Clint Goss (2015), **How Flutes Work - Native American Flute.**)



2. 空氣柱(Air Column)共振：當吹奏者對著吹孔邊緣吹氣，引發邊稜音促使空氣柱內產生振動。按壓竹笛按音孔的位置以改變內部空氣柱的長度。空氣柱越短聲音頻率較高，空氣柱越長聲音頻率較低。

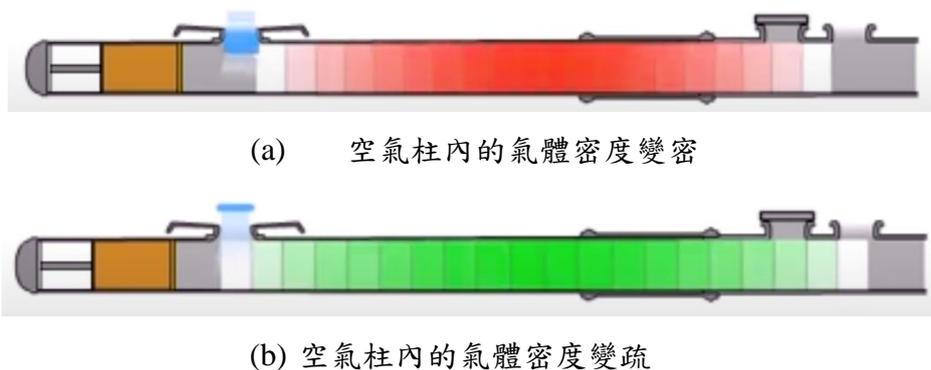


圖 4：邊稜音影響空氣柱內的氣體做疏密的振動

(資料來源：Yamaha Japan(2013)，【フルート】フルートの発音原理。)

3. 泛音(Overtone)：吹奏笛子在同一空氣柱長度下，透過邊稜音振動空氣柱產生出基本頻率，稱之為基音；若空氣柱長度不變，透過加壓提高送入的氣體流速，則此時產生的高頻聲音，稱作泛音。在同一空氣柱長度下，泛音為基音的 2 倍稱為第一泛音；若頻率基頻的 3 倍，則稱之為第二泛音。

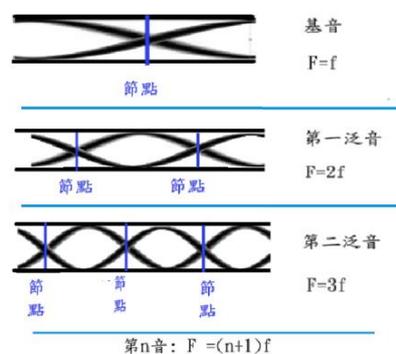


圖 5：泛音列

(資料來源：研究者自行繪製)

(四) 五線譜記譜法

樂譜上的五條水平線，用於標示不同的音高。每一條線和每一條線之間的空白都代表特定的音高。

1. 譜號：

表 5：譜號表 (資料來源：研究者自行繪製)

高音譜號 (G 譜號)	低音譜號 (F 譜號)	中音譜號 (C 譜號)

2. 音符：全音符、二分音符、四分音符、八分音符及 16 分音符等。

表 6：各種音符(資料來源：研究者自行繪製)

全音符	2 分音符	4 分音符	8 分音符	16 分音符
				
4 拍	2 拍	1 拍	1/2 拍	1/4 拍

3. 休止符：全休止符、二分休止符、四分休止符、八分休止符及 16 分休止符等。

表 7：各種休止符(資料來源：研究者自行繪製)

全休止符	2 分休止符	4 分休止符	8 分休止符	16 分休止符
				
4 拍	2 拍	1 拍	1/2 拍	1/4 拍

4. 升降記號：

- (1) 升號 (#)：將音提高半音。
- (2) 降號 (b)：將音降低半音。
- (3) 還原記號 (♮)：取消之前的升號或降號的效果。

5. 拍號：表示每小節中的拍子數和每拍的音符時值。

6. 速度：音樂速度，如 Largo (緩慢)、Allegro (快板)、Presto (急板) 等。

7. 強弱：音量強弱，如 p (弱)、f (強)、mf (中強)、pp (很弱)、ff (很強) 等。

(五) 多任務學習 (Multi-task Learning, MTL)

多任務學習是機器學習的一個子領域，將任務中學到的知識重複利用來提高模型效能，把相關任務的訓練訊號包含的資訊作為歸納偏置 (Inductive bias) 來提高模型泛化能力。目的是把多個相關任務放在一起學習，透過任務之間資訊共享來使模型能夠抵抗雜訊干擾，保持穩定輸出，並降低過擬合風險，與單任務模型相比，多任務學習提升了數據利用效率，在多任務的數據共同作用下，提供了更加廣泛的訓練信息，並在數據稀疏的任務中顯著提升模型表現。

為使任務之間資訊共享更加高效，典型的 MTL 模型架構包括以下內容：

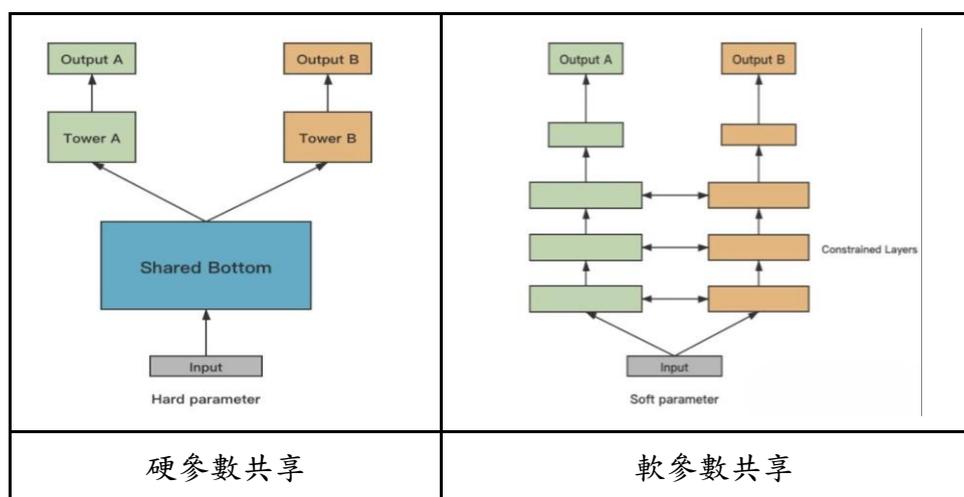
1. **共享底層網路(Shared Bottom Architecture)**：任務之間共享底層網路，並在頂層使用不同任務專用層，透過共享底層特徵來提高任務間的知識共享。

2. **分支網路(Branched Architecture)**：在模型某一層之後將分成多個分支，每個分支負責不同的任務，確保共享資訊的同時也允許任務特定的表示學習。
3. **交叉共享網路(Cross-Stitch Networks)**：在網路不同層間引入交叉共享單元，動態調整任務間的資訊共享程度。

基於上述模型架構特性，多任務學習的架構分為硬參數共享 (Hard parameter sharing) 及軟參數共享 (Soft parameter sharing)，如下表。多任務學習透過使用相關任務訓練的特徵資訊來提高模型泛化能力。硬參數共享大幅降低過擬合風險；軟參數共享使每個任務擁有獨立模型參數，並對參數間的距離進行正則化以鼓勵不同任務的參數變得相似從而促進任務之間的參數共享。

表 8：多任務學習之架構（資料來源：marketer(2021)。

當我們在談論 **Multi-Task Learning(多任務/多目標學習)**



多任務學習相對於多個單任務學習，以實用角度來看具有以下優點：

1. 模型可以在效果不下降情況下提升多任務預測效率，並共享部分網路結構，大幅減少模型參數總量，且對於某個任務而言，MTL 模型參數(共享參數和任務獨立參數)與單任務模型參數量一樣，因此模型理論表達能力並未降低。
2. MTL 可以提升小樣本任務訓練效果，因為彼此參數共享會有效降低過擬合問題，使個任務間充分學習。
3. MTL 引入 Inductive Bias 提升各任務的穩健性，對於 MTL，Inductive Bias 約束了各任務之間的效果，若約束合理，各任務都將更不容易受噪聲樣本影響，進一步提升模型穩健性。

在一張樂譜上可能有上百個需辨識物件，分類也高達 100 種以上，對於單獨模型來說訓練顯然過於困難，不僅數據集要足夠充分，還因為資訊範圍廣而導致辨識有疏漏。將

辨識樂譜的過程分開之後形成的多任務模型，減緩了單模型辨識壓力，彌補數據不足所產生疏漏、過擬合。各任務都將放大樂譜上的物件特徵，並彼此交疊，確保了大尺度上的完整，小尺度上的精確。

多任務學習上假設有 N 個任務，其中 $N \in \{T_1, T_2, \dots, T_N\}$ ，每個任務 T_i 都會有自己的損失函數 $L_i(\theta)$ ，多任務學習的目標是將學習共享的參數(或模型的特定參數 θ_i)，使所有任務的損失總和最小化，目標函數為：

$$\min (\theta, \theta_1, \theta_2, \dots, \theta_N) = \sum_{i=1}^N L_i(\theta, \theta_i)$$

θ 是共享參數， θ_i 是與任務 T_i 相關的特定參數。為了平衡不同任務的重要性，每個任務的損失函數會再加上權重 λ_i ，使多任務學習的目標函數為：

$$\min (\theta, \theta_1, \theta_2, \dots, \theta_N) \sum_{i=1}^N \lambda_i L_i(\theta, \theta_i)$$

並且多任務學習中時常與 L2 正則化一同使用，其中 α 與 β_i 是正則化參數。最終使其目標函數：

$$(\theta, \theta_1, \theta_2, \dots, \theta_N) \sum_{i=1}^N \lambda_i L_i(\theta, \theta_i) + \alpha \|\theta\|^2 + \sum_{i=1}^N \beta_i \|\theta_i\|^2$$

