

通 訊 工 程 學 系

碩 士 論 文

基於 CycleGAN 之古箏音樂風格轉換

Music Conversion for Chinese Guzheng Using CycleGAN

研 究 生：陳 秉 萱

指 導 教 授：張 寶 基 博 士

中 華 民 國 一 零 九 年 七 月

# 國立中央大學圖書館學位論文授權書

填單日期：109/07/21

2019.9 版

授權人姓名	陳秉萱	學 號	107553029
系所名稱	通訊工程	學位類別	<input checked="" type="checkbox"/> 碩士 <input type="checkbox"/> 博士
論文名稱	基於CycleGAN之古聲音樂風格轉換	指導教授	張寶基

## 學位論文網路公開授權

授權本人撰寫之學位論文全文電子檔：

- 在「國立中央大學圖書館博碩士論文系統」
  - (☒)同意立即網路公開
  - ( )同意 於西元\_\_\_\_\_年\_\_\_\_\_月\_\_\_\_\_日網路公開
  - ( )不同意網路公開，原因是：\_\_\_\_\_
- 在國家圖書館「臺灣博碩士論文知識加值系統」
  - (☒)同意立即網路公開
  - ( )同意 於西元\_\_\_\_\_年\_\_\_\_\_月\_\_\_\_\_日網路公開
  - ( )不同意網路公開，原因是：\_\_\_\_\_

依著作權法規定，非專屬、無償授權國立中央大學、台灣聯合大學系統與國家圖書館，不限地域、時間與次數，以文件、錄影帶、錄音帶、光碟、微縮、數位化或其他方式將上列授權標的基於非營利目的進行重製。

## 學位論文紙本延後公開申請 (紙本學位論文立即公開者此欄免填)

本人撰寫之學位論文紙本因以下原因將延後公開

- 延後原因
  - ( )已申請專利並檢附證明，專利申請案號：\_\_\_\_\_
  - ( )準備以上列論文投稿期刊
  - ( )涉國家機密
  - ( )依法不得提供，請說明：\_\_\_\_\_

• 公開日期：西元\_\_\_\_\_年\_\_\_\_\_月\_\_\_\_\_日

※繳交教務處註冊組之紙本論文(送繳國家圖書館)若不立即公開，請加填「國家圖書館學位論文延後公開申請書」

研究生簽名：陳秉萱

指導教授簽名：張寶基

\*本授權書請完整填寫並親筆簽名後，裝訂於論文封面之次頁。

國立中央大學碩士班研究生  
論文指導教授推薦書

通訊工程學系碩士在職專班 學系/研究所 陳秉萱 研究生  
所提之論文 基於 CycleGAN 之古箏音樂風格轉換  
係由本人指導撰述，同意提付審查。

指導教授 張寶基 (簽章)

109 年 07 月 09 日

國立中央大學碩士班研究生  
論文口試委員審定書

通訊工程學系碩士在職專班 學系/研究所 陳秉萱 研究生  
所提之論文 基於 CycleGAN 之古箏音樂風格轉換  
經由委員會審議，認定符合碩士資格標準。

學位考試委員會召集人

委

員

張文輝  
張寶基  
楊士瑩  
孫士華  
張文輝

中 華 民 國 109 年 07 月 09 日

# 基於 CycleGAN 之古箏音樂風格轉換

## 摘要

隨著科技日新月異，人們之間的距離因此縮短，文化的交流讓世界成為地球村，促使東西方的文化融合更加快速。在現代人的消遣娛樂當中，自從臺灣華語流行歌曲男歌手周杰倫與歌詞創作家方文山合力創作出《娘子》、《東風破》、《菊花台》等歌曲後，受到熱烈的回響，娛樂產業便刮起了一陣「中國風」。由於中國風的興起，更出現了西方樂曲被翻唱成中國風格的作品，翻唱（Cover）一直是樂壇中盛行之事，是由另外一位作者重新詮釋原作音樂，此種作為不僅限於歌唱，亦適用於跨樂器的詮釋，俗稱翻奏。

此篇論文，主要是提出一個基於圖像人工智慧的技術，來成就更好的音樂風格轉換，並且搭上時下流行的中國風，將音樂轉換成中國古箏的版本。透過人類感知實驗評估此方式的結果，獲得 4.3 的高分（0 分為最低分，5 分為最高分）。

**關鍵字：**深度學習、古箏、音樂風格轉換、CycleGAN

# Music Conversion for Chinese Guzheng

## Using CycleGAN

### **Abstract**

Chinese Guzheng music is popular from ancient Qin dynasty. However, from the study of modern Chinese history, Chinese people continued to pursue westernization. The Chinese Guzheng also faces the drastic changes of the external environment such as the needs of modern national orchestras. Due to changes, the Guzheng is reformed in the direction of expanding the sound range, increasing the volume. Although new design for the Chinese Guzheng instrument has been proposed, it is too hard for players to catch the music transition because of long time producing and learning new skills.

To improve this issue, we propose music conversion with CycleGAN which is a deep learning technique that involves the automatic training of image-to-image translation models without paired examples. Moreover, this method provides a simple way which can be used by whom never learn any instrument. This is a brand new research with good results.

