

第十九屆旺宏科學獎

成果報告書

參賽編號：SA19-226

作品名稱：從分類到溝通——

以機器學習分辨鳥鳴聲

姓名：洪軾凱

關鍵字：Bird Sound Classification、
Attention Mechanism、
Machine Learning

摘要

鳥類的鳴聲除了親緣關係有影響外，更重要的是行為學習。如果要進行鳥類的行為辨識與物種分類，鳥類的鳴叫聲便是重要關鍵，並且因為鳥類的聲音很容易趨同演化，甚至還有鳥類學習其他鳥種的狀況發生，因此對於鳥種利用聲音的溝通行為，用人耳來辨識是一件極困難的工作。本研究試圖利用機器學習來辨識同一種鳥類些微聲音的差異所進行的溝通行為。為了證明鳥鳴聲在不同生理時期間及不同鳥種有不同的聲音特徵，我們鳥類的生理時期分辨資料集是利用自鳥人孫清松先生花費多年所錄得之斯氏繡眼資料集來分析，而鳥類辨識是由 xeno-canto.org(XC)網站所收集的聲音資料集來嘗試鳥類種類分辨。我們將斯氏繡眼資料集分為七種不同生理時期，而 XC 資料集則是取下自 XC 網站中 B grade 前 18 大錄音資料的鳥類物種，經去除雜音、切割及經 Mel Spectrogram 及 MFCC 轉換為時頻圖後，以 Convolutional Module 為基底，各自接上 LSTM、GRU 及 Attention 的模型，和以 BERT 為基底的模型進行機器學習實驗，並嘗試將其中的 Feed Forward Network 改為 Conv1D 的模型。實驗結果發現藉由改為 Conv1D 以 BERT 為基底的模型效果最好，而且鳥類行為分析與分類兩種問題 macro F1-score 的正確率最高可達到 97.7% 和 79.0%，並且透過 t-SNE 降維法和 Visualized Attention Map，能夠再進一步證明鳥類在不同生理時期及不同物種之鳴叫聲有不同特徵。這項技術對於鳥類行為的研究與分類，有著實際深刻的應用性。

Abstract

In considering the influence of kinship, the song of birds are more important for behavioral learning. If the identification of bird behavior and the classification of species are to be carried out, we will see that the song of birds play an important role. And due to the reasons that the song of birds are easily to approach the so- called Convergent Evolution and that there are even cases in which birds learn from other bird species, it is indeed difficult to identify the communication of birds behaviors just by human ears. In this paper, we will use machine learning to identify the slight differences of bird song exhibiting in the communication behavior of same bird. In order to prove that the bird song owns the different characteristics in different physiological periods and in different species, the physiological data set of White-eye were analyzed and the set was recorded by birdman Mr. Sun Qingsong over many years, while for the purpose of bird classification the bird songs were collected by the xeno-canto.org (XC) website. We have divided the data set of White-eye into seven different physiological periods and the XC data set is collected from the top 18 B grade recordings of the bird species by the XC website. After the noise removal, cut, and Mel Spectrogram and MFCC, and the process of converting sound to a time-frequency graph, we have performed the experiments of machine learning which belong to the BERT-based model with the Convolutional Module as the base, connected the models of LSTM, GRU, and Attention and changing the Feed Forward Network into the Conv-ID model. According to the experimental results, we have found that the method, changing to Conv1D the BERT-based model, has the best effect, and that the accuracy of macro F1-score, for the two issues of bird behavior analysis and classification, can reach 97.7% and 79.0%. Through t-SNE (t-distributed stochastic neighbor embedding) method and Visualized Attention Map, we are able to further prove that the song of birds own different characteristics in different physiological periods and in different species. This provided technology is of practical and profound application for biologist in the research field of classification of bird behavior.

壹、研究動機

記得小二時有次，整班被小學的老師神秘的趕到走廊外，不知情的被帶到學校的某個角落，分享她在校內發現的領角鴞 (*Otus lettia*, 鴞科角鴞屬)。雖然當時領角鴞眼睛睜大大的，但完全無動於衷 (之後才知道他其實是在睡覺)，我和同學都對此覺得有趣。也因為牠的存在，那段時間很常和好朋友們跑到樹下玩耍，玩累了就會邊看著他睡覺邊聊天。

某天一如往常我們賽跑到那棵樹下，但非常神奇的，我們第一次聽到領角鴞短暫且急促的鳴叫聲：「嗚——嗚——嗚——」雖然老師曾經說如果聽到牠的叫聲一定要馬上告訴她，但我們就只是呆呆的站在樹下聽他的聲音，甚至希望他和我們說話。希望某天可以再聽到牠發出不一樣的聲音，不過自此之後沒過幾天他就消失在校園了。



圖 1，領角鴞 (學名：*Otus lettia*) [1]

每當聽到鳥鳴時，都會不由自主的想聽懂鳴叫聲，甚至去猜測牠鳴叫的意義，但總是沒有辦法靠自己聽出一些端倪。直到高中，因緣際會下再次碰觸到機器學習這個領域，又有幸能得到取自鳥人孫清松先生經多年錄得的斯氏繡眼資料集，(常見的綠繡眼分類地位有所改變，根據“台灣鳥類名錄”台灣及大陸的綠繡眼是斯氏繡眼，日本及菲律賓的綠繡眼是日菲繡眼) [2, 3]，決定不再拘泥於讓自己聽得懂，而是將這個重責大任交付給機器學習，希望可以證明鳥的鳴叫聲確實會因為不同的原因而有不同特徵。

貳、研究目的

- 一、嘗試利用機器學習辨識同一種鳥類不同生理時期間聲音差異所進行的溝通行為。
- 二、可否針對 xeno-canto.org(XC)網站所收集的聲音資料集來辨識出鳥類物種。
- 三、找出最有效率的機器學習模型可以辨識溝通行為及分類，以及找出相關的最佳化參數。

參、研究方法及設備

一、研究設備

- (一) 本機：MacBook Pro (Retina, 15-inch, Mid 2015)
- (二) 線上資源：TAIWANIA 2
 1. Intel Xeon Gold CPU
 2. Nvidia Tesla T100 w/32GB



圖 2 台灣杉二號 [4]

(三) Python 環境

於 Taiwania2 上 Conda 虛擬環境上安裝之套件（套件預覽於附錄一）。

二、實驗資料集

我們在實驗中利用了兩個資料集，分別從孫清松先生錄音搜集的斯氏繡眼鳴叫聲——生長時期資料集，以及從 xeno-canto.org [5] 收集的鳥類鳴叫聲——物種資料集。

(一) 斯氏繡眼鳴叫聲－生長時期資料集

斯氏繡眼 (*Zosterops simplex*)，是屬於繡眼科 Zosteropidae，雀形目，目前分佈於大部分之東亞地區如：中國、台灣、南韓、北韓、越南等，在日本及菲律賓的綠繡眼是日菲繡眼，而且因為適應力強也曾因國際性鳥類貿易出現於世界各地。全台低至中海拔區域普遍可見，通常以 20~50 隻為單位成群出現。斯氏繡眼於台灣因其外表及叫聲又有綠繡眼、暗綠繡眼鳥、青笛仔等俗名 [6]。

選擇斯氏繡眼之主要原因，除了依照現有資料相較容易取得之外，斯氏繡眼同時也是台灣常見的鳥類之一。希望藉由斯氏繡眼的聲音資料來建立基本架構，並將其模型 應用到辨識其他鳥類不同時期，甚至是辨識不同種類鳥類身上。



圖 3、斯氏繡眼（學名 *Zosterops simplex*） [7]

我們自鳥人孫清松先生拿到本資料集（以下簡稱**斯氏繡眼資料集**）。總共在約 30 個不同地點錄音，都是集中在孫清松先生台中地區的鳥園觀賞區，錄音時間集中在 1999 年至 2010 年，包含合計 65 分鐘的斯氏繡眼鳥叫聲對應到斯氏繡眼生長時期的成對資料。經過去除空白聲音等步驟後，共有合計約 3GB，30 餘分鐘的數據。

本次資料集均為斯氏繡眼之鳴叫聲，而每一筆資料各有標記錄音時間、錄音地點、當時天氣、鳴叫長度、起鳴止鳴點、各生殖階段及當時錄音者留下來之備註。數據包含七個種類（斯氏繡眼生長的七個時期），各時期的說明如下列表：

- 巢前期: 目標於巢前鳴叫之時期。
- 築巢期: 目標築巢之時期。
- 抱卵期: 目標抱卵之時期。
- 巢外育: 目標於巢外育養幼鳥時之時期。
- 產蛋期: 目標產蛋時鳴叫之時期，目標僅限於雌性。
- 育雛期: 目標育在巢內養幼鳥時之時期。
- 雛失敗: 目標育養之幼鳥去世後鳴叫之時期。

經反覆觀察聆聽，很難利用耳朵觀察不同時期之關聯性及聲音之差別（資料集預覽於附錄表 2）。

(二) XC 資料集

為了解決鳥種辨別問題，我們收集自 xeno-cento.org [5] 上的鳥種聲音。本資料集（下簡稱 XC 資料集）選取數量前 18 多的種類用以訓練分類模型，並且僅使用 B grade 以上的錄音檔，每一鳥種資料大小從 7G 到 1.7G 不等，共 55G。由於 XC 有些收音品質較低，背景噪聲較大，我們並未先裁減沒有鳥鳴的部分，反之使用較長的切片段長度，訓練模型判斷。更多實驗細節請參見研究過程及方法部分（資料集預覽於附錄表 3）。

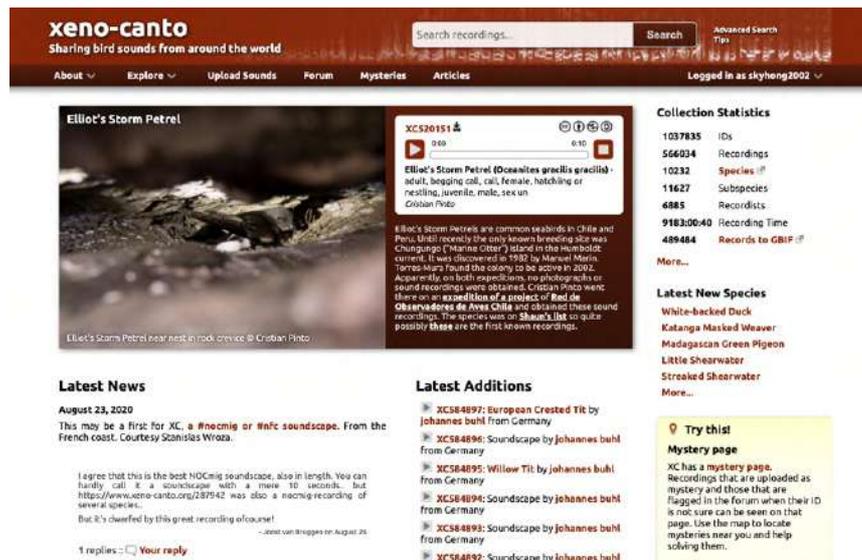


圖 4、xeno-canto.org [5] 網站截圖

這些挑選出的鳥類都是歐洲常見的鳥種，除了一種灰林鴉 (*Strix Aluco*, 鴞形目鴞科)外，其他全都是雀形目鳥類。雀形目鳥類 [8] 鳴聲多樣，又稱鳴禽 [9]，這些選取恰好都是適合用來做為鳥類鳴聲種類辨識的鳥種。