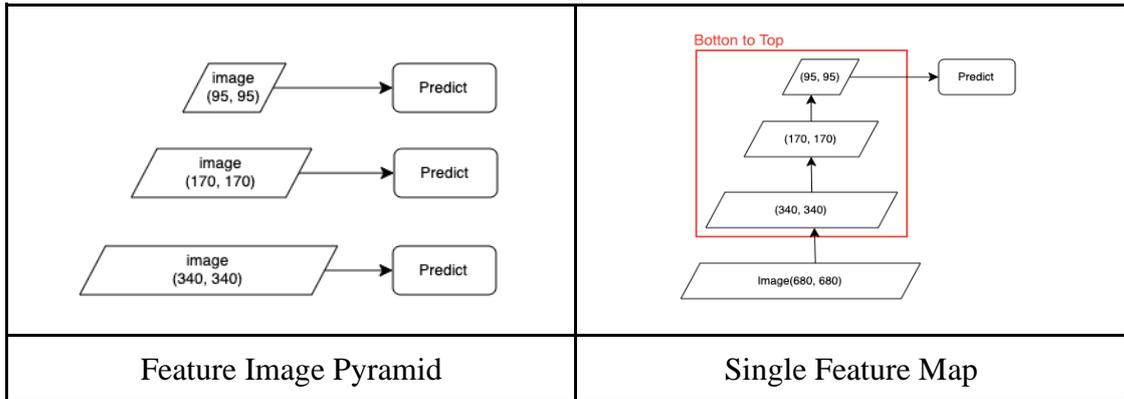
在多任務學習的優化過程中，總損失共享參數和任務特定參數 θ_i 的梯度計算為：

$$\frac{\partial L_{loss}}{\partial \theta} = \sum_{i=1}^N \lambda_i \frac{\partial L_i}{\partial \theta} , \quad \frac{\partial L_{total}}{\partial \theta_i} = \sum_{i=1}^N \lambda_i \frac{\partial L_i}{\partial \theta_i}$$

(六) 金字塔網路(Pyramid Network)

物件偵測應用上，最初都是使用 Feature Image Pyramid，當不同尺寸的影像被抽取特徵後，此方法多用在早期特徵仍以人工標記時代，但此作法等同重複做多次辨識工作，作業量增加好幾倍。單一特徵地圖 (Single Feature Map) 提出之前，在 ImageNet, COCO 等多個影像辨識比賽中熱門的卷積神經網路 (Convolution Neural Network, CNN) 或物件偵測模型皆按此設計，但這樣缺乏考量多尺度下的特徵影樣，因此競賽上依然使用 Feature Image Pyramid 的概念，這些網路都證明 Feature Image Pyramid 會有更良好的表現，但因為計算量大且耗時，因此一般應用下不太容易出現 Feature Image Pyramid。

表 9：Feature Image Pyramid 與 Single Feature Map (資料來源：研究者自行繪製)



透過 CNN 進行卷積以及降維所產生出的不同尺度特徵影像來做預測，類似於 Feature Image Pyramid，稱作 Pyramidal Feature Hierarchy，在 Single Shot Multibox Detector (SSD) 物件偵測模型中首先實現此架構，但並未使用到較高解析度的特徵影像，以高解析度特徵影像來說，特徵含義不高，但對於小物體檢測至關重要，因此 SSD 對此存在些許缺點。

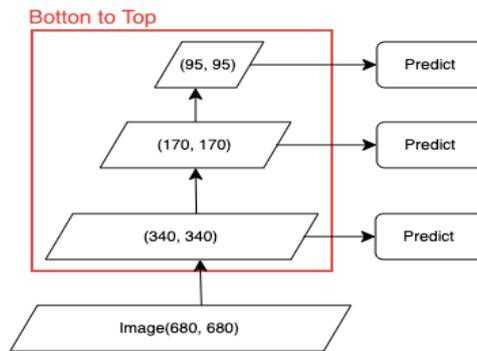


圖 6：Pyramidal Feature Hierarchy (資料來源：研究者自行繪製)

金字塔網路是一種多尺度特徵金字塔結構的神經網路，用於解決影像處理和計算機視覺中的多尺度問題，透過引入多層次特徵金字塔有效提取影像中不同尺度的信息，提升模型對各種尺寸目標的檢測和辨識能力。金字塔網路組成架構：

1. 以卷積神經網路為基底：例如殘差神經網路 (Residual Neural Network, ResNet)。
2. 包括自頂向下路徑：將高層特徵逐層上採樣並與低層特徵融合。
3. 橫向連接：將基礎網路中的特徵圖與上採樣的特徵圖進行融合。

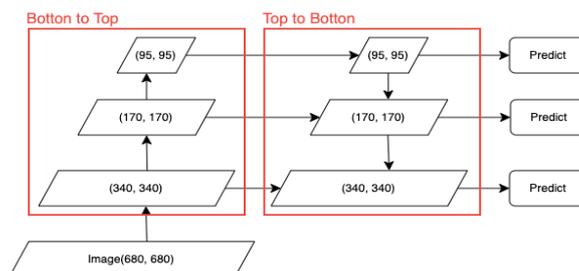


圖 7：Feature Pyramid Networks (資料來源：研究者自行繪製)

貳、研究設備及器材

一、使用設備及材料詳細

- (1) 微控制器：使用 ATMEGA328 晶片，控制直流馬達程進行氣體風箱的推送及六隻機械手指進行音階指法的按放工作，藉由程式讓送氣氣流、風門與指法可以互相配合，以改變音高並微調音準。



圖 8：ATMEGA328
(資料來源：(2014)。
atmega328-1)

- (2) 伺服馬達：風門使用 MG90S 的微型伺服馬達，伺服馬達的工作原理是利用 Pulse-Width Modulation (PWM)接腳來輸出脈波，透過輸出脈波之 Duty Cycle(工作週期)，控制伺服馬達要轉動的角度。

$$Duty\ Cycle\% = \frac{\text{工作時間}}{\text{一個週期的時間}} \times 100\%$$

運用控制伺服馬達轉動風門達到送氣或吐氣的動作。可使用微控制板的 analogWrite(x) 函式調整來工作週期，其中 $0 \leq x \leq 255$ 分別代表(0~100%)，透過調整 x 之數值改變工作週期，不同的工作週期會使轉動角度不同。

- (3) 電磁閥與矽膠手指：最新版之機械手指採用電磁閥並搭配自行灌模的矽膠手指來蓋放指孔。由於電磁閥所需的電壓為 12V，因此採用 MOSFET 隔離開關來驅動。

(三)直流馬達：採用 LT37GB90-3329 的直流馬達，其工作電壓為 DC12V，前置 1:90 的減速比的減速機，降低速度加大扭力，透過一個連桿裝置往復運動，推動風箱送氣到中國笛的吹孔，藉此發出聲音。直流馬達與機械手指一樣皆採用 MOSFET 可 PWM 開關模組來調控。

參、研究過程或方法

一、研究過程

(一) 吹氣控制

中國笛在吹氣時，需要與笛子對準孔邊的尖銳邊緣吹氣形成邊稜音，邊稜音觸發笛子空氣柱共振產生笛音。各音都需要適當調整吹氣的速度才能準確地發聲，需要讓吹氣裝置根據音頻分析結果，滾動性調整其速度，如吹高音時提高吹氣流速，要吹低音時，降低流速。吹奏笛子時，聲音中均會包含決定正確音高的主音及 2 倍或 3 倍主音頻率的泛音，控制流速可以決定壓抑泛音，使聲音更加純淨；提升泛音使聲音成為高八度音

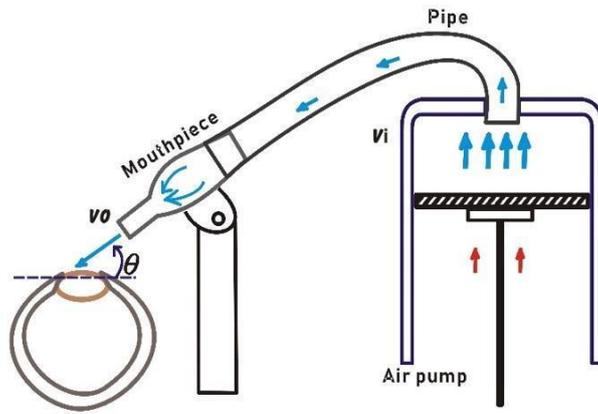
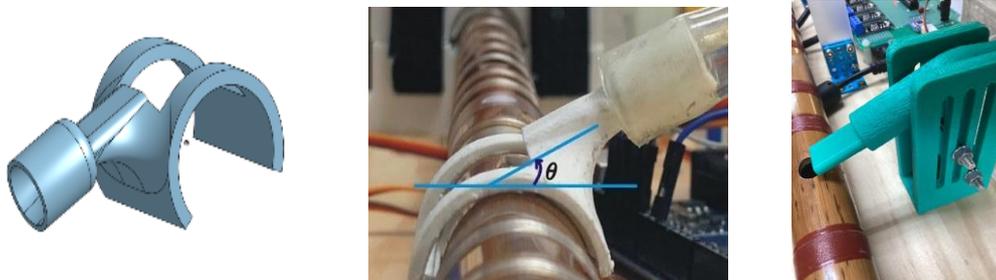


圖 9：幫浦氣量、吹嘴及角度關係示意圖（資料來源：研究者自行繪製）

1. 吹嘴 (Mouthpiece)

- (1) 設計與製作：根據上方說明，為了使笛子發音，於是需設計一個專對特定角度吹氣的吹嘴，產生邊稜音，以觸動空氣柱共振。吹嘴設計如上圖所示，其中 v_i 為進氣流速、 v_o 為出氣流速、 θ 為吹嘴角度。依據需求設定吹嘴設計圖如圖 10(a)，實際裝配於笛子吹孔如圖 10(b)，後續改進的吹嘴如圖 10(c)。