***Keyword: Deep Learning, Chinese Guzheng, Music Conversion, CycleGAN***

# 誌謝

首先，特別感謝我的碩士班指導教授張寶基教授，提供我專業領域上的建議。除此之外，兩年的碩士生涯，張教授也是我的心靈導師，在他的帶領與鼓勵之下，讓我有信心繼續於博士班深造。

感謝實驗室的梁凱雯學姊，協助我一起完成 ICCE-TW 的投稿，提供我許多專業的建議。也非常感謝實驗室的所有成員，不論是畢業的學長姐、同屆的同學或是學弟妹，都在我求學階段給予我相當多的幫助。

最後我要感謝我的家人與所有支持我、陪伴我的人，謝謝你們以一個開放的心態包容與支持我的選擇。

# 目錄

中文摘要	I
Abstract	II
致謝	III
目錄	IV
第一章 緒論	1
1-1 研究背景	1
1-2 研究動機	2
1-3 論文架構	3
第二章 風格轉換與中國古箏音樂介紹	4
2-1 翻唱、翻奏藝術之探討	4
2-2 中國古箏與音樂介紹	5
2-2-1 中國古箏樂器介紹	5
2-2-2 中國古箏音樂派別介紹	7
第三章 深度學習相關介紹	12
3-1 類神經網路	13
3-1-1 發展歷史	13
3-2 深度學習的訓練	14



3-2-1 深度學習的架構	15
3-2-2 深度學習的應用	17
第四章 實驗設計	18
4-1 音樂數據收集與前處理	19
4-2 訓練階段	21
第五章 實驗結果與分析討論	23
5-1 數據集介紹	23
5-2 評分方法	24
5-3 實驗結果	24
第六章 結論與未來展望	25
參考文獻	26
附錄	28

# 第一章 緒論

## 1-1 研究背景

古風音樂創作 [1]，是一種將現代與中國傳統相互結合的新興音樂。其擁有大量中國傳統元素，以及古典的歌詞、優美的旋律，而且非常注重意境。大多數會使用民族樂器當伴奏，以保留傳統音樂的韻味。

古風音樂不僅運用在網路遊戲的氣氛渲染，自從臺灣華語流行歌曲男歌手周杰倫與歌詞創作家方文山合力創作出《娘子》、《東風破》、《菊花台》等歌曲後，受到熱烈的回響，娛樂產業刮起了一陣「中國風」。古風音樂因網路的蓬勃發展而誕生，更有古風愛好者將網路中國風遊戲的配樂重新編曲翻唱，「翻唱」（英語：Cover）[2] 一直是樂壇中盛行之事，是將原歌曲改由他人以另一種方式重新詮釋，而翻唱的樂曲也相當有競爭力，1952 年，《芝加哥論壇報》（Chicago Tribune）對該術語進行了描述：「Trade jargon meaning to record a tune that looks like a potential hit on someone else's label.」。

## 1-2 研究動機

近年來，因為社群網站的興起，翻唱藝術更加蓬勃，其形式有很多種，例如：編曲者二度創作並且以歌唱的形式表現，或者是樂器翻奏，甚至用電腦重新編輯音樂檔案。

然而，現今人類越來越注重娛樂品質，因此音樂作品的產出量相較於過往更大。雖然音樂的人類創作有不可取代的價值，但卻無法在短時間內大量創作。另外，人類的音樂創作行為，往往侷限於接受過音樂訓練的人。

因此，此篇論文將提出一個新穎的方式，來加速音樂翻奏的產出速率，為搭上中國風的音樂風潮，特別應用於翻奏到中國古箏音樂。其音樂的轉換方式是基於人工智慧的模型，可批量處理音樂數據，也適用於沒有音樂基礎卻想翻奏出古箏音樂的使用者。

## 1-3 論文架構

本論文共分為六個章節，第一章為緒論，介紹本篇論文研究的背景、動機及目的。第二章為翻唱藝術與方式之探討，以及中國古箏音樂的介紹。第三章為深度學習相關介紹，介紹深度學習的歷史、架構與現代應用。第四章為介紹本篇論文所提出的基於人工智慧應用於古箏風格轉換的實驗設計。第五章為實驗結果以及相關的分析與討論。第六章為本篇論文研究的結論及未來展望，探討提出方法的優勢與貢獻，並討論未來的改善及應用方向。最後為參考文獻與附錄。

## 第二章 風格轉換與中國古箏音樂介紹

### 2-1 翻唱、翻奏藝術之探討

翻唱（英語：Cover），是將已經出版過的歌曲，由其他歌手重新詮釋，也被稱之為口水歌。由於翻唱音樂是將老歌新唱，基於熟悉的旋律添加新的元素，令聽眾耳目一新。

如今，用歌唱重新詮釋一首新樂曲，稱之為「翻唱」。後來也陸續出現用樂器重新詮釋一首歌唱音樂，稱之為「翻奏」。兩者有異曲同工之妙，但前者為人聲演唱，後者為樂器演奏。

近年來由於網路平台以及歌唱節目興起，有許多素人歌手從事翻唱藝術，特別要注意的是，不管是修改原著的歌詞或是修改原著的曲調，基本上都需要獲得原著的同意，或者符合法律上合理使用的範圍，否則需要承擔法律責任 [3]。

## 2-2 中國古箏與音樂介紹

### 2-2-1 中國古箏樂器介紹

古箏 [4]，是中國歷史悠久的彈弦樂器。古箏的起源，根據專家的探討，某些學派的學者認為極有可能是從瑟演變而來，漢代古箏的外型以及琴弦的編制，幾乎可視為相同，而瑟在漢代之後逐漸勢微，古箏則在魏晉南北朝之後的記載陸續增加，這項樂器可說是保存了瑟的編制，而冠上古箏的名。