以下為我們選取的種類：

表 1、XC 資料集採用種類

俗名	學名	總秒數	錄音檔數
烏鶇	<i>Turdus merula</i>	258315	2070
歌鶇	<i>Turdus philomelos</i>	232295	1387
黑頂林鶇	<i>Sylvia atricapilla</i>	166064	1778
濕地葦鶇	<i>Acrocephalus palustris</i>	160059	899
歐亞歌鶇	<i>Erithacus rubecula</i>	150121	1675
大山雀	<i>Parus major</i>	141233	2546
新疆歌鶇	<i>Luscinia megarhynchos</i>	134712	1142
布萊氏葦鶇	<i>Acrocephalus dumetorum</i>	122332	655
蒼頭燕雀	<i>Fringilla coelebs</i>	118774	1759
巉渣柳鶇	<i>Phylloscopus collybita</i>	101577	1654
灰白喉林鶇	<i>Sylvia communis</i>	94481	1021
鷓鴣	<i>Troglodytes troglodytes</i>	82810	1258
歐亞柳鶇	<i>Phylloscopus trochilus</i>	73502	953
灰林鴉	<i>Strix aluco</i>	72676	809
黃鶇	<i>Emberiza citrinella</i>	69388	874
歐亞藍山雀	<i>Cyanistes caeruleus</i>	60473	1082
家麻雀	<i>Passer domesticus</i>	60146	884
黍鶇	<i>Emberiza calandra</i>	53141	687

肆、研究過程及方法

一、研究架構

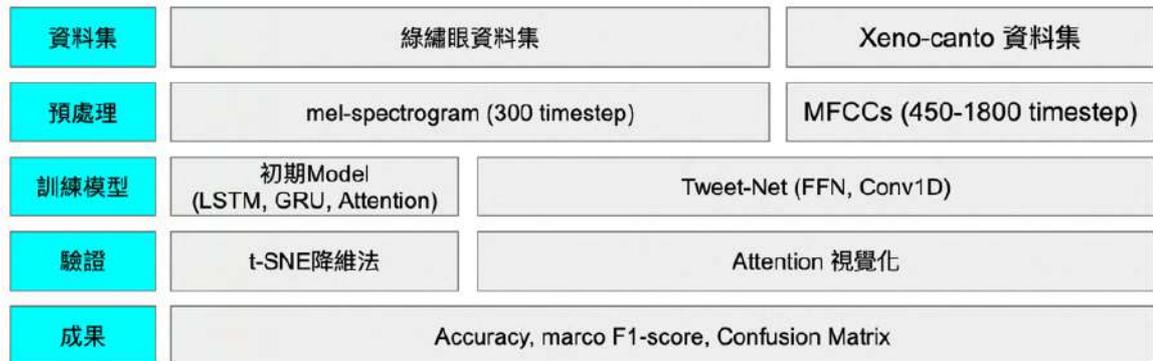


圖 5、鳥鳴聲研究的架構圖

二、資料預處理

在訓練語音辨識模型時，不會直接使用時域圖進行訓練，因為很難從時域圖中得到足夠資訊。於是我們將原聲音資料利用以下不同的時頻分析法 TFA (Time-Frequency Analysis) 後得到時頻圖以便訓練：

(一) 梅爾頻譜圖 (Log-Mel spectrogram)

梅爾頻譜圖是將原訊號經 STFT 專換後經過梅爾頻譜轉換後的結果。本次實驗大部分皆使用此頻譜圖作為資料集進行訓練。

1. 短時距傅立葉轉換 (Short-time Fourier transform, STFT)

傅立葉轉換是將一個聲音訊號的時域 (time domain) 轉換為頻域 (frequency domain) 的方法。而短時距傅立葉轉換則是傅立葉轉換的變體，不過短時距傅立葉轉換是將原訊號切成為多個片段，個別經過傅立葉轉換以 time step 接在一起的 (time step, 頻率, 強度)的三維陣列，是可以保留時間維度和頻率維度的好方法。

2. 三角濾波器 (Filter Bank)、梅爾刻度 (mel scale)

梅爾頻譜轉換是一種非線性刻度單位，目的是模擬人耳對聲音的非線性感知能力。在較低的頻率下更具辨別力，在較高的頻率下則不具辨別

力。使用梅爾刻度的三角濾波器可以將原本頻譜圖的赫茲刻度轉換為梅爾刻度，便可得到梅爾頻譜圖 (Log-Mel spectrogram)。^[10, 11]

(二) 梅爾倒頻譜係數 (Mel-frequency Cepstral Coefficients, MFCCs)

因本次處理之資料多半為較短暫的聲音檔案，因希望從音檔中提取出特徵，因此選擇使用梅爾倒頻譜(Mel-Frequency Cepstrum, MFC)轉換，步驟是將先前得到的 Log-Mel spectrogram 經過三角帶通濾波器對頻譜進行平滑化，消除諧波的作用後再將其進行離散餘弦轉換 (Discrete cosine transform, DCT)。最後再加上對數能量 (Log energy) 與增量倒頻譜參數 (Delta cepstrum) 提取特徵，得到 MFCCs。^[12, 13]。

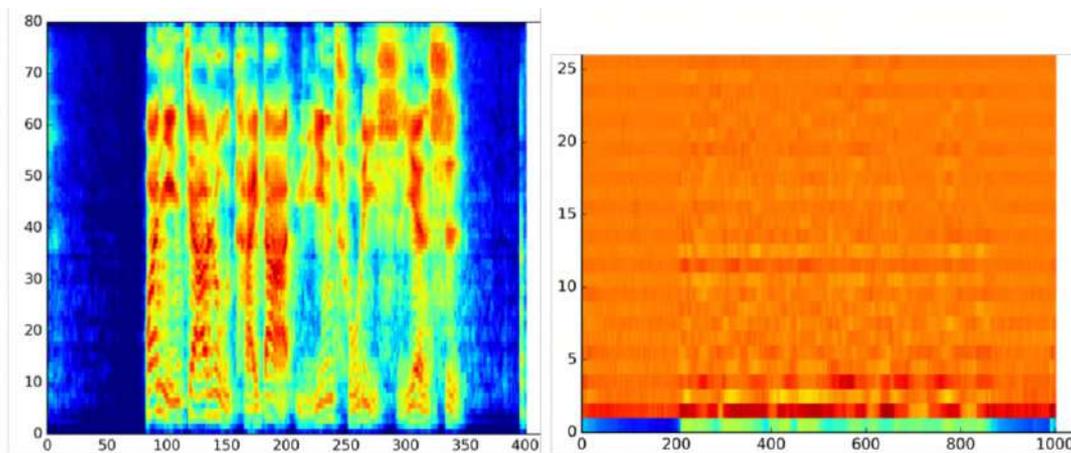


圖 6、梅爾時頻譜與梅爾倒頻譜係數示意圖 [14]

對於所有資料集，我們將錄音檔轉為梅爾時頻圖 (Mel-Spectrogram) 以及 MFCC。其中在處理斯氏繡眼資料集時，由於數據本身雜訊較少，我們直接切除錄音中沒有聲音的部分。在處理 XC 資料集時，由於此資料集雜訊較多，直接切除沒有鳥類鳴叫的聲音並不容易，因此我們並沒有切除多餘的聲音部分，並且將其一同送進模型訓練，由模型判斷之，希望能得到普遍性的分析方法。

在斯氏繡眼資料集，我們分別實驗了轉換為 Mel-Spectrogram 或 MFCC 作為模型輸入。其中 Mel-Spectrogram 在頻率上使用 128 維，MFCC 使用 64 維；在 XC 資料集中僅訓練 128 維 Mel-Spectrogram 作為輸入的模型。

三、模型設計

在我們的任務中，我們實驗了兩種模型來處理分類。分別為 1D 卷積(Conv1D)為主的模型，以及利用類似於 BERT 模型結構的 Multi-Head Attention 模型。

(一) Conv1D-Based Model

參考了一些前人的做法後 [15]，決定將模型以 1D 卷積模型為基礎 [16]，在每個 time-step 上做卷積，同時透過 Max-Pooling [17] 降低 time-step 上的寬度，並且接上 GRU、LSTM 或是 Self-Attention 層後，接上 Global Average 層取平均值，最後再接上全連接層作為輸出。結構圖解如下：

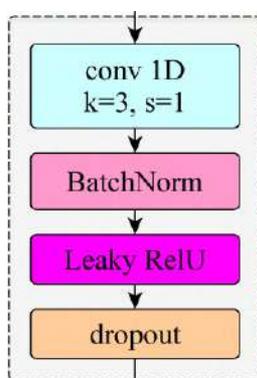
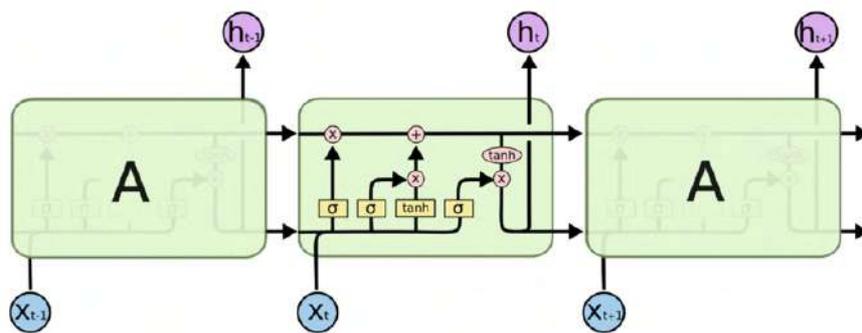


圖 7、Convolutional Module 單元示意圖

1. Long Short-Term Memory (LSTM)

一開始想到可以當作模型的發法為遞歸神經網路（Recurrent Neural Network, RNN），但因為 RNN 在處理有時序性（資料的先後互相有關係稱為時序性）的資料時，會因為較無法繼承長期記憶而導致表現不佳，於是改選擇長短期記憶模型（Long Short-Term Memory, LSTM）加入模型 [18]。

訓練資料與測試資料均為 6:4。嘗試使用 64 個 unit 及 128 個 unit 進行訓練。所有訓練方式 Dropout rate 均為 0.5 [19]，且皆使用 Adam Optimizer [20, 21] 及使用 Tensorflow.keras 預設學習率。



The repeating module in an LSTM contains four interacting layers.