(a) 吹嘴的 3D 模型 (b) 吹嘴與角度關係 (c) 可調角度之吹嘴

圖 10：使用 3D 列印設計中國笛的吹嘴（資料來源：研究者自行拍攝及繪製）

- (2) 測試：使用雙動式幫浦送氣，微調吹嘴角度 θ ，逐步沿 θ 角旋轉微調，運用頻譜分析程式依發聲音量的大小及頻率，來判斷最佳的角度。在正確的角度也可以使笛的不只音量提升及音準也會更趨於準確。

測試音以按兩孔，音階為 E4 的音調，標準的音頻為 659Hz，空氣幫浦送氣的氣量與速度固定，使用頻譜分析軟體「Spectrum View」偵測主音的音量及音頻，來判斷最佳角度。圖 11 所示，可以發現 A、B、C 三個吹嘴角度，音頻變化不大大約若在 650Hz 及 700Hz 之間，音量參考尺度表，顏色依紅橙黃綠靛藍及各色由表達由大到小的強度，因此深紅最強、深藍最弱，強度的單位為 dBFS (Decibels Relative to Full-Scale)，此數值只有負值，0 代表最高強度。

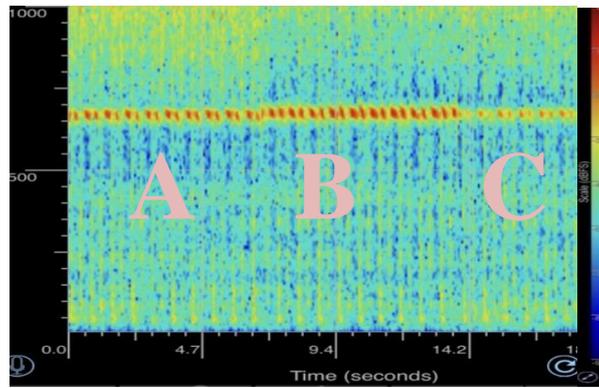
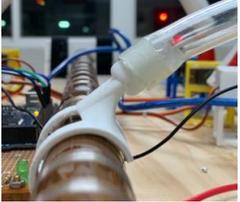
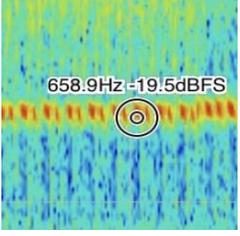
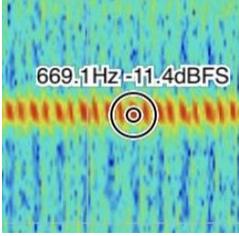
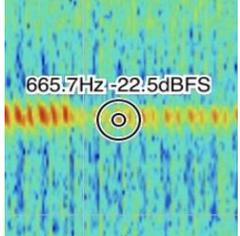


圖 11：吹嘴在各角度之音頻檢測（資料來源：研究者自行截圖）

分析使用 θ 為各個角度進行測試，下方列舉出其中三個角度，依照 Spectrum View 分析音量及音頻，記錄如下表所示。

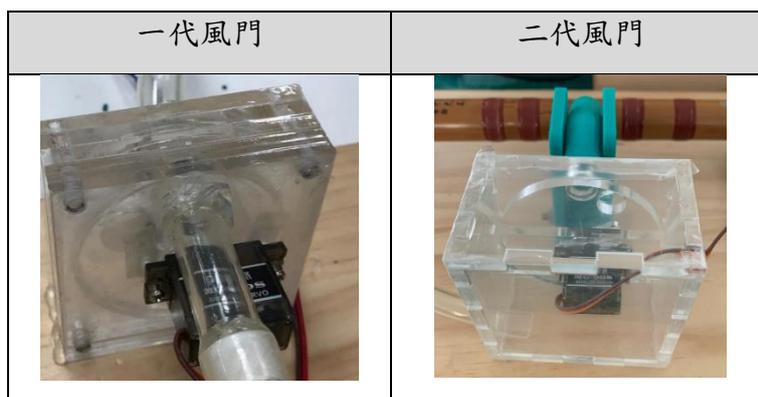
表 10：吹嘴角度與音量、音準的關係測試(資料來源：研究者自行拍攝及收集)

段落	A	B	C
角度	38°	33°	16°
實體圖			
數據點			
音量	次大	最大	最小
音頻	658.9Hz	669.1Hz	665.7Hz

實驗結果：

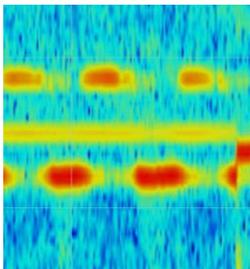
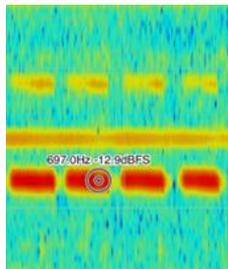
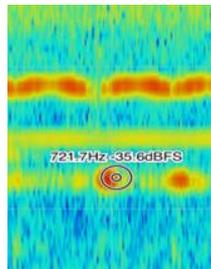
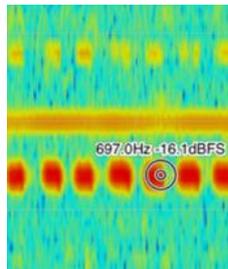
- (1) 同一個送氣裝置下透過吹嘴吹奏笛子，角度接近 33°時，音量為-11.4dBFS 音量最大，其次為 38°為-19.5dBFS，16°為-22.5dBFS 音量最小。
 - (2) 角度對音頻高低差異較小，角度 38°時音頻略低於 33°，角度大會使音頻偏低。
2. 風門：在設計雙動式風箱時休止符及音樂結尾時會有餘從環緩慢的排出而發出聲音，藉由通道阻擋通道來避免多餘的氣體進入中國笛，增加風門後改善了以下幾點：
- (3) 用於堵氣及送氣，使演奏聲音可以準確發音及停止。模擬笛子的吐音演奏技巧。
 - (4) Spectrum View 觀察聲音頻譜，加上風門後聲音間隔明確雜音減少，測試如下圖。
 - (5) 透過增加圓周半徑的方式使得伺服馬達以較小的轉動角度即可控制風門，達到在 120ms 內的快速連續吐音。

表 11：風門設計(資料來源：研究者自行拍攝及收集)



實驗結果：使用風門控制 4 分音符及 8 分音符的輸出，比較有無風門的差別，可以發現不管是 4 分或 8 分音符只要有風門控制，頻譜都呈現各音分明，泛音微弱的良好狀態；反之沒有風門則聲音渙散、音準不穩定及泛音強的狀況。

表 12：風門測試(圖片來源：自行擷取、拍攝)

4 分音符 無風門	4 分音符 有風門	8 分音符 無風門	8 分音符 有風門
			
間格不明確雜音多	間格明確	間格不明確雜音多	間格明確

3. 空氣幫浦 (Air pump)

(1) 設計與製作：有了吹嘴之後需要設計一個可以模擬人類肺臟送氣的裝置，並且這個裝置還要可以調整吹氣時的氣量及流速。為了使 Arduino 可以控制馬達轉速，設計了電路，使其達到最高轉速時將 PWM 設為 255(100%)。氣體流速、流量及管截面積的關係如下，可見管截面積越大流量越大，截面積與流速成反比。

$$Q = A \times v \quad (Q: \text{流量}, A: \text{截面積}, v: \text{流速})$$

初步使用餅乾桶設計單向的空氣幫浦，經測試時發現每個音階中間都會有很大的停頓，主要馬達折回時只有進氣沒有出氣。若改成雙動式活塞，也就是來回的時候皆會送氣，雖有停頓但較為持續。然而，仍不符合我們的要求，所以我們設計了雙動式風箱，他不僅可以雙向供氣，還搭配了壓力調節風箱，如下表。

表 13：幫浦種類(資料來源：研究者自行繪製)

種類	單向空氣幫浦	雙向空氣幫浦
結構圖		

由實際的吹笛樂曲，通常會有長短不一的長音，然不管是單向或雙向的幫浦，無法持續一拍，雙向式風箱解決了此問題，但仍因為幫浦運動，會形成規律的噪音，是極需要解決的問題。

使用雙動式活塞，測試馬達轉速與音量及音頻的關係：以上述測試的架構，利用改變 PWM 改變馬達轉速，比較音量大小及音頻準確的程度。其中紅色的文字為最接近標準值的音頻。

實驗結果：由表 15 之實驗數據，推斷出如下表之簡表，當流量及流速低時，音高就低，且不穩定，聲音亦小；當流速低但流量高時，音高偏低。

表 14：音頻與音量測試 (資料來源：研究者自行收集)

音階	標準頻率	馬達轉速(PWM)									
		195		230		245		250		255	
		音量	音頻	音量	音頻	音量	音頻	音量	音頻	音量	音頻
A	440	-18.5	436.1	-17.0	436.1	-8.5	436.1	-10.0	444.7	-10.6	444.7
B	494	-18.6	487.5	-17.9	496.1	-10.5	487.5	-7.5	508.1	-7.4	494.4
#C	554	-18.4	573.2	-16.7	556.1	-14.0	547.5	-10.1	559.5	-9.1	559.5
D	587	-18.2	590.3	-13.0	590.3	-10.3	590.3	-9.5	597.2	-7.0	593.8
E	659	-19.8	676.4	-8.3	658.9	-12.0	676.0	-8.0	665.7	-7.4	665.7
#F	740	-13.0	753.1	-10.7	761.7	-12.3	753.1	-11.0	758.2	-7.0	765.1
G	784	-20.3	795.9	-19.1	813.0	-18.8	804.5	-11.6	789.6	-10.1	813.0

4. 風箱 (Bellows)

為解決送氣維持時間過短，無法構成長音的樂句的問題，在參考 YouTube 影片有關自製管風琴 (Organ) 的送氣裝置，使用風箱送氣，給我們很大的啟發。



圖 12：Youtube 上關於自製管風琴使用風箱送氣的展示影片

(資料來源：FlpBusker(2011), **Demonstration Organ**)