箏的琴弦數量，由古至今有五弦、六弦、七弦、九弦、十弦、十一弦、十二弦、十三弦、十四弦、十五弦、十六弦等。隨著古箏演奏曲的內容日益豐富，對於音域有新的需求，近代改良過的古箏琴弦擴增至有二十一弦、二十三弦、二十五弦等規格。而琴弦的材質，古代有蠶絲製成，也有金屬琴弦，而目前流行的琴弦為尼龍材質纏於金屬弦之外，可以保護琴弦，延長使用壽命，還可以上色方便演奏辨認琴弦。現代箏體是一個由木板構成的共鳴箱，外觀呈扁長型但有些弧度，底部擁有兩到三個音孔。框板的材質甚多，不同的木頭材質，可呈現的音質也不同，最好的木頭為紫檀木，也可用紅木以及花梨木等等，外部可能會鑲嵌貝殼，或是在木頭上做雕飾。古箏的頭尾有前梁和後梁。琴面則有琴弦，側邊則擁有調音的軸。古箏的面板上，每條琴弦都擁有一個琴柱，因為

其外觀猶如群雁在飛翔，故又被稱為雁柱，雁柱是可移動的，用以取得不同的音高，雁柱的材料，會用材質較硬的材料製成。古箏是以五聲音階定弦，也可以調成七聲音階，也有其他定弦方法，音域跨了三個八度。

彈奏古箏時，需要使用義甲，現代較常見到的是使用玳瑁製作的瓜子形義甲。傳統的古箏技法，是「右手彈聲，左手表韻。」但由於現代箏曲日益複雜，有時左手也須右移彈奏 [5]。



圖 1. 古箏示意圖

## 2-2-2 中國古箏音樂派別介紹

此段落將針對不同的箏派，介紹其發源地、歌曲特色、著名曲目等。依序會介紹五種箏派：山東箏派、河南箏派、陝西箏派、浙江箏派與潮州箏派。

### 一、山東箏派

山東箏曲 [6] 又稱「齊魯箏曲」，其特色為剛勁有力，以及古樸、典雅的演奏風格。山東箏派的代表曲目有：《四段錦》、《漢宮秋月》等。山東箏派的左手傳統按弦技法，主要特點是「虛點實按，揉拈攝空」，山東箏派較多地使用滑音，尤其是上滑音，演奏滑音時的速度也較快，左手作韻時注重餘音的處理，強調滑按之音的細微變化，這又使得有些山東箏曲具備了古樸典雅的風格。山東箏派的這一演奏特色與山東琴書的特點及山東的方言有著密不可分的關聯，山東箏派傳統樂曲多為歡快的情緒。

### 二、河南箏派

河南箏 [6] 的表現，是箏樂中影響力較大的一個派，集敘事、狀物、抒情於一體。河南箏派與山東箏派一樣，都是北方箏派，曲調非常富有



歌唱性，於清新流暢、活潑開朗中又顯露出抑揚頓挫、雄壯豪邁的韻味。其樂曲風格高亢、粗獷，與河南人的性格和語言風格相似，且受到戲曲音樂和民間說唱音樂的深刻影響。在旋律的進行中，左手的慢滑急顫常常都是與右手的游搖相配合而完成的，在演奏中要特彆強調按滑音時音色的厚重感，指法中最有特色的就是左手的滑按、小顫、滑顫、大顫等技巧了。河南箏派代表作品有《蘇武思鄉》、《河南八板》、《打雁》及由戲曲故事《二度梅》改編而成的箏曲《陳杏元落院》、《陳杏元和番》等樂曲。

### 三、陝西箏派

陝西秦箏 [7] 的歷史可回溯到公元前一百多年前，陝西是古箏的發源地，在陝西榆林一代，古箏大多是拿來伴奏用，古箏在樂隊裡用來掌握節奏跟美化旋律，使音樂聽起來豐富而細膩。在古代用於辦堂會、唱小曲來喧鬧氣氛，陝西箏派旋律的上行跳進使人的情緒高亢激昂、起伏多變，下行級進又使得人的情緒深沉厚重、哀怨纏綿，如：《秦桑曲》、《姜女淚》、《迷糊調》、《香山射鼓》等。陝西箏，陝西地區是中國箏的發源地，但現在“真秦之聲”的箏樂卻幾乎近於絕響，這裡有豐富多采的戲曲和民間音樂，其間傳統的慣用樂器均不為少，但唯獨無箏或極少用箏；只有在榆林地區才用箏作為伴奏樂器，跟洋琴，琵琶，三弦

一起在榆林小曲的伴奏中出現。著名秦箏理論家，教育家曹正先生說，這就是“秦箏的餘緒”。箏在榆林雖有三百多年的歷史，但榆林地處偏僻，文化比較落後，箏的技藝還停留在古法彈奏上，右手主要用大食兩指，左手滑顫音很少，這正是箏的最初奏法。箏在陝西地區已漸漸失傳，其傳譜傳人都無資料可尋。20世紀50年代末期，陝西箏家提出了“秦箏歸秦”的口號，從理論研究到彈奏技藝，從伴奏地方戲曲到編制秦韻風格箏曲，做的工作，使陝西人民重新熟悉 1957 年，榆林箏名家白葆金參加了全國人民的民間音樂調演和陝西省第三屆民間戲曲匯演，並獨奏了《掐蒜苔》，《小小船》箏曲。1961 年全國古箏教材會議在西安召開以及會議對陝西迷胡箏曲的肯定，都對秦箏在陝西的發展起了有力的促進作用。後來湧現了像《秦桑曲》，《姜女淚》，《香山射鼓》，《三秦歡歌》，秦腔和迷胡清曲等史料中關於箏的敘述，給陝西箏人以啟示，後來又有箏名家曹正，高自成，王省吾等人的巨大努力，眾多箏同仁的幫助，經過近三十年的實踐，陝西已漸漸地恢復箏奏秦聲。七聲階中的四級音偏高，七級音偏低。所謂偏，當然不是半個音。這是陝西風格箏曲的鮮明特色，首先是音律上的特殊性和二個變音的游移性。一般來說，是向下滑動緊靠下一級音的；其次，在旋律進行上，一般是上行跳進，下行級進的。第三，在彈奏時的左手按弦使用大指較多，這是出於旋律進行需要而必然使用的技術。第四，風格細膩，委婉中多悲怨；慷慨急