圖 8a、LSTM cell 示意圖 [22]

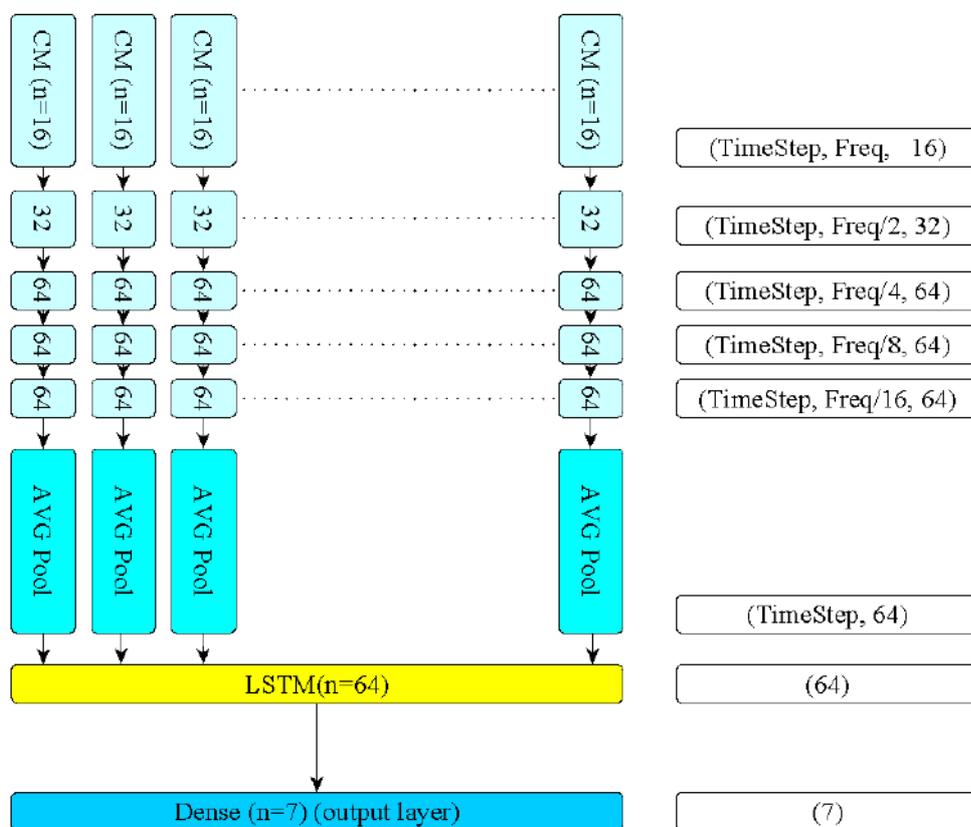
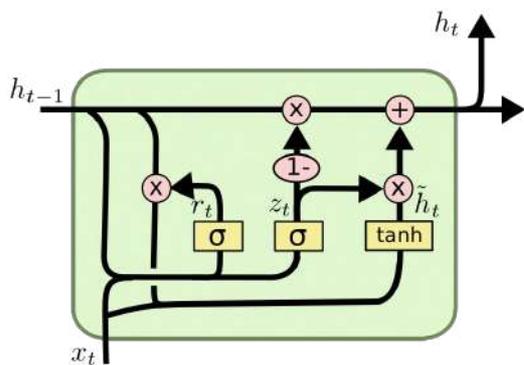


圖 8b、LSTM 架構示意圖 (n=64)

2. Gated Recurrent Unit (GRU)

此為 LSTM 的一種改良版。GRU 將 LSTM 中的遺忘閥 (forget gate) 與輸入閥 (input gate) 用一個更新閥 (update gate) 取代，並把單元狀態 (cell state) 和隱藏狀態 (H_t) 進行合併，計算新資訊的方式和 LSTM 也有所不同。



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

圖 9a、GRU 單元示意圖及計算方式 [22]

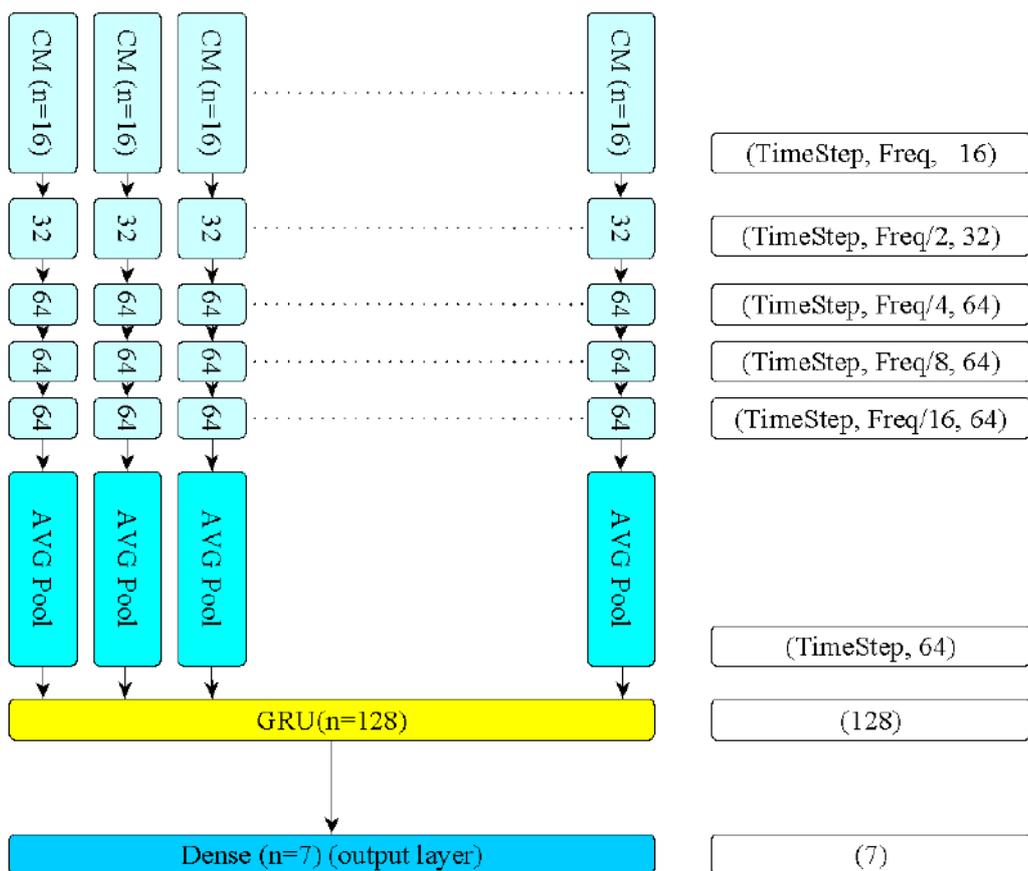


圖 9b、GRU 架構示意圖

3. Attention

相較循環神經網路（Recurrent neural network）只能看到之前的資料，且當輸入片段較長時很容易忘記之前的內容。而 Self-attention 藉由平行運算可以一次看完所有內容並得出結果 [23]。模型架構如下：

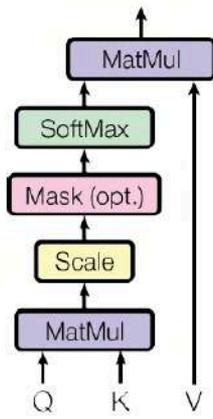


圖 10a、Attention 單元示意圖 [24]

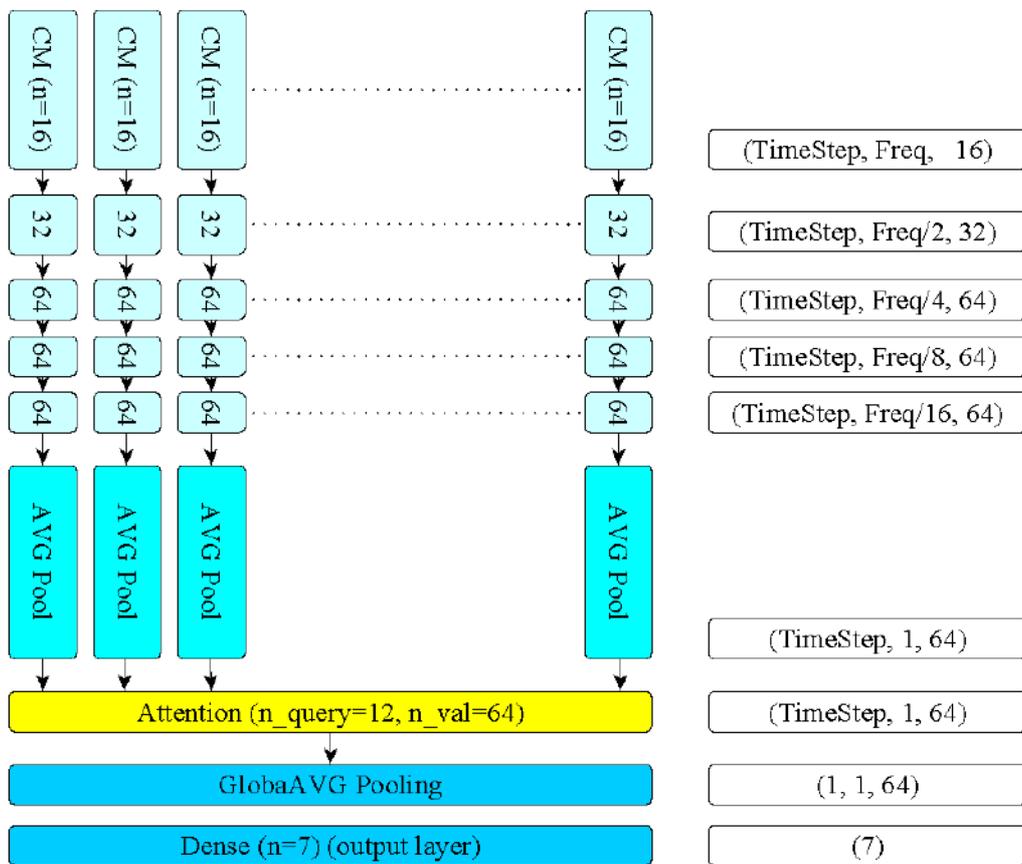


圖 10b、Attention 架構示意圖

(二) BERT-based Model

我們初步利用 Attention 架構嘗試解決鳥種分辨問題，但因為表現不佳，決定開發新的模型。為了解決鳥種分辨問題，我們接受能夠解決大部分 NLP (Natural Language Processing, 自然語言處理) 任務的 BERT (Bidirectional Encoder

Representations from Transformers) 啟發，想到是否可以將聲音分類看作為序列分類任務，再參考前人文獻後 [25-27] 決定將 Transformer 的 Encoder 當作基底開發。

我們在梅爾時頻圖前新增 <CLS> token (即全部相同的數字，在實驗中為 60.0)，並經過一層全連接層調整形狀至 256 維，並輸入至與 Transformer 的 Encoder 相同的網路結構中，一層多頭注意力層、即一層前饋網路，重複數次的結構。

先前實驗的訓練和驗證的資料比例為 6:4，而 Accuracy 則是依照驗證資料算出來的。但因為這樣的切割方式無法保證訓練出的模型會是最佳的，於是我們將訓練、驗證和測試資料比例調整為 7:1:2，而將測試資料用來測量最終模型的結果。[28]

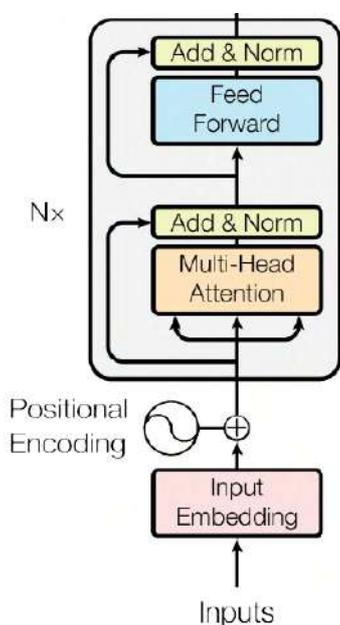


圖 11a、Transformer 的 Encoder 示意圖

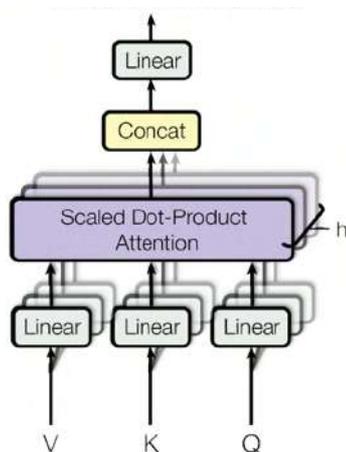


圖 11b、Multi-Head Attention 示意圖 [24]