由上述的啟發，製作出類似管風琴的風箱，不僅較原本的單或雙動式活塞輕巧效率更高，使用皮帶輪搭配馬達與將扇型風箱不間斷地進出氣，並在前面增加一個調壓風箱，達到類似濾波的效果，即可達到連續吹氣之目的，如果沒有增加調壓風箱，氣流則會像雙向幫浦一樣斷斷續續。

下圖為雙動式風箱運作示意圖，左方兩個風箱上下交替進氣與出氣，配合逆止閥控制進氣與堵氣，使氣流可以正確及完整的擠壓到調壓風箱，藉由調壓風箱上頭的厚重木板的重力，將氣體擠壓到出口送到笛子的吹嘴。

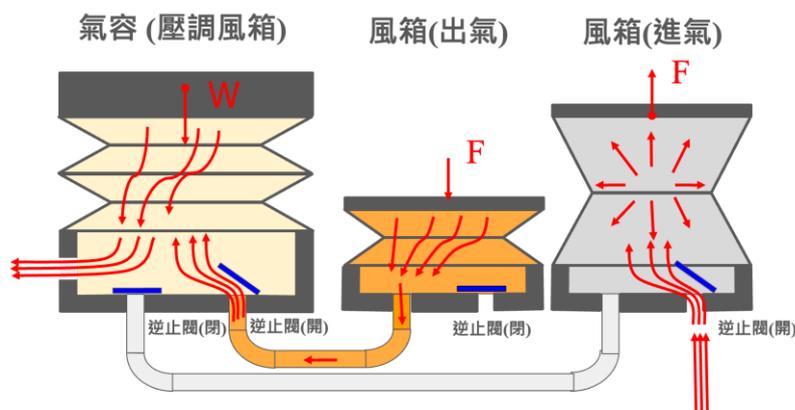


圖 13: 雙動式風箱運作示意圖 (圖片來源：自行繪製)

(1) 風箱設計製作：使用雷射切割與木板製作扇形鼓風箱，模仿 Youtube 製作過程，使用羊皮作為風箱接合材質，配合皮革膠黏著密合接著的效果良好。所製作的成品相較其他材質，羊皮擁有更加的密閉性，且柔韌度較高。

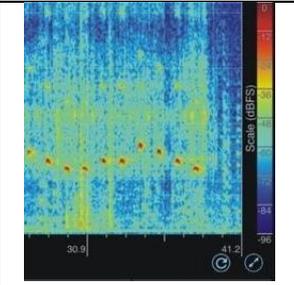
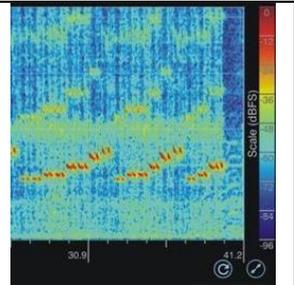
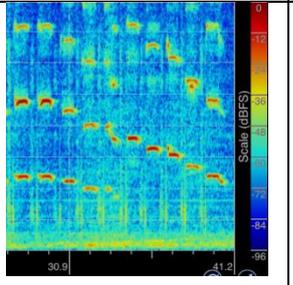
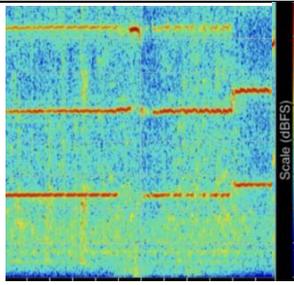


圖 14：雙動式風箱製作(圖片來源：自行擷取及拍攝)

(2) 分時頻率強度測試：為確定幫浦產生之氣流與音階之間的關係，利用軟體 SpectrumView 測試各音階各頻率分時的強度。

實驗結果：畫面如下表，扇形鼓風箱相較於單向及雙向幫浦可以達到更穩定送氣度的目標，當使用兩個扇形鼓風箱同時運作並配合上氣容，則聲音可以連續輸出。

表 15：風箱音階頻譜製作(圖片來源：自行擷取、拍攝)

單動式幫浦	雙動式幫浦	單向風箱	雙向式風箱 (含壓調風箱)
			
持續性短、不連續	持續性短、較連續	持續性長、不連續	聲音連續明確輸出

(3) 使用雙動式風箱，測試同指法在風門大小改變時所產生的音頻範圍：以上述測試的架構，利用改變伺服馬達角度改變風門大小，測試音階音頻範圍並選定各個音階最佳風門大小，其中 30 度為全開；53 度為關閉。

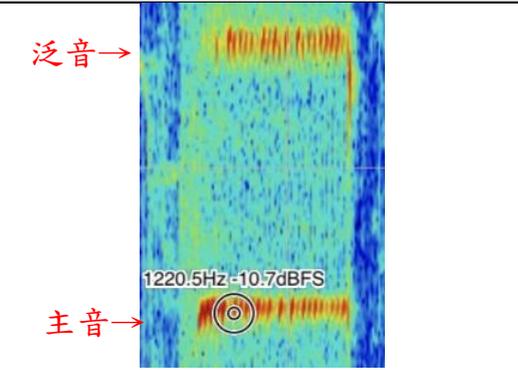
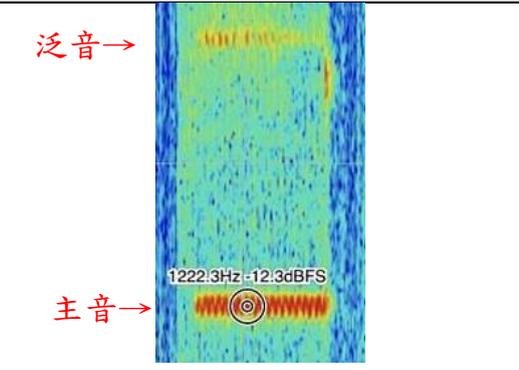
為確定最佳音準，必須針對每個音進行調校，調校的原則如下：

原則一：主音應該遠大於泛音。

原則二：主音越接近標準音頻越好。

實驗結果：下表為音頻調校在不良與目標結果之差異比較，左圖泛音大於、接近或等於主音，則音色不良；右圖調校結果泛音不明顯，音色良好且靠近標準音階。

表 16：音頻調校結果 (資料來源：研究者自行收集)

未達目標音頻結果	達到目標音頻結果
 <p>泛音→</p> <p>1220.5Hz -10.7dBFS</p> <p>主音→</p>	 <p>泛音→</p> <p>1222.3Hz -12.3dBFS</p> <p>主音→</p>
調整前:35 度時泛音大於等於主音	調整後:40 度時泛音小於主音且接近標準音階

接下來針對各音進行音頻校正，使風箱轉速固定，針對風門大小進行調整，風門設計使用伺服馬達轉動圓盤，改變出氣孔的大小。伺服馬達角度越大則出氣孔越小；角度越小則出氣孔越大。

實驗結果：依上述的調校原則，針對每個音進行調校，結果如表 15 並在最佳的風門角度取五次之平均：

表 17：音頻與音量實驗結果（資料來源：研究者自行收集）

音名	標準頻率(Hz)	實測音頻 Hz(音量(dBFS))					風門角度(度)	
		平均(大小差)	1	2	3	4		5
C5#	554	550 (-16.51)	554	570	554	530	566	50
D5	587	607 (-14.16)	605	607	606	606	608	45
E5	659	670 (-13.46)	671	668	670	670	671	50
F5#	740	744 (-12.87)	755	758	756	755	730	45
G5	784	821 (-11.57)	817	818	802	839	840	40
A5	880	900 (-11.27)	899	886	907	903	914	40
B5	988	991 (-10.04)	969	1012	993	993	1011	42
C6#	1108	1146 (-9.31)	1098	1098	1193	1141	1141	37
D6	1244	1200 (-9.11)	1196	1202	1204	1201	1201	40
E6	1319	1361 (-8.89)	1352	1358	1361	1370	1360	35
F6#	1480	1522 (-8.65)	1528	1519	1516	1527	1528	35
G6	1568	1614 (-8.41)	1622	1634	1640	1647	1580	30
A6	1760	1802 (-8.22)	1791	1800	1809	1813	1809	30

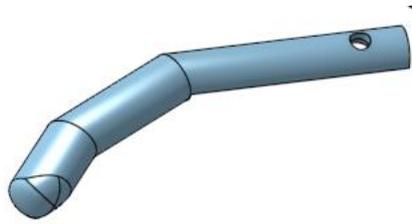
由上表之實驗數據，推斷出如下結論，當風門大時，音頻變高，當風門小時，音頻變低。且在特定頻率(角度)會提升一個八度的音階，而部分數據相較標準音頻略高，主要為竹笛的關係，且由於使用的竹笛無法調音因此會有略高或略低的情況。

上述實驗竹笛的音準調校已可以讓依照風門大小調校改變音高，使趨近於音準。然竹笛調音還是有其困難之處，以本次實驗為例，若使竹笛音色飽滿則整體音高將近高半個音，而實驗用的竹笛又無調音環將笛子的音柱拉長。解決的方法應該換一隻有調音環的笛子，使笛子可以在音色飽滿的情況下吹出準確的音準。

(二) 指法運動

1. 機械手指

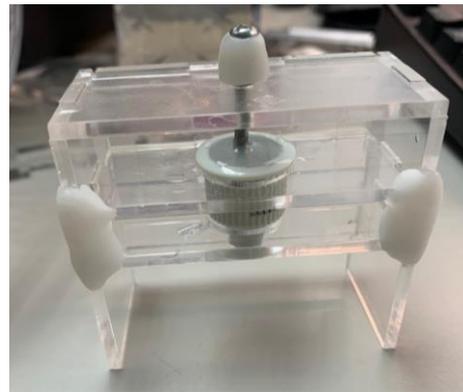
(1) 設計與製作：為了使機械手指能夠模擬出類似人類的指法，於是利用 3D 列印機來印出了一個模擬人類的手指。其中的結構是由伺服馬達的接合處控制來手指本體，借由其改變角度以控制指孔的開閉。其中密封在機械手指當中是極為重要的，找尋了各種材料包含矽膠、塑膠墊、熱熔膠等等，最終經由測試發現選用了大氣泡紙會有較佳的效果，其氣泡的特性幫助了密封。