楚，激越中有抒情。已故著名浙派箏家王巽之先生談古箏流派時說“陝西派多抒情”，說明了陝西箏曲的風格特點。

#### 四、浙江箏派

浙江箏派 [8] 又可以稱為「武林箏派」，雖然它起源於浙江的杭州地區，但真正盛行的地方卻是在上海。浙江箏派在中國多個古箏流派中也具有廣泛的影響力。浙江箏派在技法上除了繼承傳統的技法外，積極的借鑑揚琴、琵琶甚至是一些西方樂器的演奏技法，這就使得浙江箏派在演奏上有所突破，另人耳目一新。浙江箏派演奏上要求以右手為主，與其他箏派很不同的是不太強調左手的「以韻補聲」，傳統的浙江箏曲比較忌諱雙手繁鬧的彈奏手法，所以也就形成了含蓄、淡雅、秀美、清麗的風格。不過，儘管如此，浙江箏也還有著它獨特的左手作韻技法，這就是「快速點滑」。「快速點滑」的裝飾按音要迅速、乾淨利落，不可拖泥帶水。這就形成了浙江箏曲的演奏風格。浙江箏派的代表曲目有《將軍令》、《月兒高》、《四合如意》、《雲慶》、《三十三板》、《燈月交輝》等。

## 五、潮州箏派

潮州箏派 [9] 典雅、輕柔、細膩，體現於「輕六」、「重六」、「活五」。潮州箏的起源主要可分成兩種，一種是潮州絃詩樂，例如：《寒鴉戲水》、《昭君怨》、《錦上添花》，另外一種是潮州細樂，以清新細膩著稱，例如：《胡笳十八拍》。

### 第三章 深度學習相關介紹

早在 1950 年代，人工智慧的思維已經出現，當時創造出比人腦更強的運算電算機。而近年來人工智慧的議題再度受到熱議，因為 AlphaGo 打敗了知名圍棋手。許多科技大公司陸續投入了深度學習的研究，例如：Google、Facebook，深度學習 Deep Learning [10] 默默的成為人類生活中不可或缺的技术，目前的應用範圍有：自然語言處理、推薦系統、生醫資訊、影像辨識、語音辨識。

傳統的機器學習 Machine learning，需要各領域的資料科學家，對資料做深入的理解與研究，進而創造出好的機器學習演算法，此方式就是 Feature engineering。深度學習是其中的一個分支，可以自動抽取特徵是它與機器學習最大差別之處。早在 90 年代深度學習已經出現，但由於電腦的運算能力不夠強大，所以在當時效果不彰。深度學習希望把資料透過多個層中的線性或非線性轉換而自動找出特徵，所以被視為一種特徵學習，大大提升效率。

## 3-1 類神經網路

類神經網路是科學家受到生物神經系統的啟發，創造的數學模型，每個神經元都有輸入與輸出，將一大堆這樣的神經元相互連接，而成為了類神經網路。而這些神經元通常有好幾層，每一層緊密相連。而整個類神經網路的數量、階層、連接方法與使用哪一種活化函數的類型，這些參數設定攸關整個類神經網路的效能表現，簡而言之，整個類神經網路的學習過程，就是在調整到最好的參數設定。

### 3-1-1 發展歷史

早在 1960 年代，資料科學家藉由了解及模仿動物的神經傳導，將這樣的技術實現在電腦上，因而提出多層的類神經網路（Artificial neural network）。然而，在那個年代受限於電腦的計算能力，及不易取得大量數位資料，所以類神經網路在當時並沒有蓬勃發展。在 1980 年到 2010 年，類神經網路並未在此時有突破性的應用，但開發出許多不同的類神經網路架構，例如卷積神經網路（Convolutional neural network, CNN）、受限波茲曼機（Restricted Boltzmann machine, RBM）、遞歸神經網路（Recurrent neural network, RNN）等等。

2000 年後，深度學習開始被使用在影像辨識，特別是運用在辨識手寫的數字。在 2010 年之後，深度學習更是在 MNIST（一個手寫數字的資料集）的判斷上，超越人類的精準度。

而在近幾年，深度學習的運用越來越廣闊，不僅僅只是應用傳統的機器學習來做分類與回歸的處理，也被運用在降維，或者是讓電腦自動對話、寫詩、製圖等仿真人的應用。

### 3-2 深度學習的訓練

深度學習的訓練可分為三步驟：第一，定義網路架構；第二，定義學習目標；第三，透過數值方法進行訓練。

「定義網路架構」就是先選出一些可能的函數，接著來進行深度學習訓練。定義了適當的網路架構，才能經由訓練來產生一個好的深度學習模型。

「定義學習目標」就是用一個數值，來定義學習的好壞，又被稱之為適性函數或者是目標函數，這對機器學習與深度學習皆是重要的一環。

實際的訓練過程就是要找出最好的參數組，讓學習目標的數值越小越好。在深度學習中，可用隨機梯度下降法的變型減少終止在 Local optimum 的可能，這樣能使深度神經網路的學習效果更好。