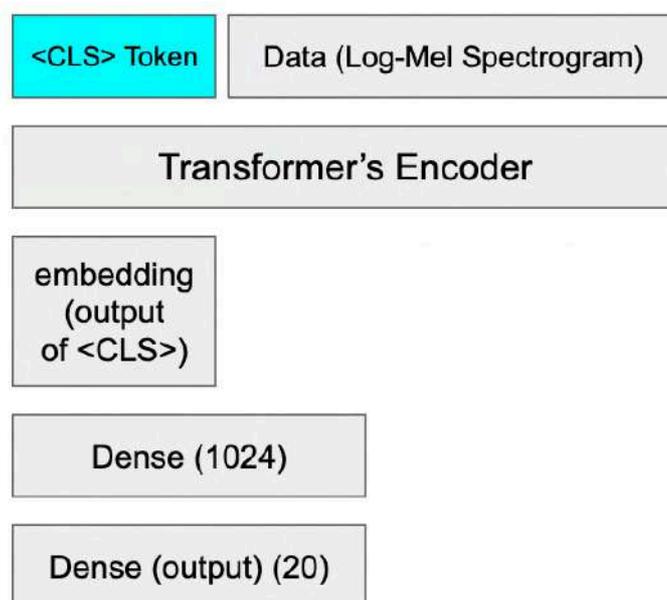


圖 11c、BERT-based Model 模型示意圖

1. 鳥種分辨問題

由於 mel spectrogram 及 MFCC 之序列長度較長，為了使訓練過程控制在合理的顯存使用量之內，我們將層數縮減至 8 層以下。我們枚舉不同層數的 Encoder 和不同個數之 Multi-Head Attention，實驗 (4, 2)、(4, 4)、(4, 8)、(8, 4) 的組合。

除了訓練原始之 BERT 結構，我們也嘗試將前饋網路 (Feed Forward Network, FFN) 換成兩個 kernel size = 5、stride = 1、並且填充 padding 的 Conv1D (在頻率維度上卷積)，嘗試獲得較高的準確度。

2. 鳥類生理時期分辨問題

經過多重測試後，我們直接利用 Conv1D (kernel size = 5, stride = 1, padding = 2) 代替前饋網路，並使用 4 個 Multi-Head Attention 的 4 層 Encoder 進行實驗。

四、驗證

(一) t-SNE 模型視覺化

t-隨機鄰近嵌入法 (T-distributed Stochastic Neighbor Embedding) 是一種非線性的機器學習降維方法 [29]，可以將資料視覺化。我們利用此方法將 Attention 模型

將為視覺化，驗證斯氏繡眼在不同生理時期是否有不同特徵，且藉由 Attention 模型是否能夠訓練出能夠辨認聲音特徵。

(二) Attention map 權重視覺化

在 BERT-based model 中，為了確認訓練出來的模型能夠準確辨別鳥類聲音，於是將模型中注意到 <CLS> Token 權重的 Attention map 視覺化，預期模型權重會對應到在聲音出現的地方，印證訓練模型的正確性 [30]。

(三) Confusion Matrix

在各個表現良好的模型中，我們將畫出其混淆矩陣 (Confusion Matrix)，將模型容易搞混的地方釐清原因比找尋解決辦法。

伍、研究結果

一、Conv1D-Based Model

以下是斯氏繡眼資料集中分別是用 Conv1D-Based Model 套上 LSTM、GRU 和 Attention 的結果。

表 2、Conv1D-Based Model 不同模型之表現

Model	LSTM(n=64)	LSTM(n=128)	GRU	Attention
Total Accuracy	49.20%	56.09%	85.20%	95.52%

在 LSTM 的結果中，n=64 之最高 Accuracy 為 49.20%，n=128 時最高 Accuracy 則為 56.09%。在附錄表 4 中可以看出，當 n=64 時會因為 node 太少而收斂效果不好 (Underfitting)；當 n=128 時擬合過度 (Overfitting) 的情況也很嚴重。在 GRU 的結果中，雖然情況有好轉，但表現最高只能夠達到 85.20%；而在 Attention 模型中收斂速度極快，Overfitting 之情況也不高，且測試資料準確度提高至 95.51%。以下為 Attention 模型中的訓練 Loss 及 Accuracy 隨訓練 Epoch 之表現：

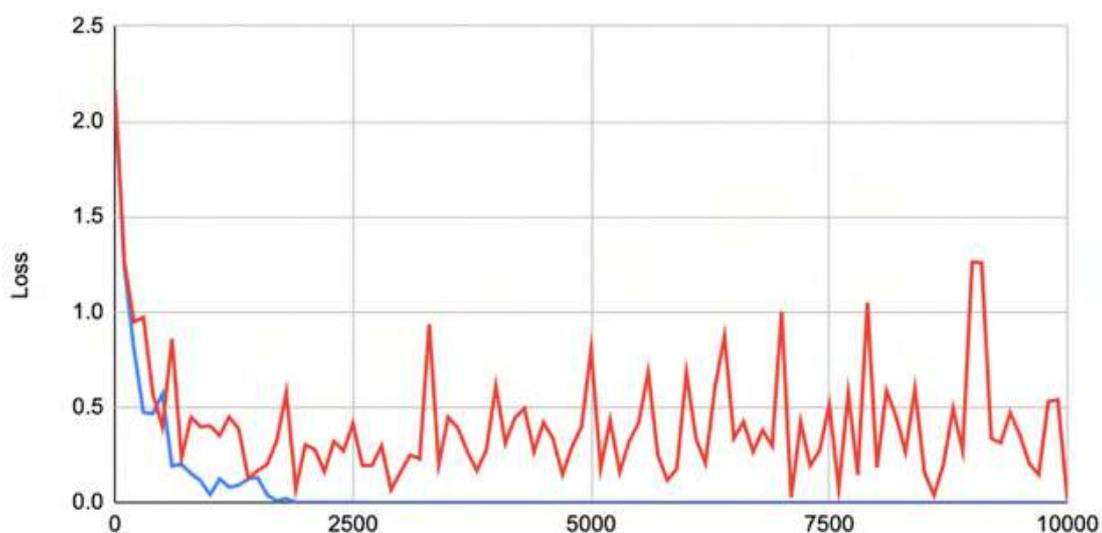


圖 12a、Attention 模型訓練過程，藍線: training，紅線: validation (橫軸: Epoch, 縱軸: Loss)

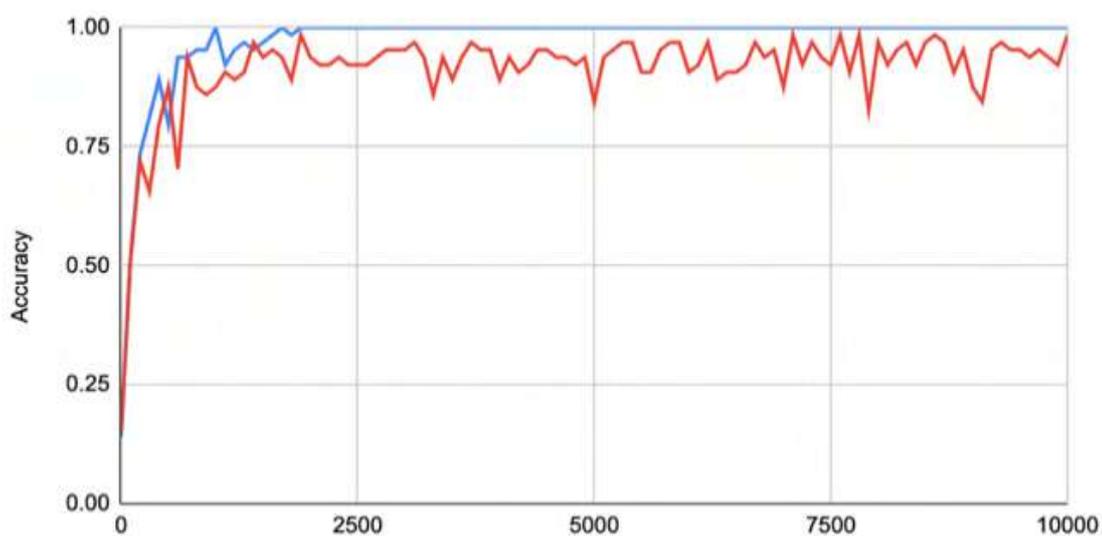


圖 12b、Attention 模型訓練過程，藍線: training，紅線: validation (橫軸: Epoch, 縱軸: Accuracy)

二、BERT 模型

(一) 鳥種分辨問題

1. Feed Forward Network 之結果

表 3、FFN 枚舉 Encoder Layer 及 Attention head 個數表現

(Layers, Heads)	(4, 2)	(4, 4)	(4, 8)	(8, 4)
Total Accuracy	63.52%	66.69%	69.63%	64.86%
Average Accuracy	62.69%	65.02%	68.83%	62.80%
Marco F1-score	61.62%	64.57%	68.34%	62.14%

2. Convolutional Layer (kernel size=5, stride=1) 之結果

表 4、Conv1D 枚舉 Encoder Layer 及 Attention head 個數表現

(Layers, Heads)	(4, 2)	(4, 4)	(4, 8)	(8, 4)
Total Accuracy	78.95%	75.94%	79.62%	78.04%
Average Accuracy	78.74%	75.60%	78.97%	77.71%
Marco F1-score	78.85%	75.07%	<u>79.00%</u>	77.26%

經由實驗結果得知，利用 Convolutional Layer 1D 代替 Feed Forward Network 能夠有效提高表現，且在當 Encoder Layer = 4, Head = 8 時，Marco F1-score 能夠達到 79.00%。

(二) 生理時期分辨問題

表 5、Conv1D Transformer 及 Attention 表現

(Layers, Heads)	(4, 4)
Total Accuracy	97.57%
Average Accuracy	97.66%
Marco F1-score	<u>97.78%</u>

在 Conv1D (kernel size = 5, stride = 1)，使用 4 個 Multi-Head Attention 的 4 層 Encoder 的結果之 Marco F1-score 最高可達到 97.78%。

三、驗證

(一) t-SNE 模型視覺化

將 Attention 模型訓練結果進行 t-SNE 降維視覺化後，可以發現模型清楚的將各個時期的鳥鳴聲分類，因此可以證明藉由 Attention 模型能夠訓練出能夠辨認聲音特徵。