(a) 機械手指 3D 設計圖 (b) 機械手指實際照片搭配密封材料

圖 15：機械手指圖片（資料來源：研究者自行拍攝及繪製）

- (2) 機械手指氣密改進：舊版是使用包裝紙來當作按壓指孔的材料，但不夠密合，於是我們調配矽膠手指，混合了 SJ3200 的 AB 矽膠，A 與 B 1：1 混和，經測試可以達到密合的效果，混和後我們會將其倒入眼藥水蓋子中，並在中間置放螺帽，待其固化即可完成，如下圖所示。



(a) 初期使用矽膠製作機器手指無法密合 (b) 採眼藥水蓋當模具灌製手指可以完整密合

圖 16：機械手指灌模製作（資料來源：研究者自行拍攝）

- (3) 機械手指速度改進：由於舊版之力臂及速度問題，架設了平台測試了速度相關之數據。

實驗結果：使用伺服馬達來回最快只要超過 196ms，就使機械鍵盤讀值異常，而使用電磁閥之手指則可以來回最快 78ms，大約 1 秒鐘可以按壓 12~13 次。

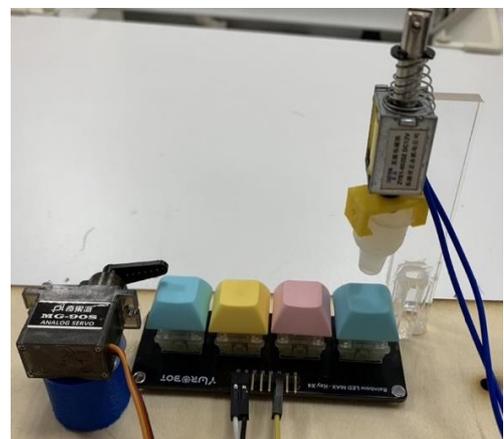


圖 17：各版機械手指速度測試（資料來源：研究者自行拍攝）

- (4) 設計與製作：為了使機器人能按照樂譜辨識之結果吹奏，於是就多了指法程式撰寫的必要，整體程式運作方法如下圖，使用者只需要透過網頁介面上傳樂譜，樂

譜程式辨識完即會透過序列傳輸傳送音高及音長給微控制板，詳細如下圖，當接收到樂譜資料後，微控制板會加以解析，例如回傳的資料為 080500，則代表吹奏 E5 這個音持續 500 毫秒，吹奏時也同時控制幫浦速度，來達到正確音高音長。

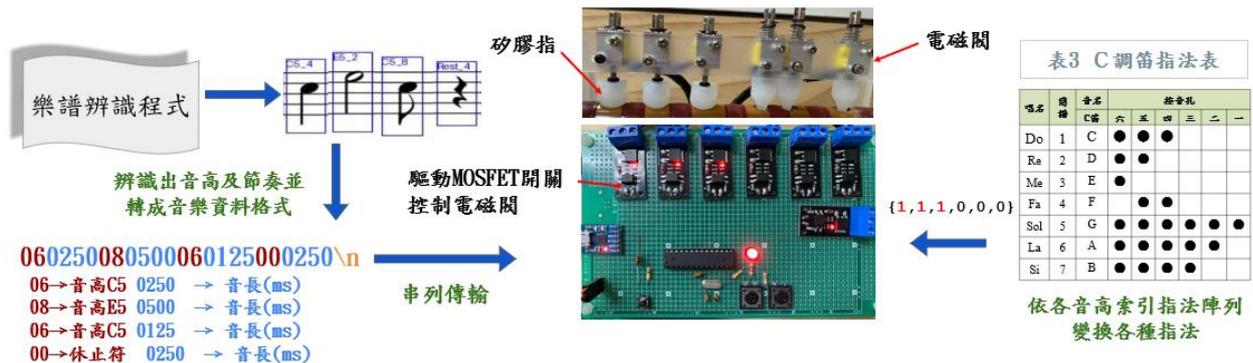


圖 18：機械手指整體運作架構 (資料來源：研究者自行繪製及拍攝)

(三) 樂譜辨識程式開發與實作

五線譜樂譜辨識複雜度很高，若採用所有不同音高與音長的音符分類，光以五線譜 5 線 4 間，加上音長全音符到十六分音符、休止符，及上述符號的附點符號，就已經有至少 100 個分類，且尚無法涵蓋所有的音符，樂譜辨識程式使用單一神經網路有很大的難度，未來擴充分類上會面臨標記及訓練的工作份量及時間上的困難。

因此規劃辨識過程規劃為二個階段：第一階段主要辨識譜線，第二階段辨識音符、音長、有無附點，這樣的分類規劃將使各種分類少於 30 個，具有較高的彈性及可擴充性與較少的訓練週期。

1. 樂譜辨識樣本目標：因開發上樣本收集及標示工程頗大，初期使用 5 線 4 間的範圍收集樣本，試行可行後，擴增到以上下三線，分成兩組數據集，數據一包含譜線、GCF 譜號與速度記號。數據二包含影像音高、音長、有無附點共三種分類結果。

表 18：數據一樣本(資料來源：研究者自行繪製)

譜線	其餘物件

表 19：數據二樣本(資料來源：研究者自行繪製)

音高	音長	附點

2. 半監督式學習：採用「半監督式學習」的策略，先人工產生樣本及模型，經訓練並應用後產生新的樣本，再擴充樣本重複訓練，逐步提高模型的精確度及召回率，流程如下圖所示。



圖 19：訓練流程圖 (資料來源：研究者自行繪製)

網路收集適合笛子吹奏的樂譜影像作為數據一，使用 LabelImg 針對譜線進行標記，產出譜線樣本作為數據二，再針對音符進行標記產生音符樣本。後再進行初審以維護品質。

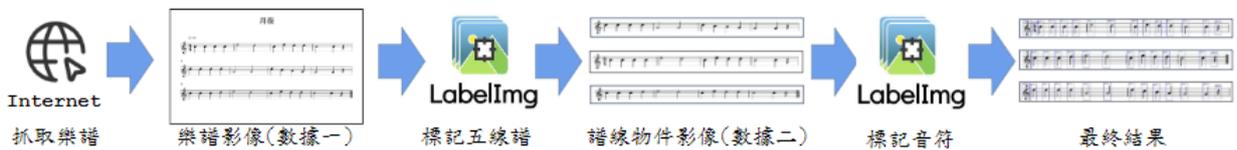


圖 20：樣本初始標記流程圖 (資料來源：研究者自行繪製)

將整張樂譜分成小段進行標記物件，並訓練產生出兩個模型，透過模型應用將整張樂譜辨識後產生新數據，數據透過人工方式審查，確保數據正確。

樂譜影像放入模型一標記基本物件並記錄其座標。取出譜線物件進行整理後一併放入模型二進行辨識，取得其中音符等物件。物件座標會轉存為(class, x, y, w, h)格式的文字數據。



圖 21：數據集處理程序 (資料來源：研究者自行繪製)

3. 多任務學習：採用多任務學習能夠提高模型泛化能力、降低過擬合，並通過特徵共享來降低計算成本，本研究規劃如下：

- (1) 模型一：僅使用一個任務 (Task) 進行辨識譜線、譜號、速度物件的辨識，其中譜線辨識結果將丟入模型二進行其他 Task 的辨識。

- (2) 模型二：使用 Task A、Task B 分別辨識譜線內音符音高、音長，另外由於音符的「附點音符」辨識較為特別，為避免附點音符增加 Task B 的負擔，增加了 Task C 獨立辨識有無附點的工作。

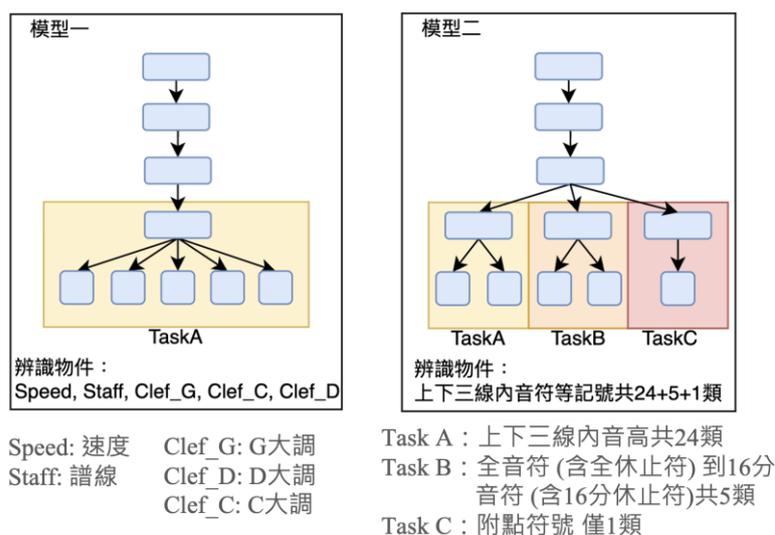
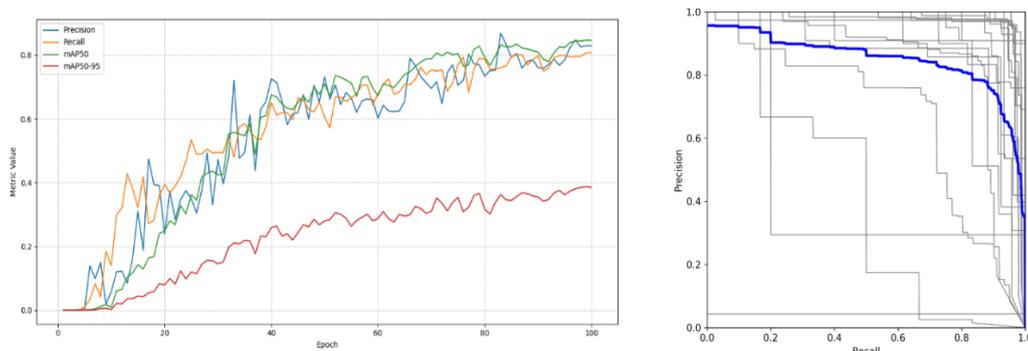