### 3-2-1 深度學習的架構

卷積神經網路（Convolutional Neural Networks）因為利用 Convolutional layer 以及 Pooling layer 來強化 Pattern recognition 與周邊資料的關係，所以是在處理聲音與影像的資料可以得到好的效果，它也是很常見的深度學習架構。例如：運用深度學習來玩電子遊戲時，常會使用卷積神經網路，來分析影像的內容，像是 AlphaGo，使用了變化過的卷積神經網路結合蒙地卡羅樹搜尋演算法，得到令人振奮的結果。

卷積層是卷積神經網路最重要的部份，是一組平行的 Feature map，把圖上的每一個點當作中心，取四周  $n \times n$  變成一個 filter，每個 filter 會對不同的 Image pattern 進行強化，訓練的過程中，同時會強化 filter 的影像模式，所以不同的問題，卷積層會產生出不同的 filter 效果。

在卷積層之後，通常會接池化層，它與訊號處理中的 Down sampling 相似，可在保留重要特徵的基礎之下，縮減整個模型的大小，



以提高計算的速度。池化層大多用於影像分辨的神經網路，常見的動作為 Max pooling。

遞歸神經網路（Recurrent Neural Network）是比 Feedforward Neural Network 更加進階的神經網路，其與傳統的類神經網路相當不同。遞歸神經網路計算後的輸出，又會回傳給該層作為自己的輸入，它的神經元內有暫存的記憶空間，可把先前輸入的資料產生的狀態儲存在暫存空間內，再根據之前的狀態計算而輸出不同的值，所以遞歸神經網路是有記憶性的，特別適合處理時間序列（Time series）相關的問題，例如：自然語言處理以及語音辨識等應用。

遞歸神經網路看似方便，但在實際訓練上卻存在這一些問題，例如：越前面的資料對於越後面的資料決策影響越小，當中間經過的時間序列過多，前面的資料造成的影響可說是趨近於零。另外，遞歸神經網路權重組合的空間形狀對不利於隨機梯度下降法，會產生不穩定的結果，平緩的地方容易停留在局部的最佳解，產生梯度消失（Vanishing gradient）的問題，而非常陡峭的部分，極可能讓隨機梯度下降法在更新後的數值超出正常的範圍。

前段提到，RNN 越前面的資料對於越後面的資料決策影響越小，這時就會需要長短期記憶神經網路（Long-short term memory, LSTM）[11]，就是利用三個控制閥（Gate）來決定記憶的去留，它們擁有各自的參數，

會依照輸入的資料，經過權重的計算後，來決定控制閥得開關，分別是寫入（input）、遺忘（forget）、輸出（output）。

### 3-2-2 深度學習的應用

目前人工智慧的應用主要分成三大領域：一、自然語言處理。二、影像識別。三、聲音辨識。

在自然語言處理的部份，如 Apple 公司的 Siri，Google 公司的 Google Assistant，Amazon 的 Alexa 等語音助理服務陸續問世，都是自然語言處理的技術，希望電腦能夠如人類的正常對話。