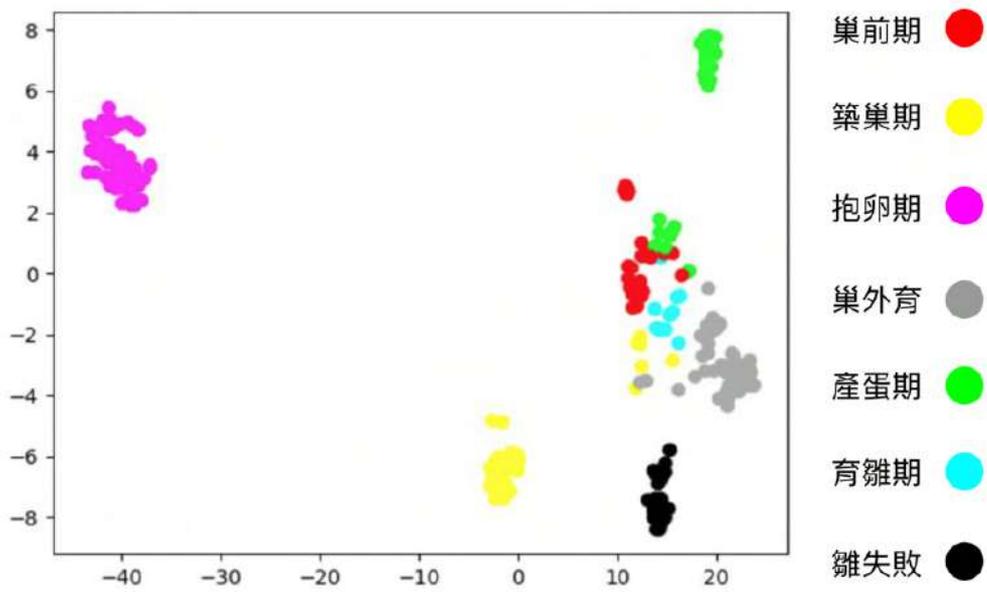
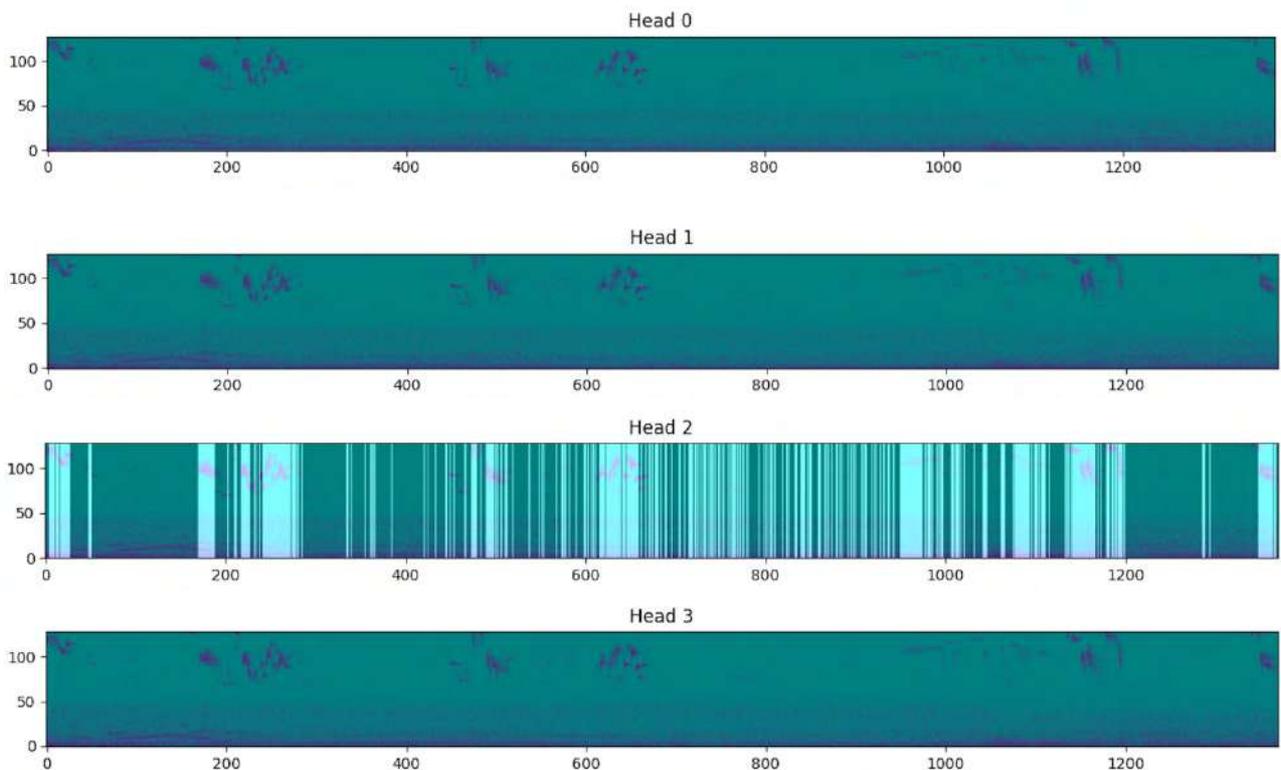
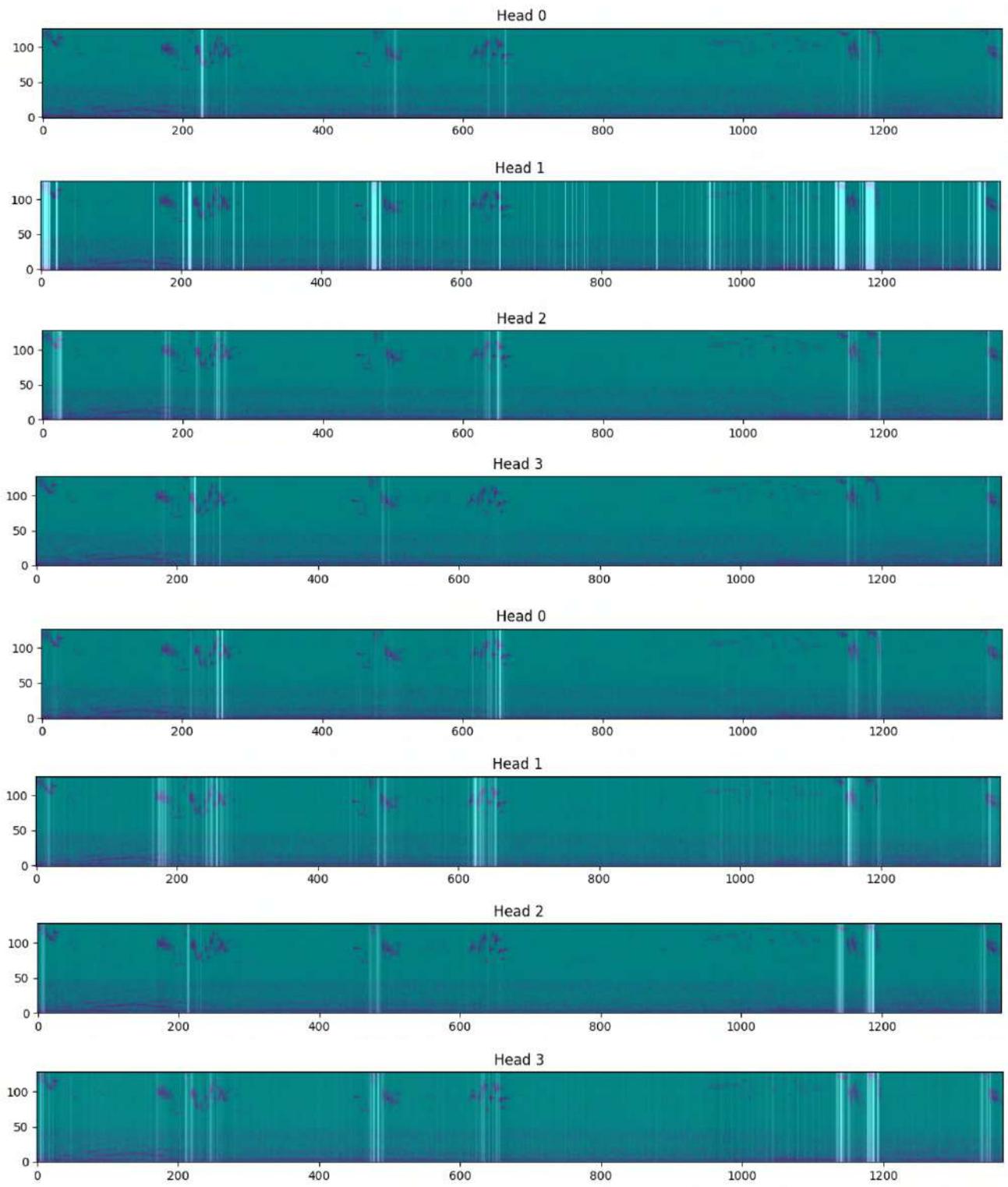


圖 13、Attention 結構經 t-SNE 降維視覺化之結果

(二) Attention Map 權重視覺化

藉由結果可以發現 <CLS> Token 的注意力都與鳥鳴叫之聲音範圍內的特徵有關，證明 BERT-based Model 有學習正確。以下展示生理時期辨別問題的 XC 資料集的其中一筆資料波形圖疊加上的 Visualized Attention Map (4 Layers, 4 Heads)：





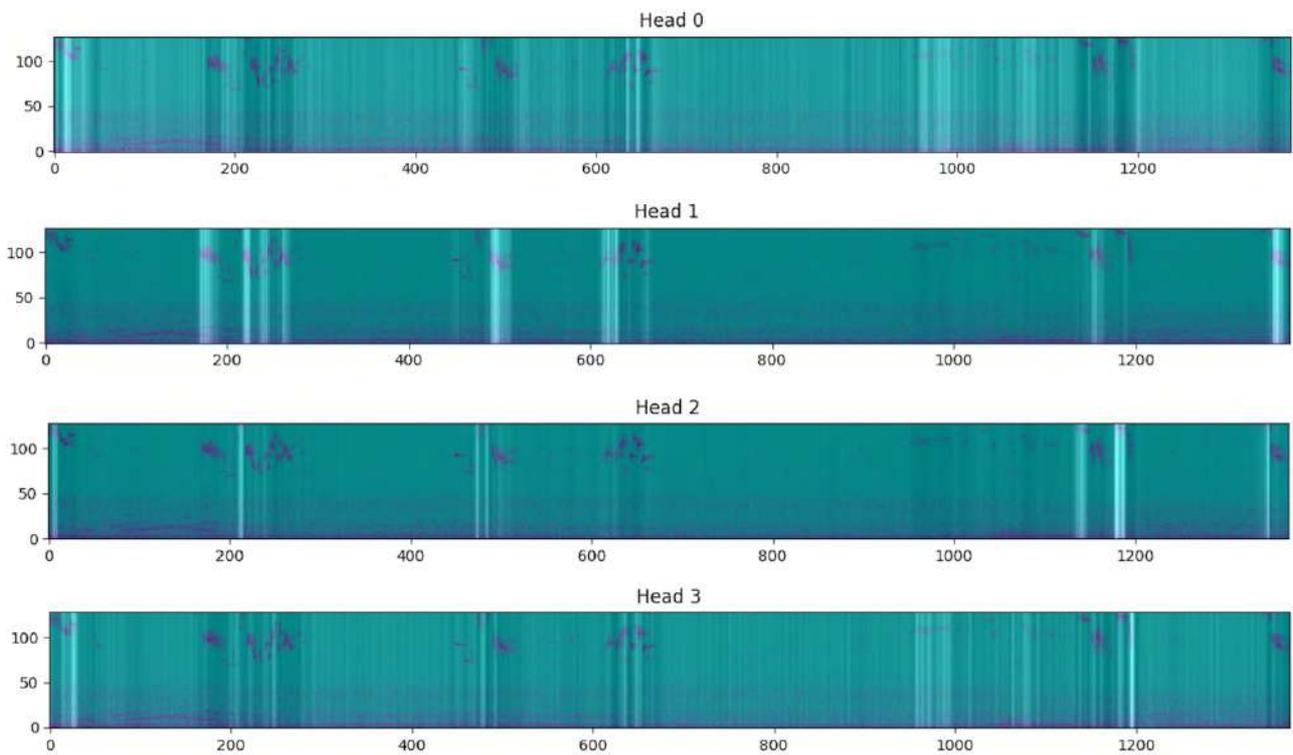


圖 14a~14p、Visualized Attention Map 示意圖

(三) Confusion Matrix

藉由 Confusion Matrix 可以清楚列出每個 Test data 的 Actual label 和 Predict answer 各為何者，並藉由比較可以發現由 BERT-based Model 解決鳥類生理時期辨別問題之 Confusion Matrix 較 Conv1D-based Model 來的好。

以下展示 Conv1D-based Model 解決生理時期辨別問題和 BERT-based Model (Conv1D) 分別解決鳥類生理時期辨別問題及鳥種辨別問題的 Confusion Matrix：