圖 22：模型架構及各 Task 的辨識分類 (資料來源：研究者自行繪製)

4. 訓練成效：模型訓練過程中產生各項指標，包括 Precision, Recall, mAP50, mAP50-95 指標，將其繪製成圖以便觀察確保模型成效。



- (a) 訓練的品質透過圖表判斷模型二的 3 個 Task Precision, Recall 及 mAP50 均 $\geq 80\%$
(b) P-R 曲線凸顯訓練模型大部分有準確召回

圖 23：模型評估圖 (資料來源：模型自行產生)

(四)多任務學習規劃

上述的樂譜辨識程式採用三種模型分階段進行辨識，三模型雖然是同一種樣本形式，但卻各自獨立，雖已初具多任務學習的特性，然而模型間參數的毫無共享。為了更加強讓任務間的參數可以相互共享，提升訓練及辨識效能，發揮多任務學習的特性，使同一類型的模型可透過共享實現多任務的特性，進行實作如下：

1. 數據產生方式及樣式：與前述的方法相同透過軟體 LabelImg 將網路蒐集或自製的

樂譜做標記物件分類，每張樂譜會產生三份數據集，分別是譜線樣本、音符樣本、節奏樣本將所有譜線標記後產出譜線樣本，再從其內標記音符等元件形成音符樣本，最後根據吹奏音符進行二次分類，產生出節奏樣本。訓練音符樣本將擴大到 5 線 4 間加上 3 線 3 間及下 3 線 3 間，也就是 G3 到 B5 的全音，共 20 類；節奏(音長)擴充到全音符、二分音符、...、十六分音符及外加附點共 10 類。

2. 模型架設階段架構：模型大至可分成三個任務，分別目標是提出、辨識以及分類。任務一將樂譜譜線進行提出，所有譜線都將單獨進入任務二進行音符等元件的辨識，辨識出來的吹奏音符會再單獨被帶入任務三分類出該音符的音長。可發現在每個任務階段都會將部分特徵共享到下一任務，實現多任務學習的共享特性。

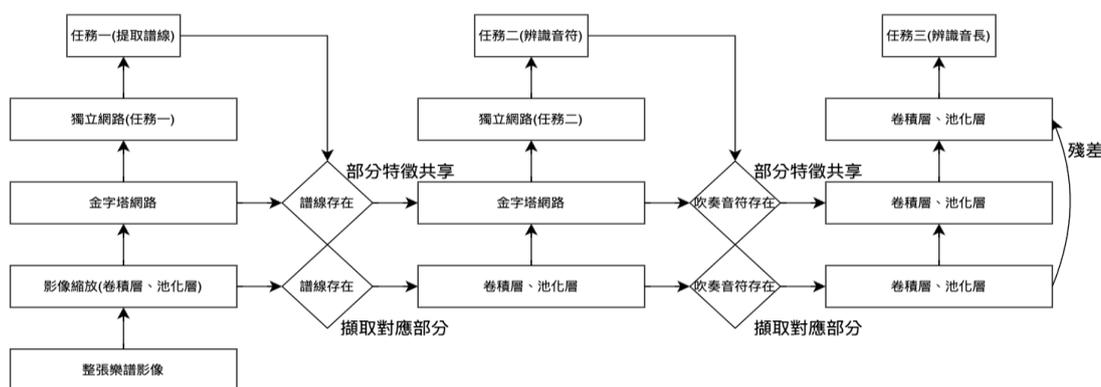


圖 24：多任務模型架構圖（資料來源：研究者自行繪製）

而其中使用到金字塔網路，主要將影像以金字塔的形式從下而上進行卷積、池化等操作如下圖，再從上而下重複進行，相同高度下金字塔會朝下一個連接，不斷的特徵變換、彌補下輸出三個不同尺度的辨識結果，將三尺度的結果重新縮放至統一大小後，進一步計算損失並反向傳播，任務二與任務三的反向傳播將回到任務一並對其造成影響，從而形成參數共享。

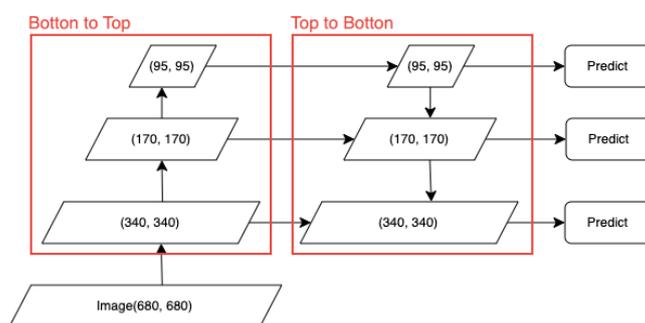


圖 25：金字塔網路架構（資料來源：研究者自行繪製）

3. 模型訓練：

- (1) 數據集處理：各份數據集以 7:2:1 的形式隨機分割，產生出訓練集、驗證集與測試集，訓練集會將數據進行模糊、噪點處理，使圖片發生些微改變。
- (2) 設定及方法：使用 YAML 文件做出模型的設定檔，一份初始設定、一份階段自訂義設定，並細分化文件架構，使訓練、撰寫輕鬆。

4. 模型評估：模型訓練過程中紀錄損失值，在測試狀態紀錄召回率、準確率並匯成圖。且為防止梯度消失、爆炸，過擬合等一系列問題，模型的實時評估至關重要，當模型出現問題時可以立即暫停並查找可能問題出現點。

模型一採用 Adam 優化器，理想情況損失會呈現收斂狀態，直至趨近於零，模型二採用的是 Adam 的改良版 Adadelta，理想情況下前期損失變化率大，中期會逐漸找到學習方式並開始收斂，對於複雜的辨識也能夠找到合適的學習方法，模型三使用 SGD 優化器，收斂更加平緩，適合用於簡易模型中。

5. 網頁設計：為應用上述訓練及評估可用的模型於輸入樂譜的辨識，設計一個網頁介面，可以讓使用者，透過前端的網頁將樂譜影像圖檔上傳到後端，再透過後端的辨識程式進行辨識，最後產生可供輸入到機器人的音樂資料，進行笛子演奏，說明如下：

- (1) 前端設計：以 React 為前端框架搭配 Flowbite 進行介面設計，結合使用者介面 (User Interface)、使用者體驗 (User Experience)，提供使用者更好的體驗環境。
- (2) 後端處理：以 Flask 構建後端，接收前端所發送的樂譜資料並丟入模型辨識，儲存結果、管理數據集。

肆、研究結果

一、第一代：雙動式幫浦之吹笛機器人

右圖第一代機器人，使用雙動式幫浦將氣送至吹嘴，控制電路控制機械手指按壓笛孔。由於送氣是靠馬達旋轉上下推送活塞，因此在上下推氣同時，才会有氣送到吹嘴，而不管活塞往上或往下，只有在末端送氣的瞬間才会有氣。因此只能吹奏間格相等且時間速度較慢的音樂，在這個開發階段尚無樂譜辨識，所有的音樂資料，都是使用程式輸入。

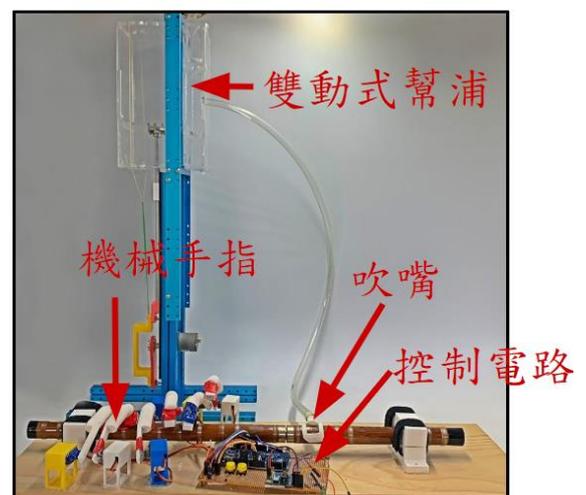


圖 26：第一代實體圖（研究者自行拍攝）

二、第二代：樂譜辨識雙動式風箱吹笛機器人

下圖第二代機器人採用雙動式風箱，搭配一個氣容，可以讓氣流綿綿不絕地送氣。加入了樂譜辨識程式，透過串列傳輸音高及音長之資料給為控制板進行解析，解析完成之後再找到對應的指法、風門大小及風箱速度，進行吹奏。

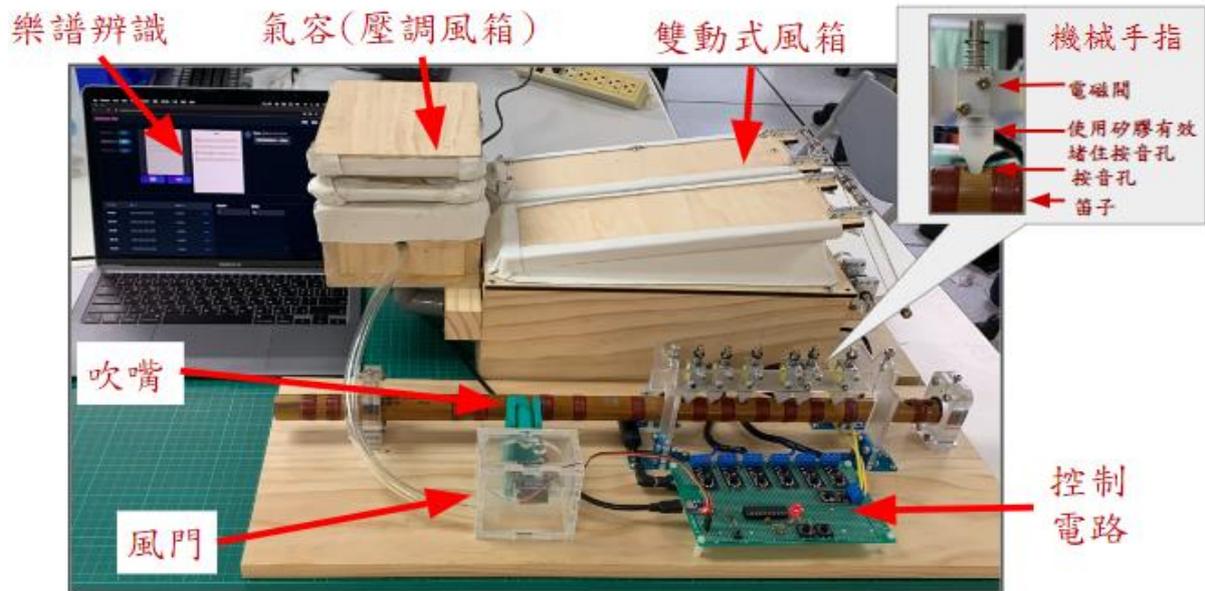


圖 27：第二代實體圖 (研究者自行拍攝)