影像辨識的部分，例如：中國大陸的天網，可以快速且正確地利用臉部辨識抓犯人、又或者是對於人類行為的監控，監控家中老人是否跌倒等等。

聲音辨識，則有應用在聲音的情緒辨識，或是曲目推薦系統，又或者是場景辨識。

## 第四章 實驗設計

此章節針對數據的收集與前處理，以及訓練參數的設定做詳細介紹。

在 4-1 節將介紹對數據的前處理，在 4-2 節將介紹訓練階段。

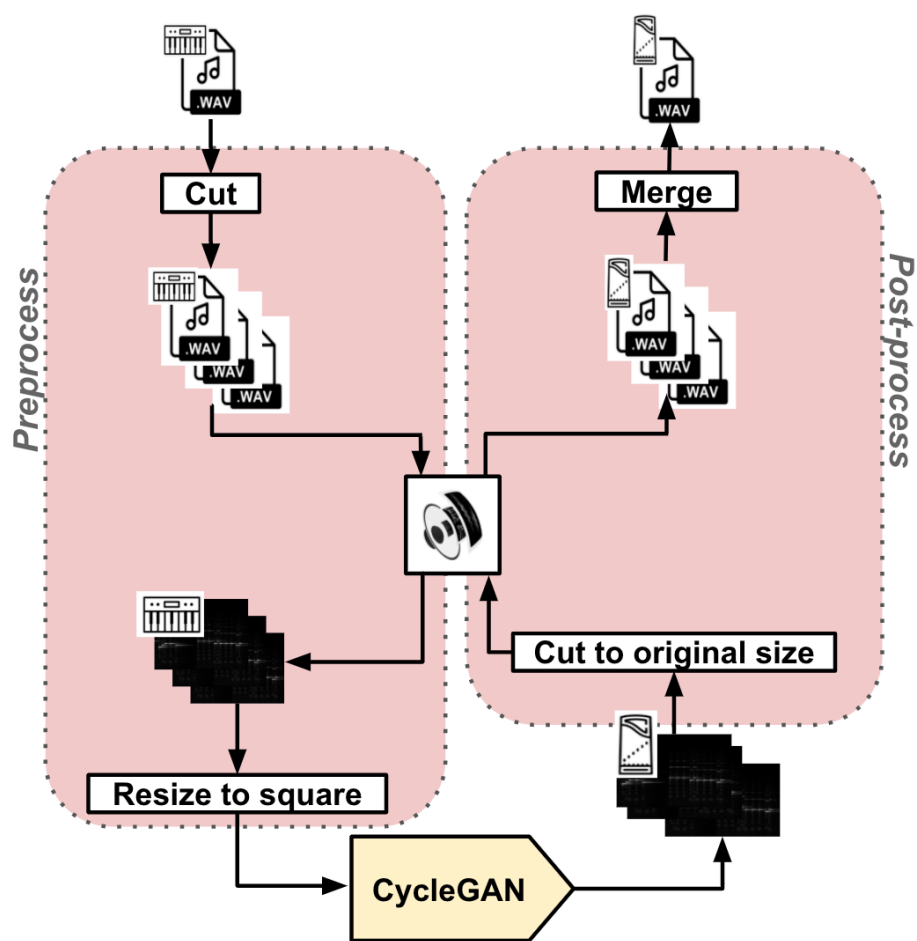


圖 2. 以 CycleGAN 做古箏音樂風格轉換流程圖

## 4-1 音樂數據的收集與前處理

由於 CycleGAN 是圖像對圖像的轉換，所以我們必須將收集到的音樂檔案妥善地轉換成圖像的檔案。在前處理的部份，首先，我們以錄製的方式收集到音樂檔案，並且將收集到的 wav 檔案，以每拍為單位做切割。切割完成之後，接著再使用 Photosounder，將檔案轉換成頻譜，Photosounder [12] 是第一個採用完全基於圖像的音樂創建和編輯方法的音頻編輯器或合成器。因為其強大而無所不能的合成算法，它能夠產生任何可能的聲音。最後，由於方便 CycleGAN 的訓練，若圖像並非邊長為 4 像素的倍數，我們將圖片以黑色畫面將其補足，原檔是以置中的方式貼在黑色方形圖片上（請參考圖 3, 4）。

在後處理的部份，首先，將 CycleGAN 產出的圖片截回原本的大小，再透過 Photosounder 將圖片模擬回聲音檔，最後再將以 1 秒為單位的 wav 檔利用 Audio joiner [13] 合成為完整的曲子。

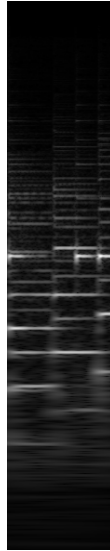


圖 3. 以 1 秒為單位的 wav 檔經由 Photosounder 轉換出的 bmp 檔

( Dimensions: 100\*571 )

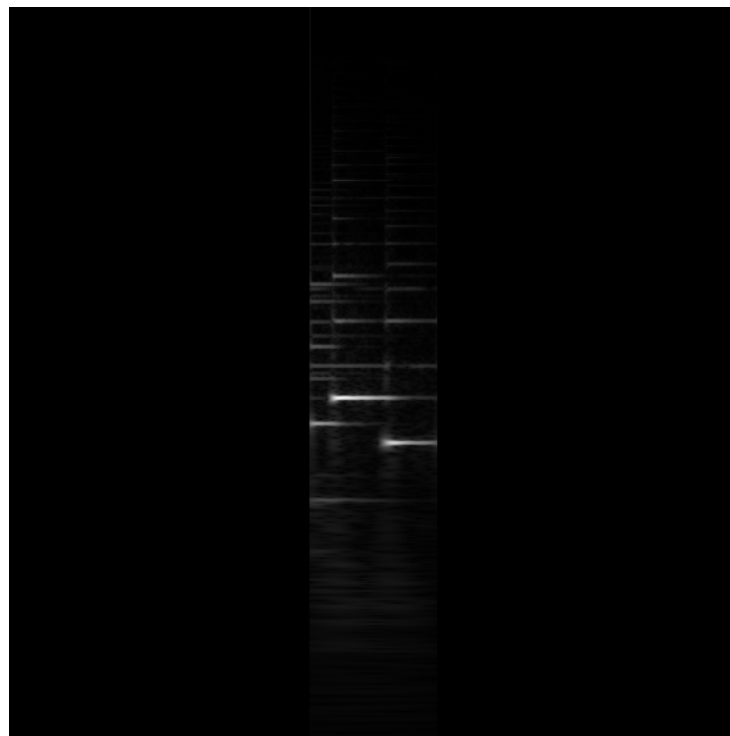


圖 4. 經過黑底補圖過的 bmp 檔 ( Dimensions: 1024\*1024 )

## 4-2 訓練階段

在這篇論文中，我們應用 CycleGAN [14] 生成古箏音樂。為了確保產生的質量，這模型結構包括生成器和鑑別器。這兩種模式一直試圖互相挑戰，直到生成器創建的數據與真實數據幾乎沒有區別的數據集。

$G_{\text{SynthPiano\_Guzheng}}$  的生成器，它有助於將圖像從模擬鋼琴領域轉移到古箏域。另一個生成器稱為  $F_{\text{Guzheng\_SynthPiano}}$ ，它有助於將圖像轉換回合成鋼琴領域。生成器是編碼器與解碼器之模型架構。該系統包括兩個 stride-2 卷積，用於訓練圖像的九個殘差塊從合成鋼琴到中國古箏的轉換，高於  $1024 \times 1024$  分辨率和兩個 1/2 步幅卷積步幅。剩餘的塊是確保之後的層有保留先前圖層的屬性。對於區分者  $D_{\text{Guzheng}}$ ，是  $70 \times 70$  的 PatchGAN 用於分類圖像補丁之真假（式 1）。循環一致性損失是使用輸入到  $G_{\text{SynthPiano\_Guzheng}}$  的圖像與  $F_{\text{Guzheng\_SynthPiano}}$  輸出的圖像進行相應更新以減少圖像的差異（式 2）。完整的目標函數如（式 3），建議  $\lambda$  設置 10，所有網絡的學習率均為 0.0002。因為使用來自源域和目標域在任何情況下都不需要關聯，這樣的建構很適合應用於音樂風格翻奏，因為此類課題的自源域與目標域不容易找到相對應的資料。