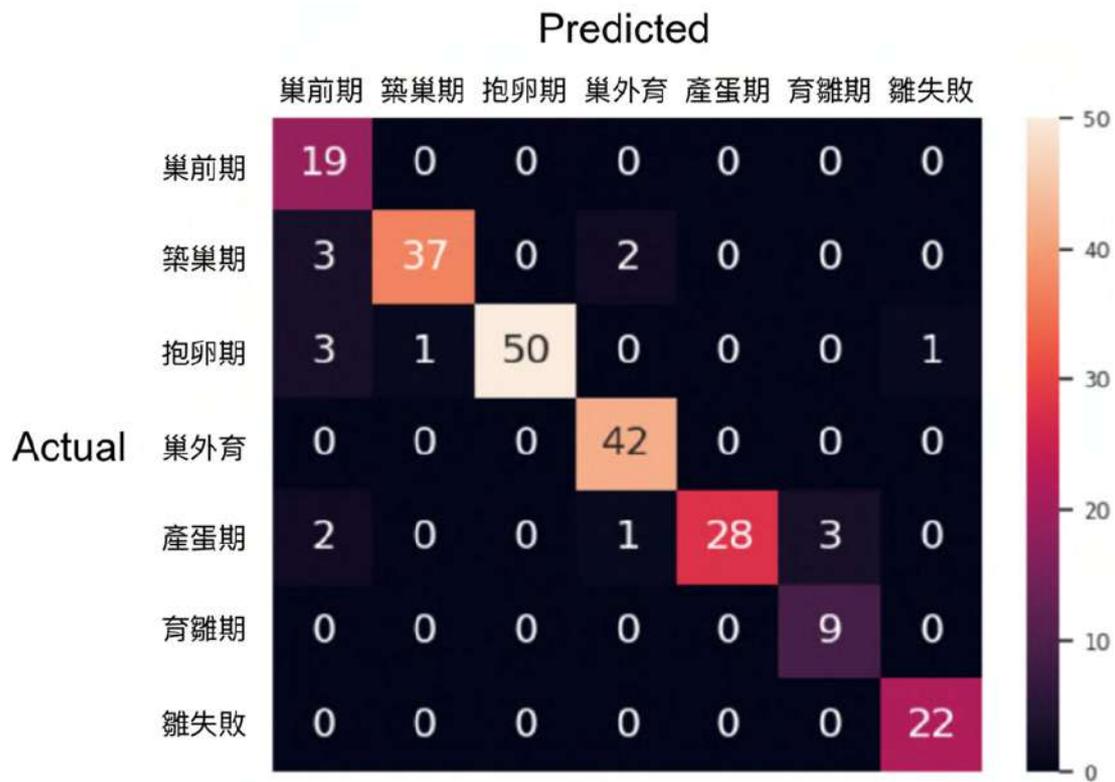


圖 15a、生理時期辨別問題 Confusion Matrix (Conv1D-based, Attention)

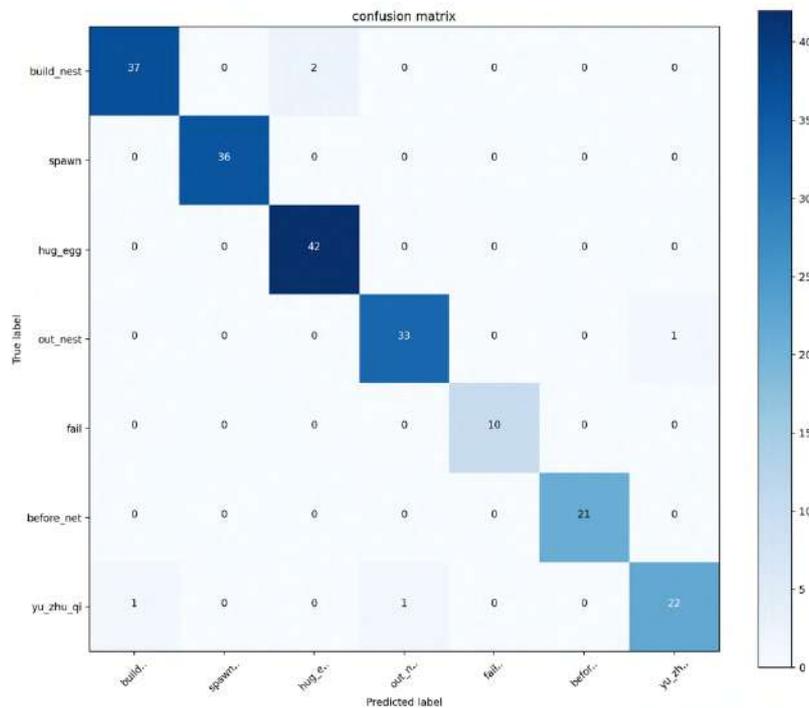


圖 15b、生理時期辨別問題 Confusion Matrix (BERT-based, Conv1D, 4 Layers, 4 Heads)

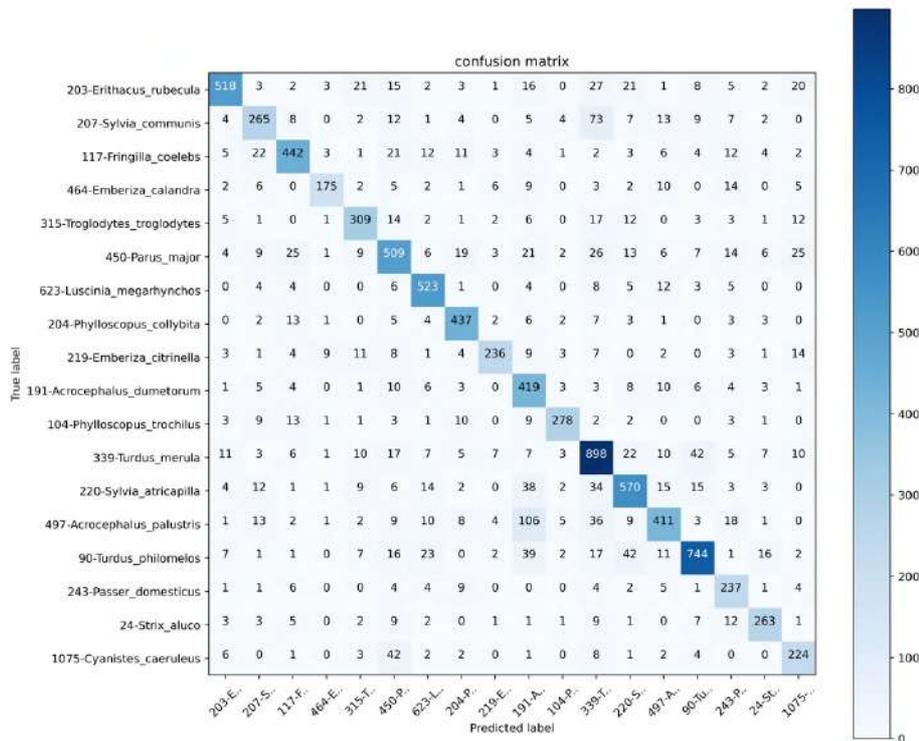


圖 15c、鳥種辨別問題 Confusion Matrix (BERT-based, Conv1D, 4 Layers, 8 Heads)

陸、討論

一、實驗結果討論

(一) Conv1D Model

藉由測試 LSTM 模型後，可以知道利用 LSTM 訓練效果不盡理想：當模型太小時 loss 值無法收斂；模型太大時會因為資料集太小無法訓練太多 Node 導致 Overfitting。而藉由測試 Gated Recurrent Unit 模型後，可以發現經由 GRU 優化後可以有效減少所需 Node 大幅改善 Overfitting 的問題。

藉由使用 Attention 之結果，將此技術套用到此問題分析獲得很大的改善。將其結果進行 Confusion Matrix 上可以發現每個時期之間都有很高的辨識率，唯獨產蛋期及育雛期有些相似，但就狀態上來看因為也是相近的；將訓練結果進行 t-SNE 視覺化後，也可以看到分明顯的分塊現象。綜合以上結果，確定斯氏繡眼在不同時期的鳴叫聲的確有不一樣的特徵。

(二) BERT-based Model

BERT 原來是用來處理 NLP 任務為主。雖然我們不能確定鳥鳴是否為自然語言，但因為和自然語言一樣都是 sequence，故能夠藉由 BERT-based Model 訓練出結果。以下為原本 Confusion Matrix 經過 Normalize 後之結果：

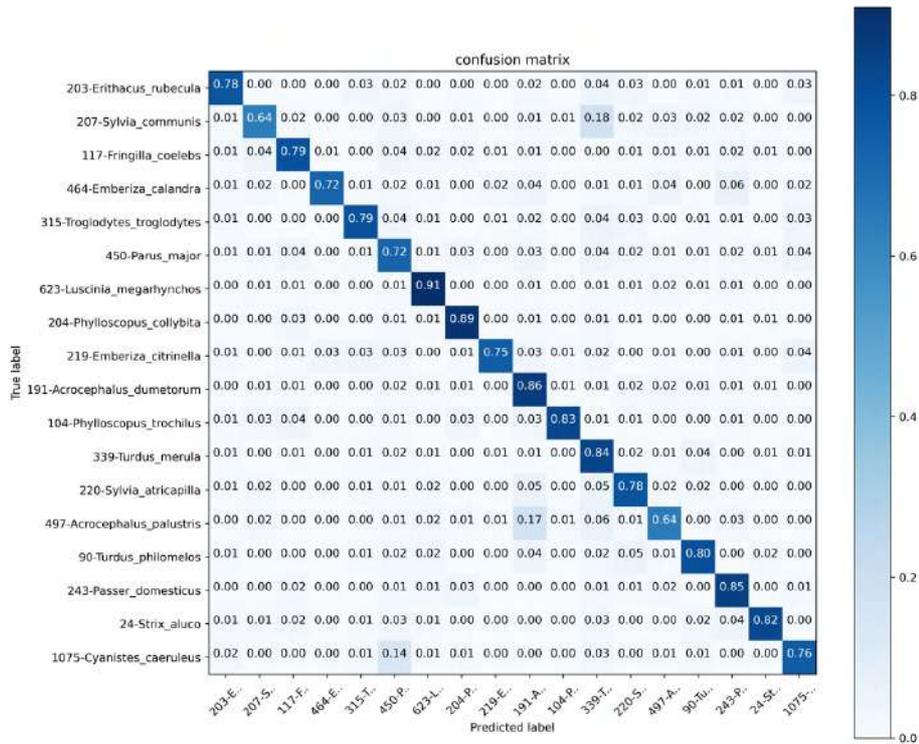


圖 16a、Normalized Confusion Matrix (Conv1D, 4 Layers, 8 Heads)

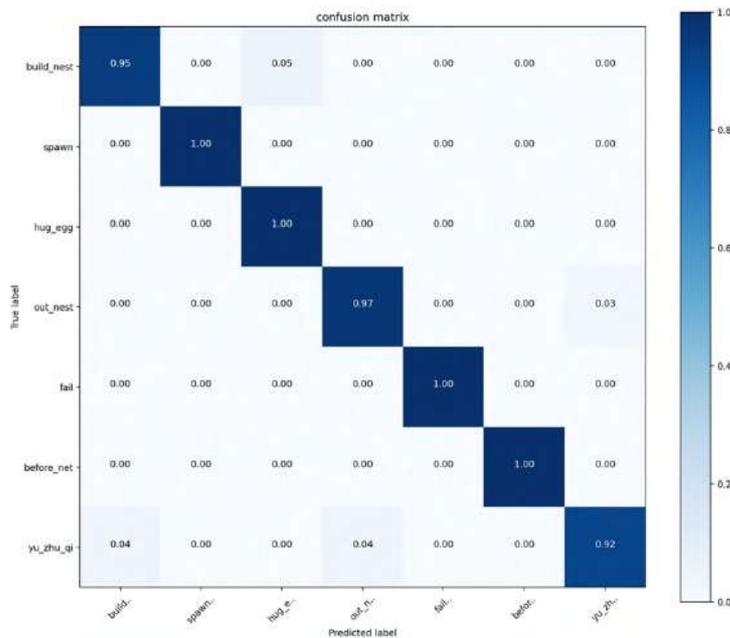


圖 16b、Normalized Confusion Matrix (Conv1D, 4 Layers, 4 Heads)