三、完整音階演奏

在整體經過調整讓每個音確定可以正確的發音後，測試演奏上行及下行音階，確定機器人可以演奏 B3~A5 的音高，經過 Spectrum View 測試錄製結果如下圖。

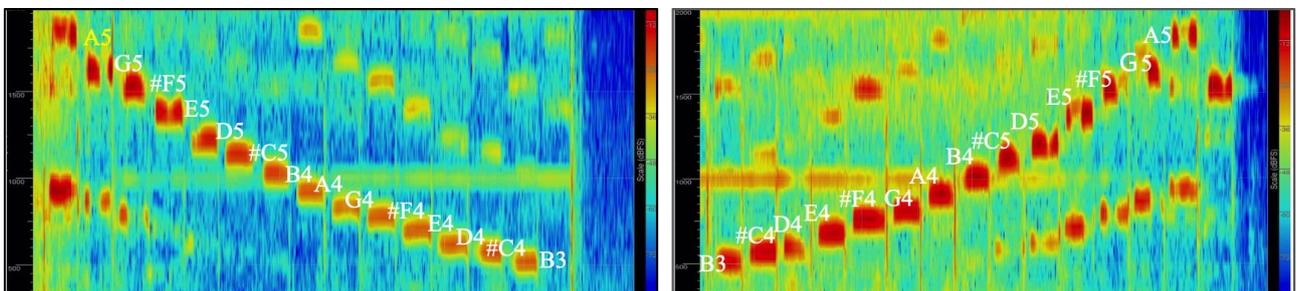


圖 28：測試完整的音階下行及上行結果 (作者繪製及拍攝)

四、模型訓練成果

以下是各版本樂譜辨識的差異，表(a)為數據樣本真實範例，表(a)為數據樣本之說明，表(b)為各版實際比較。

表 20：各數據版本比較 (資料來源：研究者自行撰寫)

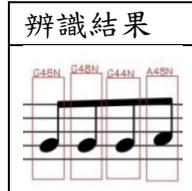
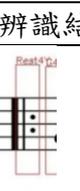
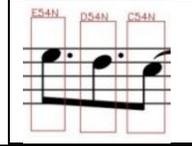
項目	數據 A	數據 B	數據 C
樣本數量	25	135	267
前處理	模糊、噪點	模糊、噪點	模糊、噪點、傾斜
樣本來源	● 潔白電子樂譜	● 網路電子樂譜	● 網路電子樂譜 ● 真實影像樂譜

表 21：各模型訓練版本比較 (資料來源：研究者自行撰寫)

項目	Precision	Recall	mAP50	mAP50-90	範圍	數據
V1.0	35.5%	20.3%	無資料	無資料	五線四間	A
V2.0	45.2%	25.0%	無資料	無資料	五線四間	A
V2.1	61.7%	24.1%	無資料	無資料	五線四間	A
V2.2	60.3%	62.4%	57.3%	24.3%	上下三線	B
V2.2.1	75.4%	74.4%	75.3%	35.0%	上下三線	C
V2.2.2	77.9%	79.0%	76.1%	39.1%	上下三線	C
V2.2.3	85.8%	78.8%	84.0%	39.7%	上下三線	C
V2.2.4	88.9%	81.5%	88.1%	43.3%	上下三線	C

經訓練後發現使用裁切會使模型退化而無法學習，因此在各版本中皆不裁切。而使用網路電子樂譜與部分的真實影像樂譜，提升了模型在雜訊上的抑制，有效提高辨識率，添加部分影像傾斜使其能夠適應在未來使用攝像機拍攝樂譜所產生的角度偏差。

表 22：各模型辨識結果比較

			
錯誤樣態		錯誤樣態	
辨識結果	錯誤	辨識結果	錯誤
	都是 8 分音符 (6N)，結果有辨識成 4N 及 8N。		應為反覆符號，但模型沒有這個樣本，所以誤判為休止符。
	未辨識到附點符號，音長應為 8 分音符(6N)，結果都是 4N。		
V2.2 音符分類多數有誤		V2.2.4 音符分類大多正確	

在上表兩個版本中發現 V2.2 版本的音長及附點符號大多都辨識錯誤，而再加強樣本的模糊、噪點及傾斜處理後，召回率及泛性大幅提升。

五、樂譜辨識程式

設計了網頁介面，讓使用者選取樂譜圖檔，當樂譜上傳到後端後，此時後端會應用以多任務學習為架構訓練模型，驅動模型進行樂譜辨識。辨識後會產生下圖右方的樂譜檔。對於吹笛機器人來說，也會在辨識後接收到音樂檔，接著進行演奏。

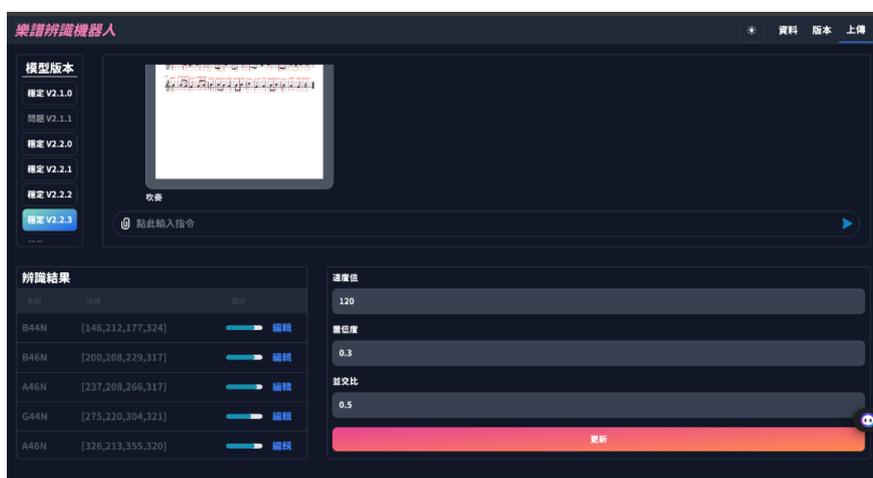


圖 29：網頁介面(資料來源：研究者自行拍攝)

下圖為辨識實例兩首，輸入的樂曲經過辨識後，可以看到大紅線框出樂譜的每一行譜線，這是任務 1 的辨識結果(最右側(含熱力圖))，接著在任務 2 較小的紅線框出譜線上的每一個音符，在辨識出音高之後，接著用同樣的音符進行任務 3 辨識節奏(音長)。經辨識成功的曲子選擇在上下各三線範圍內的曲子，以下以窗邊的故事及紫竹調為範例。

表 23：辨識結果(資料來源：研究者自行截圖)

原始樂譜 / 辨識結果 / 熱力圖		
<p>窗邊的故事</p>	<p>窗邊的故事</p>	<p>窗邊的故事</p>
<p>紫竹調</p> <p>(Lyrics) (Composer)</p>	<p>紫竹調</p> <p>(Lyrics) (Composer)</p>	<p>紫竹調</p> <p>(Lyrics) (Composer)</p>

伍、討論

本作品創新技術方法及構想：

表 24：討論 (資料來源：研究者自行撰寫)

項目	創新技術	方法	構想
吹部	壓力調節風箱	增加在兩個風箱前面	參考濾波電路的電容發想到增加氣容是否能讓氣流更穩定。
指部	電磁閥手指	利用電磁閥來按壓指孔	按壓時，搭配自行灌模的矽膠手指，可以使按壓更加密合，且使用電磁閥可以大幅加快按壓速度。
眼部 1	可插拔架構	將模型拆分成階段式	可替換其中一個部件模組進行樂譜辨識，每部件都有屬於自己的專屬特徵，只需確保輸入輸出格式相互對應。
眼部 2	多任務學習	採用多任務學習模型訓練	獨創以多任務學習來辨識樂譜。

陸、結論

一、吹笛機器人的研究結果

本研究基於真實吹笛者的身體架構進行吹氣、手指的仿生，建構以雙動式扇型風箱可連續送氣，以多任務學習模型進行高效的訓練策略及精簡可擴充的樂譜辨識模型，完成一部具有識譜能力的六孔竹笛吹笛機器人，成果如下：

1. 使用雙動式風箱搭配氣容(壓調風箱)，可吹連續的長音，針對休止符設計電子風門放氣及堵氣。
2. 定義音高及音長的資料，提供給機器人控制音階指法，以吹奏不同音高及音長。
3. 笛子各音應調節馬達速度及風門大小控制，以吹到正確的音高及音準。
4. 運用多任務學習，簡化成兩模型識別五線譜譜線、20 個音高、10 種音長及附點等四個工作分類。
5. 採用半監督式的訓練策略，運用初始標記樣本及模型的辨識結果產生新樣本逐步提升訓練效能。