$$Loss_{adv}(G, D_y, X, Y) = E_{y \sim P_{data}(y)}[\log Y(y)] + E_{x \sim P_{data}(x)}[1 - D_y(G(x))] \text{----} (式 1)$$

$$Loss_{cyc}(G, F) = E_{x \sim P_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)}[\|F(G(y)) - y\|_1] \text{----} (式 2)$$

$$Loss_{full} = Loss_{adv} + \lambda Loss_{cyc} \text{----} (式 3)$$

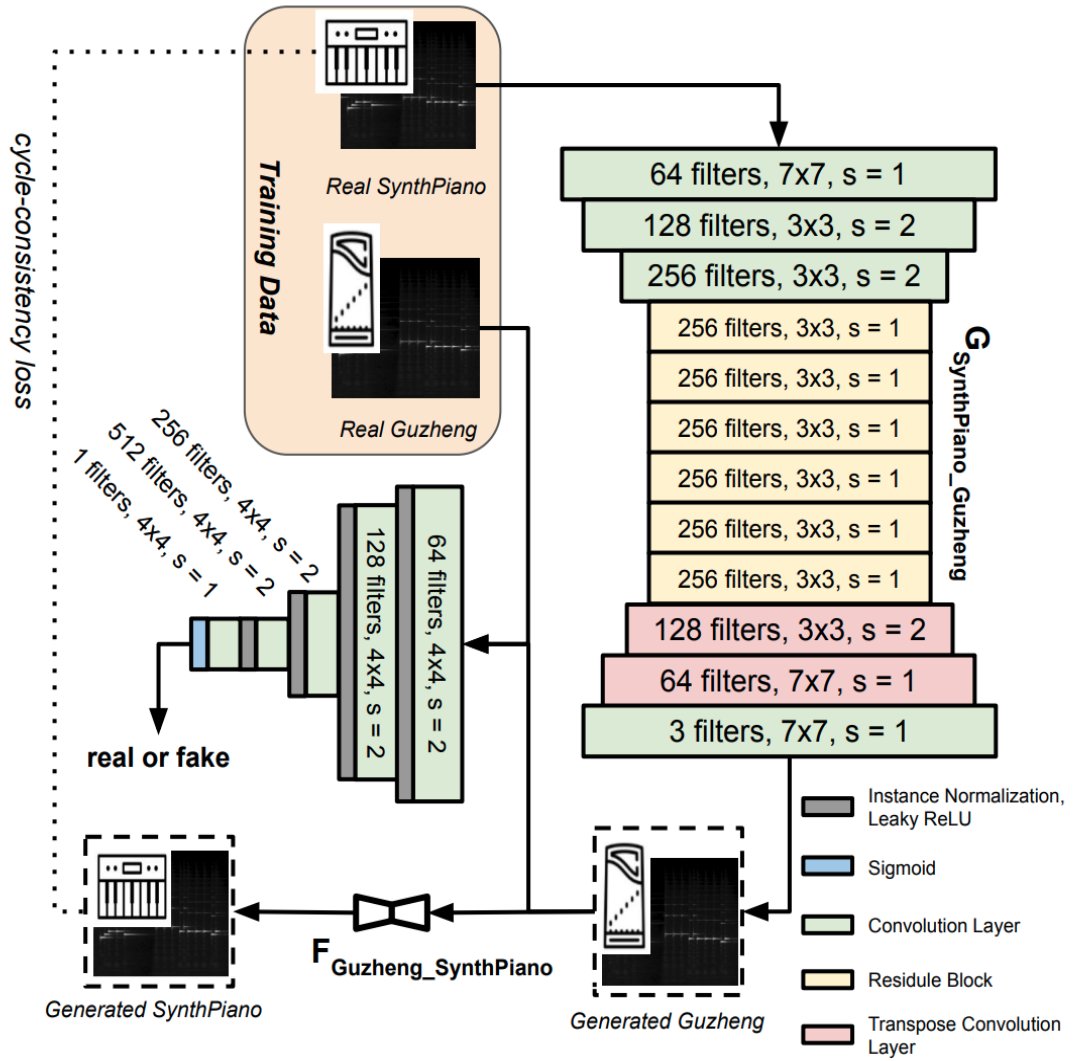


圖 7. 訓練架構圖

## 第五章 實驗結果與分析討論

本章節將介紹本研究所提出之方法的實驗結果，並進行相關的討論分析和比較。主要可以分為三個部分，5-1 數據集介紹，5-2 評分方法，5-3 實驗結果。

### 5-1 數據集介紹

在本實驗，總共收集了五首陝西風格的古箏曲，以及相對應的鋼琴版本，為了讓模型可以學到更好的陝西腔，另外也特別收錄陝西按放音的練習曲。為了增加資料量，除了收錄樂曲原調音樂，另外也收集了上下增減 1 key 的版本。

表 1. 訓練收錄的曲目

曲目	曲長
<秦桑曲>	5'24"
<百花引>	5'28"
<懨>	9'06"
<雲裳訴>	12'54"
<秦土情>	8'20"
<按放音練習曲>	3'19"



## 5-2 評分方法

我們採用人類的感知來調查本文的方法是否具有保留音樂作品的能力。問卷讓實驗對象聽取我們所設計的方法產生之古箏音樂，受試者總共有 40 位，分別是 20 位曾經學過古箏，以及 20 位從未學習過古箏。然後由受試者依照他們對於中國古箏音樂的印象來比較相似度，給予一個分數，分數從 1 到 5，分數越高意味著更高的相似性。