二、資料集

(一) 斯氏繡眼資料集

在斯氏繡眼資料集中，每個資料都被分別標上七個不同生理時期，而經過實驗後也證明不同生理時期有特定特徵可以分析。雖然我們沒有辦法藉由鳥鳴聲理解當下鳥想表達的意思，但因為有不同特徵的鳥鳴，因此此分析確實能夠成為鳥之間溝通的媒介。巢前期是目標在鳥巢前鳴叫，同時也是為了吸引異性，而發出與眾不同的聲音，而我們訓練的結果中，巢前期的分辨率也是七種時期中最高的。即便我們仍然不曉得在其他時期中斯氏繡眼想要透過鳴叫聲表達什麼，但我們可以確定的是鳥在鳴叫時除了引起注意外，一定還有其他訊息想要傳達。

(二) CX 資料集

除現今 18 種再加上麻雀 (*Passer montnaus* 雀形目麻雀科) 和歐亞柳鶯 (*Phylloscopus trochilus*, 雀形目柳鶯科) 兩種，便是我們一開始搜集選 XC 資料集選擇 20 種資料集多的鳥類進行訓練。但因為麻雀和柳鶯訓練效果不佳，因此將其放棄。但是到後面分析檢討時 [31]，才發現機器學習模型將 104-歐亞柳鶯 (*Phylloscopus trochilus*) 及 41-歐亞柳鶯亞種 (*Phylloscopus trochilus trochilus*) 認為是相同機率 0.86、以及 824-麻雀 (*Passer montnaus*) 及 243 家麻雀 (*Passer domesticus*) 認為是相同機率 0.36，如圖所示，我們認為這是機器學習強大的地方，一般亞種的分類，有些是聲音相近，外貌不同，所以才細分亞種，例如 104-歐亞柳鶯及 41-歐亞柳鶯亞種 0.86 相同，而相同雀科時，家麻雀和麻雀其實叫聲也是相近，所以機器學習視為 0.36 相同，這點是我們模型設計精準所致，此點未來可以給生物學家更加細緻分類的意見。

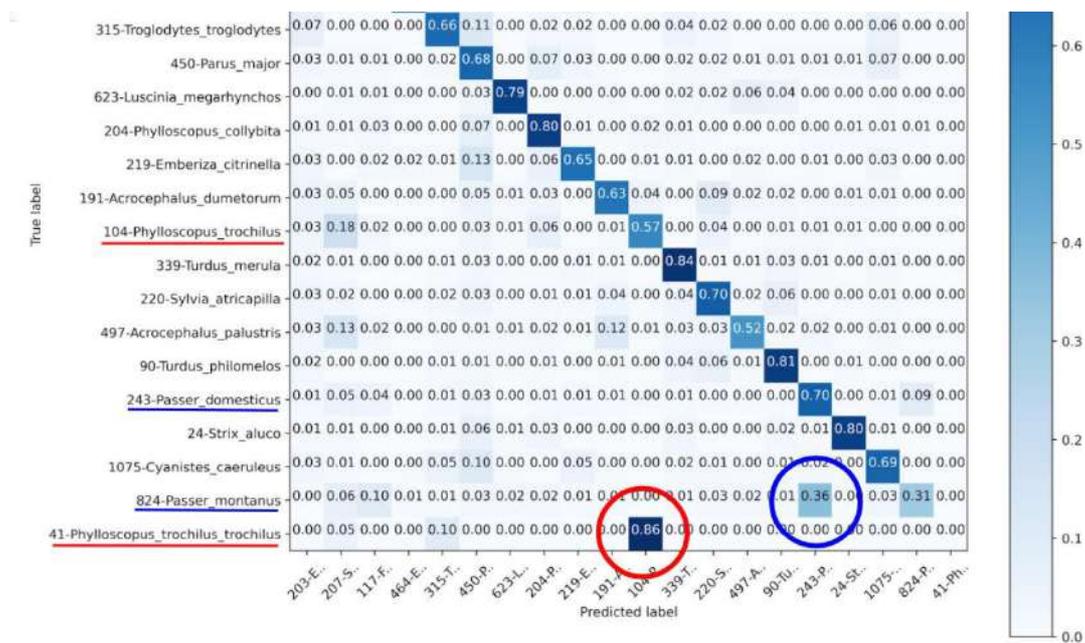


圖 17、104-歐亞柳鶯及 41-歐亞柳鶯種、824-麻雀及 243 家麻雀對角線外高機率示意圖
(亞種問題與同科問題)

三、實驗誤差

(一) 資料集大小

斯氏繡眼資料集內容僅收錄斯氏繡眼鳴叫聲及其七種生活的時期之標籤，且已資料數量上看，若是要計算更加複雜的內容，計算量勢必將不足以負荷；XC 資料集種類僅採用 18 種也是因為雖然在 xeno-canto.org [5] 上的種類高達 27674 種，但在此同時卻只有 927 種鳥類的錄音次數超過 100 次，且這些數量是包含 A, B 等級以外的品質，C 等級以外的聲音，人耳初步辨識品質相較不佳 [32]。

(二) 錄音品質

雖然經由改良後的模型已經有明顯進步，但以現在的資料集來說，只要是錄音效果不同都有可能影響到實驗結果。未來訓練時應當於資料預處理時加入不同程度之雜訊，除了增加訓練資料集外也可以更接近實際應用情況。

(三) 環境變因

在斯氏繡眼資料集中，如果將錄音地點移動到不同地形，亦或是使用不同的錄音設備，皆會影響到其正確率。因此如果訓練驗證資料各自取自不同來源，驗

證之可信度會更高。雖然斯氏繡眼資料集所使用的錄音設備及錄音地點均一致，但若是使用不同的錄音設備或不同環境，不保證能夠保持高準確率。

四、未來展望

(一) 模型、資料集優化

在本次研究中，並沒有將資料集做特殊處理，或經由不同手續增加測試資料。除了從根本上增加資料集，在未來的研究也將把資料集做不同處理，讓控制變因增加，較能預期實驗效果；在研究模型上，也希望在未來能夠得到更高的訓練效果，因此也將測試不同的結構，讓成功率提升。

依照目前的構想，假設可將鳥鳴視為一種自然語言，或許能夠當作 NLP 任務看待並以 Transformer 當作模型訓練。未來若要實作 Transformer，可以將所有 xeno-canto.org 上的所有鳥叫聲音當作預訓練集，再將 mel spectrogram 或 MFCC 等 sequence 當作訓練資料集進入 Transformer 內，訓練效果可能會更加提升，更可以一次解決不同種類的鳥鳴辨識問題 [33, 34]。

(二) 比較不同物種之關係

因資料集的限制，鳥類生理時期辨別問題僅能限縮在斯氏繡眼的不同時期之鳴叫聲上，但我認為這種特徵不僅是在同一隻鳥的不同時期會有效果，不同物種之間說不定也會有相似的情況，甚至可以將兩相似之物種比較。希望未來可以比較各種鳥類於不同時期鳴叫聲之分別，並比較不同鳥類之間之相似度，藉此驗證不同鳥類之間的演化關係，藉此可以在生物分類學上提供我們的一份心力。

(三) 建立親緣關係樹或行為物種間學習關係圖

除了應用在不同的物種之間外，我認為這種關係也有可能與個體和個體之間的親緣關係有所連結。如同聲紋一般，當兩個體之聲線、習慣越相同，除了可能生活在相同地點外，也有可能彼此有親緣關係。但我們後來認為鳥類的聲音很容易趨同演化，甚至還有鳥類學習其他鳥種的狀況發生。希望未來可以藉由更大量的資料及驗證此假說 [35]。

(四) 聲音特徵符號化之可能性

雖然能夠初步發現各生理時期之特徵，但仍舊無法明確標示出各特徵出現的時間點或特徵內容。如果能夠找到某種方法表示所有鳴叫聲的發音規則，也就可以為每個聲音找到各自的符號，也就能反推不同符號代表的意思，也就能讓我們即使聽不懂也能夠藉由符號表推敲情緒及其意義 [36]。

(五) 未來應用

在 xeno-canto.org 的資料尚未齊全，錄音數量也較稀少。雖然現今能夠透過排程錄音機來獲得大量的錄音資料，但是現在幾乎都是由人耳來辨識其內物種，非常曠時費日，若能夠自動辨識，或許能夠讓資料集指數成長並更完整。

柒、結論

- 一、斯氏繡眼於不同生理時期之鳴叫聲雖然以未受訓練之普通人無法分辨出區別，但藉由實驗得知，斯氏繡眼在不同時期之鳴叫聲可以依不同特徵被分辨出來。
- 二、我們發現機器學習模型可以將亞種問題分辨出來，或者是將同科不同種有些是聲音相近定義出機率不同，這點是我們模型設計精準所致，此點未來可以給生物學家更加細緻分類的意見。
- 三、藉由 Confusion Matrix 及 t-SNE 降維視覺化及 Attention 視覺化之結果，可以進一步藉由類神經網路證明不同生理時期及不同鳥種之鳴叫聲各具不同特徵。
- 四、相較於 LSTM 及 GRU 之訓練模型，使用 Convolution Module 後配合 Attention 之模型能夠達到至少 95%以上之分辨率。
- 五、利用 BERT-based Model 訓練出之模型最佳 F1-score 分別可以達 97.7%及 79.0%，能夠有效解決生理時期分別問題和鳥種分辨問題，這種模型可以提供未來鳥類生物在聲音分類學上的依據。

捌、參考資料及其他

- [1] 領角鴞圖片 <https://zh.wikipedia.org/wiki/%E9%A0%98%E8%A7%92%E9%B4%9E>
- [2] 于蘇英。鳥人孫清松的心路與奮鬥。環境資訊中心。Feb. 10, 2005。
<https://e-info.org.tw/node/2943>
- [3] 台灣鳥類名錄: <https://www.bird.org.tw/basicpage/87>。
- [4] 台灣杉二號圖片 <https://technews.tw/2019/08/22/ai-national-teams-get-new-contract-they-will-improve-taiwania-ii-computing-ability/>
- [5] XC 原網址 <http://xeno-canto.org>
- [6] 斯氏繡眼詳細資訊 <https://taicol.tw/pages/75407>
- [7] 斯氏繡眼圖片 <https://www.guodong.net/?p=12549>
- [8] 雀形目鳥類 <https://zh.wikipedia.org/wiki/%E9%9B%80%E5%BD%A2%E7%9B%AE>
- [9] 鳴禽 <https://zh.wikipedia.org/wiki/%E9%B3%B4%E7%A6%BD>
- [10] 梅爾頻譜圖實作 <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>
- [11] 梅爾頻譜圖 <https://hackmd.io/@sysprog/fourier-transform>
- [12] MFCC <http://blog.maxkit.com.tw/2019/12/mfcc.html>
- [13] 林育德、郭鈴莉、劉宗憲、崔致豪。語音辨識 Speech Recognition (MFCC 應用)。逢甲大學自動控制工程學系。June, 2008。
- [14] Mel spectrogram, MFCC 圖片 <http://speech.ee.ntu.edu.tw/~acetylsv/thesis.pdf>
- [15] Mike Smales. Sound Classification using Deep Learning. Feb 27, 2019.
<https://link.medium.com/zQPNqbeMI6>
- [16] Brandon Rohrer, Jimmy Lin. Data Science and Robots. May 25, 2017.
<https://brohrer.mcknote.com/zh-Hant>
- [17] Kun Li. 圖像分類中 max-pooling 和 average-pooling 之間的異同. Mar 3, 2018.
<https://blog.csdn.net/u012193416/article/details/79432668>
- [18] Teng Yuan Chang. 比較長短期記憶模型 (LSTM) 與改良後的遞歸神經網路模型: GRU. Feb 17, 2019. <https://link.medium.com/tXL6ht36H6>
- [19] Airwaves. 使用 TensorFlow 了解 Dropout. Sep 5, 2018. <https://medium.com/手寫筆記/使用-tensorflow-了解-dropout-bf64a6785431>
- [20] GGWithRabitLIFE. SGD, Momentum, AdaGrad, Adam Optimizer. Aug 5, 2018.
<https://link.medium.com/u;8T60BFP6>
- [21] Ryan Lu. Learning Model : Gradient Descent -Learning Rate Schedules 和深度學習的學習率學習方法. Apr 1, 2019. <https://link.medium.com/jnCoWPYGP6>