6. 運用網頁介面上傳樂譜圖檔並辨識，並轉成資料以串列送入微控制器使機器人進行演奏，機器人目前可以順利辨識笛子的樂曲並演奏。

二、未來展望

多任務學習具有共享資訊的特性，在文獻探討過程中，有一篇有關於多任務學習運用於音樂情緒辨識的模型。在本研究有模型有三個任務，若再引入情緒辨識，讓表情記號、強度記號等辨識形成的音樂情緒，表現在吹笛機器人上，讓吹奏出來的音樂帶有更適切的情緒表現。笛子在表現樂曲情緒上，除了音階外，氣流速度、氣量大小

的控制表現音樂的強弱，在機器人的架構上尚須具有控制風門的大小的機構，才能構成完整吹笛表演的各項參數，表

現音量大小控制音樂的起伏。若上述機構完成，配合上樂譜辨識後所產生的表情分類，將可達到更接近人類演奏表現，更接近具有模仿真人演奏的機器人。

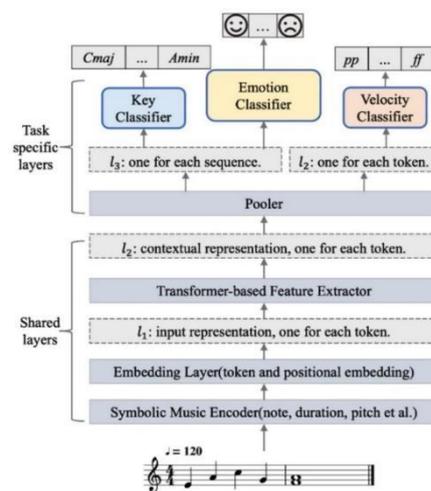


圖 30：音樂情緒辨識的模型 MT-SMNN 架構
(資料來源：Jibao Qiu, C. L. Philip Chen and Tong Zhang(2022), **A Novel Multi-Task Learning Method for Symbolic Music Emotion Recognition**)

柒、參考資料及其他

- [1] Ben Hu (2020), (深度學習)ResNet 之殘差學習, <https://reurl.cc/oRyjaj>
- [2] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia (2018), **Path Aggregation Network for Instance Segmentation**, <https://arxiv.org/abs/1803.01534>
- [3] Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao (2024), **YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information**, <https://arxiv.org/html/2402.13616v1>
- [4] 人間別久不成悲 (2022), **FPN、PAN、BiFPN 的區別**, <https://www.cnblogs.com/wensieric/p/16617794.html>
- [5] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie (2016), **Feature Pyramid Networks for Object Detection**, <https://arxiv.org/abs/1612.03144>
- [6] Shikun Liu, Edward Johns, & Andrew J. Davison (2019), **MTAN: Multi-Task Attention Network**, <https://shikun.io/projects/multi-task-attention-network>
- [8] Sebastian Ruder (2017), **An Overview of Multi-Task Learning in Deep Neural Networks**, <https://reurl.cc/NQIWVm>
- [9] fengbingchun (2021), 深度學習中的多任務學習介紹, <https://reurl.cc/0vd3Y6>
- [10] Jorge Solis, Kenichiro Ozawa, Maasaki Takeuchi, Takafumi Kusano, Shimpei Ishikawa, Klaus Petersen, Atsuo Takanishi (2014), **Biologically-Inspired Control Architecture for Musical Performance Robots**, <https://journals.sagepub.com/doi/epdf/10.5772/59232>
- [11] John Wiley & Sons Ltd (2022), **Analysis on the mechanism of sound production and effects of musical flue pipe**, <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/ccs2.12048>

- [12] 蔡自偉 (2004) 。樂譜辨識 **Music score identification** 。
<https://image.cse.nsysu.edu.tw/2004student/Wei/Automatic%20Recognition%20of%20Printed%20Music%20Score.pdf>
- [13] 鱷魚(2014) 。鱷魚的手作笛子心得筆記-(2)邊稜音及笛子之發聲原理 。
<https://genepantom.pixnet.net/blog/post/43863640>
- [14] yunxingluoyun (2021) 。圖像去噪(圖像平滑) 。<https://reurl.cc/1vb6x9>
- [15] KUNG-LI LIN (2021) 。空氣真的能吸到“肚子”裡去嗎? <https://reurl.cc/VzMA3Q>
- [16] Clint Goss (2015), **How Flutes Work - Native American Flute**, <https://reurl.cc/9v3erj>
- [17] 陳中申 (1985) 。如何控制笛的音量、音高、音色。玉笛春風。
- [18] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia (2018) 。**Path Aggregation Network for Instance Segmentation**, <https://arxiv.org/abs/1803.01534>
- [19] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie (2016) 。**Feature Pyramid Networks for Object Detection**, <https://arxiv.org/abs/1612.03144>
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015) 。**Deep Residual Learning for Image Recognition**, <https://arxiv.org/abs/1512.03385>
- [21] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg (2015) 。**SSD: Single Shot MultiBox Detector**, <https://arxiv.org/abs/1512.02325>
- [22] Jae-Won KIM, Hochong PARK (2021) 。**Multi-Task Learning for Improved Recognition of Multiple Types of Acoustic Information**, <https://reurl.cc/vaogx1>
- [23] 廖威凱 (2007) 。基於類神經網路及樂理知識限制之印刷樂譜辨識系統。
<https://hdl.handle.net/11296/gwhfc2>
- [24] 蔡自偉(2004) 。印刷樂譜辨識系統。
<https://image.cse.nsysu.edu.tw/2004student/Wei/Automatic%20Recognition%20of%20Printed%20Music%20Score.pdf>
- [25] 李豪韋 (2016) 。自動樂譜辨識 **Automatic Music Score Recognition**,
<https://hdl.handle.net/11296/5wd5mb>
- [26] Jibao Qiu, C. L. Philip Chen and Tong Zhang (2022), **A Novel Multi-Task Learning Method for Symbolic Music Emotion Recognition**,
<https://www.semanticscholar.org/reader/9d0cb790952e76e1973e553b82f294888d15d97f>
- [27] Marketer(2021) 。當我們在談論 **Multi-Task Learning(多任務/多目標學習)**,
<https://hypergrowths.com/entrepreneur/digital-transformation/16734/topic-352428655/>
- [28] BotJunkie (2008). **Waseda Flutist Robot**,
https://youtu.be/jx8U1FgILCE?si=_QtJhz6IKF1Bc8tj
- [29] Yamaha Japan(2013) , **【フルート】フルートの発音原理** 。<https://reurl.cc/4rd1QD>
- [30] Clint Goss (2015), **How Flutes Work - Native American Flute**.
- [31] 330 ohms(2014), **atmega328-1**, <https://reurl.cc/nv1eY2>

附錄

評估檢測模型的性能時，Precision, Recall, mAP50, mAP50-95 都具有重要意義，反映了模型不同方面的性能。

1. Precision (精確度)

定義模型預測正樣本中實例為正例的比例，公式為：

$$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)}$$

2. Recall (召回率)

定義實際為正例樣本中模型正確檢測的比例，公式為：

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)}$$

3. mAP50 (mean Average Precision at 50% IoU)

當IoU閾值在 50%時，各類別平均精確度(AP)的均值，具體來說AP是 Precision Recall 曲線下的面積，mAP50 則是所有別AP的平均值，公式為：

$$mAP_{50} = \frac{1}{N} \sum_{i=1}^N AP_{IoU=50}^{(i)}$$

4. mAP50-95 ((mean Average Recision at IOU 50% to 95%))

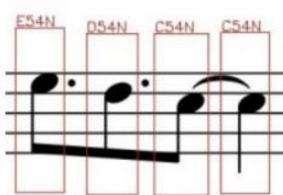
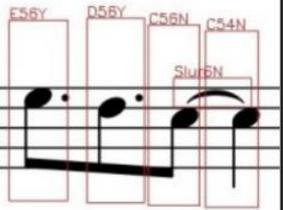
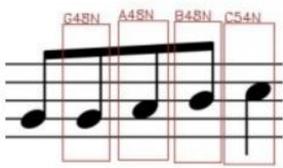
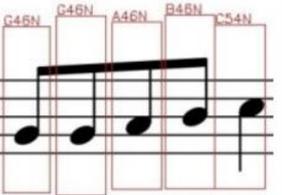
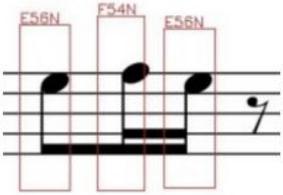
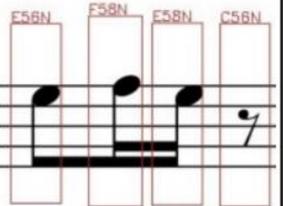
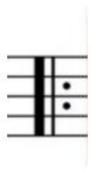
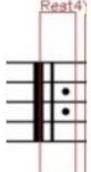
定義在多個IoU閾值(50%~95%, 5%步長)下，計算各類別AP的平均值，公式為：

$$mAP_{50-95} = \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{j=1}^K AP_{IoU=0.5+0.05*(j-1)}^{(i)}$$

上述方程式中， N 代表類別數量， K 是IoU閾值數量(0.50, 0.55..., 0.95 共 10 個)， $AP_{IoU}^{(i)}$ 中 i 代表第 i 個類別在IoU = 50%時的AP值。

以下是不同模型辨識結果，檢視不同版本辨識差異與改進：

表 25：模型版本差異（資料來源：研究者自行撰寫）

項目	V2. 2. 0	V2. 2. 4	說明	改進方法
音符組A			未辨識到附點，音長應為8分音符(6N, 有附點為6Y)結果都是4N且連音未標記。	加強附點數據與連音記號的辨識
音符組B			都是8分音符(6N)結果有辨識成8N及甚至未能標記出，後續版本已改進。	平衡分類標籤數量
音符組C			休止符未標記，且八分音符(8N)辨識成4N與6N，在後續版本已改進但休止符仍有部分分類錯誤。	擴充音符實例樣本並加強訓練
非目標組			應為反覆符號，但模型沒有這個樣本，所以誤判為休止符。	應擴充更多情況下的樂譜同時增幅至信閾值
整體	四分音符辨識以外辨識度差	精確度大幅上升，小部份仍錯誤		