## 5-3 實驗結果

我們將 20 位曾經學過古箏的受試者對於相似度的給分做平均，得到了 4.1 分，而 20 位不曾學過古箏的受試者之給分平均，得到了 4.5 分，而最後得到的總平均為 4.3 分。

Score \ Subject	20 people <i>have experience for playing Guzheng</i>	20 people <i>no experience for playing Guzheng</i>	Total experiment subject
Average score	4.1	4.5	<b>4.3</b>

表 2. 評估結果

## 第六章 結論與未來展望

在本文中，我們提出了一個新穎的古箏音樂風格轉換方法。由實驗評分可得知此方法具有產生良好結果的能力。而且，即使不會演奏古箏的人，都可以使用此方法產生中國古箏的音樂。關於未來展望，因這個模型只能從模擬鋼琴轉換到古箏音樂，希望在未來能夠探索更靈活的架構，以適應更多的樂器。並且，也希望未來可以使用人工智慧的方式來解決本論文聲音與圖像相互轉換的部份。而在資料收集的部份，此篇論文所收集的音樂並未在專業的錄音室收錄，期望在未來能夠使用更專業的器材與空間，以避免外在干擾以及聲音反射等問題。

## 參考文獻

- [1] 何孟翰，數位音樂創作中編曲配器之探討：以〈塵煙〉一曲為例談古風歌創作，<https://hdl.handle.net/11296/dyu3mv>，2018。
- [2] 趙麗妍，當代台灣國語翻唱歌曲現象研究（1990~2011），2012。
- [3] 林育嬋，改作權問題之研究，<https://hdl.handle.net/11296/pxahsy>，2014。
- [4] 張晏萍，新型式古箏製作研究，<https://hdl.handle.net/11296/4c8bne>，2016。
- [5] 黃金燕，古箏演奏法在箏樂流派之運用，2014。
- [6] 李怡瑩，說唱音樂在傳統箏曲上的應用-以山東箏派及河南箏派為例，2018。
- [7] 黃暉閔，當代陝西箏派作品之研究—以《姜女淚》、《雲裳訴》、《秦土情》為例，2013。
- [8] 許紫庭，浙江箏派的演奏技法對當代箏曲創作的影響，<https://hdl.handle.net/11296/2n8f2f>，1998。
- [9] 洪涇綾，從《柳青娘》各家版本演奏譜探討潮州箏樂的彈奏藝術，<https://hdl.handle.net/11296/j46aaq>，2000。
- [10] 張榮麗，訓練深度學習網路之探討，<https://hdl.handle.net/11296/53dtbs>，2017。
- [11] S.Hochreiter, J.Schmidhuber, “Long short-term memory Neural computation,” 9(8):1735–1780, 1997.
- [12] Photosounder. Retrieved from <http://photosounder.com/>
- [13] Audio joiner. Retrieved from <https://audio-joiner.com/>

- [14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” *Proc. Int. Conf. Computer Vision*. [Online], Available: <https://arxiv.org/abs/1703.10593>, 2017.
- [15] G. Brunner, A. Konrad, Y. Wang, and R. Wattenhofer, “MIDI-VAE: Modeling Dynamics and Instrumentation of Music with Applications to Style Transfer,” *arXiv :1809.07600*, 2018.
- [16] Si. Huang, Q. Li, C. Anil, X. Bao, S. Oore, and R. B. Grosse, “TimbreTron: A Wave Net(CycleGAN(CQT(Audio))) Pipeline for Musical Timbre Transfer,” *arXiv:1811.09620*, 2018.

# 附錄

## 附錄一：以每拍為單位做切割的程式碼範例

```
from pathlib import Path
import argparse
from pydub import AudioSegment
import wave
import contextlib
import glob
import os

def split(input, output, segment_duration):

    for file in os.listdir(input):

        current_file = os.path.join(input, file)
        sound = AudioSegment.from_file(str(current_file))
        base = os.path.basename(str(current_file))
        frames = sound.frame_count()
        rate = sound.frame_count(ms=1000)
        duration = int(frames / float(rate))
        start_time = 0 - segment_duration
        end_time = 0
        split_file_num = 0

        while start_time < duration * 1000 - 2 * segment_duration:
            while end_time < duration * 1000 - segment_duration:
                while split_file_num <= duration * 1000 / segment_duration:
                    start_time += segment_duration
                    end_time += segment_duration
                    split_file_num += 1
                    sound[start_time:end_time].export(str(output) + "/" + str(base) + '_seg'
+ str(split_file_num) + '.wav', format = "wav")
                    print_name = str(base) + '_seg' + str(split_file_num) + '.wav'
                    print(print_name, start_time, end_time)

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument('--input', type = Path, required = True,
                        help='The directory of file you want to split!')
    parser.add_argument('--output', type = Path, required = True,
                        help = 'The directory you want to put the finished-split data')
    parser.add_argument('--ms', type = float, default = 100,
                        help = 'Milliseconds of segment duration')

    args = parser.parse_args()
    split(args.input, args.output, args.ms)

if __name__ == '__main__':
    main()
```

## 附錄二：BMP 檔黑底補圖程式碼範例

```
from PIL import Image
import os

def make_square(im, min_size=1024, fill_color=(0, 0, 0, 0)):
    x, y = im.size
    size = max(min_size, x, y)
    new_im = Image.new('RGBA', (size, size), fill_color)
    new_im.paste(im, (int((size - x) / 2), int((size - y) / 2)))
    return new_im

x = 0
for file in os.listdir("./"):
    try:
        original_image = Image.open(file)
        new_image = make_square(original_image)
        new_image.save('P_square_' + str(x) + '.bmp')
        print(x)
        x += 1
    except:
        pass
```