- [22] Cell 示意圖 <http://hemingwang.blogspot.com/2019/09/lstm.html>
- [23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000 – 6010, 2017.
- [24] Attention 圖片 <https://nocater.github.io/2019/05/05/transformer%E8%AF%A6%E8%A7%A3/>
- [25] W. Boes and H. V. Hamme, Audiovisual Transformer Architectures for Large-Scale Classification and Synchronization of Weakly Labeled Audio Events, in Proc. of Int’ l Conf. on Multimedia (MM’ 19), 2019.
- [26] Q. Kong, Y. Xu, W. Wang, and M. D. Plumbley, Sound Event Detection of Weakly Labelled Data with CNN-Transformer and Automatic Threshold Optimization, arXiv preprint, arXiv:1912.04761, 2019.
- [27] Prakhar Ganesh, Yao Chen, Xin Lou, Mohammad Ali Khan, Yin Yang, Deming Chen, Marianne Winslett, Hassan Sajjad, and Preslav Nakov. Compressing large-scale transformerbased models: A case study on BERT. arXiv preprint arXiv:2002.11985, 2020.
- [28] 資料分區 <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>
- [29] t-SNE <https://towardsdatascience.com/deconstructing-bert-part-2-visualizing-the-inner-workings-of-attention-60a16d86b5c1>
- [30] L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9, p2579-2605, 2008.
- [31] 中華民國野鳥學會 <https://www.bird.org.tw/>
- [32] Willem-Pier Vellinga, Robert Planque. The Xeno collection and its relation to sound recognition and classification. CLEF2015. January ,2015.
- [33] Ruixiong Zhang, Haiwei Wu, Wubo Li, Dongwei Jiang, Wei Zou, Xiangang Li . Transformer based unsupervised pre-training for acoustic representation learning. arXiv preprint arXiv:2007.14602. Jul 29, 2020.
- [34] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition, arXiv preprint arXiv:1912.10211, 2019.
- [35] 江郁宣。從發音結構與 DNA 序列探究擬金眼鯛科(Pempheridae)之親緣關係。國立中山大學海洋生物研究所。2010。
- [36] 林威廷。塔米爾文語音辨識系統之設計研究。資訊科技國際期刊第八卷第一期第 60~69 頁。2014。
- [37] 齋藤康毅。Deep Learning 用 Python 進行深度學習的基礎理論實作。台北市基峰資訊。2017。

附錄

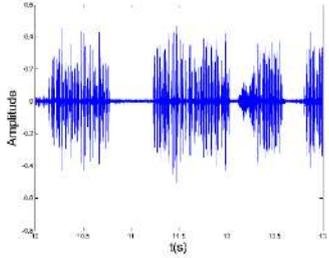
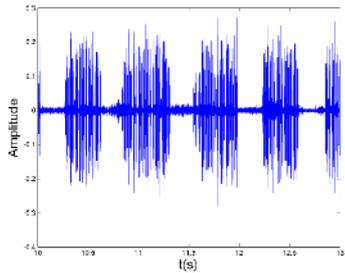
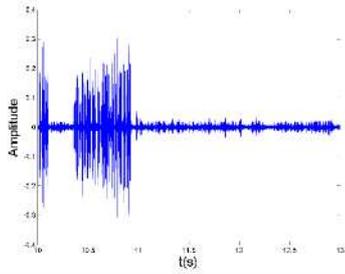
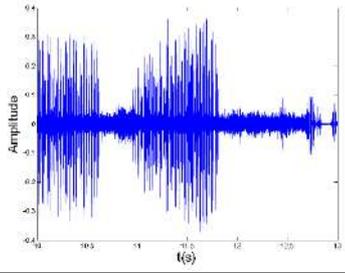
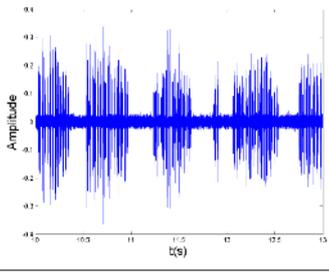
附錄表 1a、虛擬環境套件列表

# Name	Version	Build	Channel
_libgcc_mutex	0.1	main	
absl-py	0.9.0	pypi_0	pypi
appdirs	1.4.4	pypi_0	pypi
astunparse	1.6.3	pypi_0	pypi
audioread	2.1.8	pypi_0	pypi
bzip2	1.0.8	h7b6447c_0	
ca-certificates	2020.6.24	0	
cachetools	4.1.1	pypi_0	pypi
cairo	1.14.12	h8948797_3	
certifi	2020.6.20	py37_0	
cffi	1.14.1	pypi_0	pypi
chardet	3.0.4	pypi_0	pypi
cycler	0.10.0	pypi_0	pypi
decorator	4.4.2	pypi_0	pypi
expat	2.2.9	he6710b0_2	
ffmpeg	4.2.2	h20bf706_0	
fontconfig	2.13.0	h9420a91_0	
freetype	2.10.2	h5ab3b9f_0	
fribidi	1.0.10	h7b6447c_0	
gast	0.3.3	pypi_0	pypi
glib	2.63.1	h5a9c865_0	
gmp	6.1.2	h6c8ec71_1	
gnutls	3.6.5	h71b1129_1002	
google-auth	1.20.0	pypi_0	pypi
google-auth-oauthlib	0.4.1	pypi_0	pypi
google-pasta	0.2.0	pypi_0	pypi
graphite2	1.3.14	h23475e2_0	
graphviz	2.40.1	h21bd128_2	
grpcio	1.30.0	pypi_0	pypi
h5py	2.10.0	pypi_0	pypi
harfbuzz	1.8.8	hffaf4a1_0	
icu	58.2	he6710b0_3	
idna	2.10	pypi_0	pypi
importlib-metadata	1.7.0	pypi_0	pypi
joblib	0.16.0	pypi_0	pypi
jpeg	9b	h024ee3a_2	
keras-preprocessing	1.1.2	pypi_0	pypi
kiwisolver	1.2.0	pypi_0	pypi
lame	3.100	h7b6447c_0	
ld_impl_linux-64	2.33.1	h53a641e_7	
libedit	3.1.20191231	h14c3975_1	
libffi	3.2.1	hd88cf55_4	
libgcc-ng	9.1.0	hdf63c60_0	
libopus	1.3.1	h7b6447c_0	
libpng	1.6.37	hbc83047_0	
librosa	0.8.0	pypi_0	pypi
libstdcxx-ng	9.1.0	hdf63c60_0	
libtiff	4.1.0	h2733197_1	
libuuid	1.0.3	h1bed415_2	
libvpx	1.7.0	h439df22_0	
libxcb	1.14	h7b6447c_0	
libxml2	2.9.10	he19cac6_1	
llvmlite	0.33.0	pypi_0	pypi

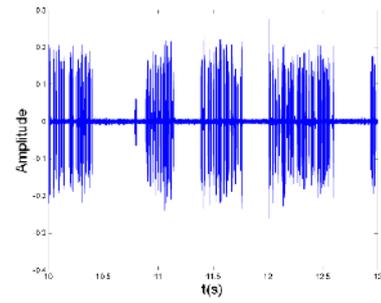
附錄表 1b、虛擬環境套件列表

lz4-c	1.9.2	he6710b0_1	
markdown	3.2.2	pypi_0	pypi
matplotlib	3.3.0	pypi_0	pypi
ncurses	6.2	he6710b0_1	
nettle	3.4.1	hbb512f6_0	
numba	0.50.1	pypi_0	pypi
numpy	1.19.1	pypi_0	pypi
oauthlib	3.1.0	pypi_0	pypi
openh264	2.1.0	hd408876_0	
openssl	1.1.1g	h7b6447c_0	
opt-einsum	3.3.0	pypi_0	pypi
packaging	20.4	pypi_0	pypi
pango	1.42.4	h049681c_0	
pcre	8.44	he6710b0_0	
pillow	7.2.0	pypi_0	pypi
pip	20.1.1	py37_1	
pixman	0.40.0	h7b6447c_0	
pooch	1.1.1	pypi_0	pypi
protobuf	3.12.2	pypi_0	pypi
pyasn1	0.4.8	pypi_0	pypi
pyasn1-modules	0.2.8	pypi_0	pypi
pycparser	2.20	pypi_0	pypi
pydot	1.4.1	pypi_0	pypi
pydub	0.24.1	pypi_0	pypi
pyparsing	2.4.7	pypi_0	pypi
python	3.7.7	hcf32534_0_cpython	
python-dateutil	2.8.1	pypi_0	pypi
python-graphviz	0.14.1	pypi_0	pypi
pyyaml	5.3.1	pypi_0	pypi
readline	8.0	h7b6447c_0	
requests	2.24.0	pypi_0	pypi
requests-oauthlib	1.3.0	pypi_0	pypi
resampy	0.2.2	pypi_0	pypi
rsa	4.6	pypi_0	pypi
scikit-learn	0.23.1	pypi_0	pypi
scipy	1.4.1	pypi_0	pypi
setuptools	49.2.0	py37_0	
six	1.15.0	pypi_0	pypi
soundfile	0.10.3.post1	pypi_0	pypi
sqlite	3.32.3	h62c20be_0	
tensorboard	2.2.2	pypi_0	pypi
tensorboard-plugin-wit	1.7.0	pypi_0	pypi
tensorflow	2.2.0	pypi_0	pypi
tensorflow-estimator	2.2.0	pypi_0	pypi
termcolor	1.1.0	pypi_0	pypi
tf	1.0.0	pypi_0	pypi
threadpoolctl	2.1.0	pypi_0	pypi
tk	8.6.10	hbc83047_0	
tqdm	4.48.0	pypi_0	pypi
urllib3	1.25.10	pypi_0	pypi
werkzeug	1.0.1	pypi_0	pypi
wheel	0.34.2	py37_0	
wrapt	1.12.1	pypi_0	pypi
x264	1!157.20191217	h7b6447c_0	
xz	5.2.5	h7b6447c_0	
zipp	3.1.0	pypi_0	pypi
zlib	1.2.11	h7b6447c_3	
zstd	1.4.5	h9ceee32_0	

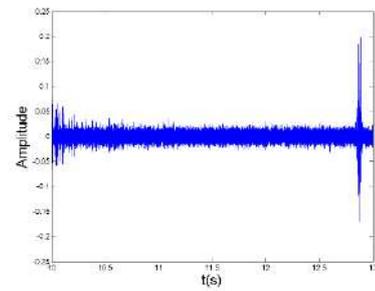
附錄表 2、斯氏繡眼資料集部分原始資料之預覽

日期	天氣	鳴者代號	時期	波形圖 (僅顯示約前 13 秒)
2001/4/3	陰時雨	右 B.W 左 O	抱卵期	
2001/4/6	晴	右 BW 左 O	抱卵期	
2001/4/9	陰	右 BW 左 O	育雛期	
2001/5/28	雲	右 BW 左 O	育雛期	
2001/4/10	晴	右 BL 左 R	巢前期	

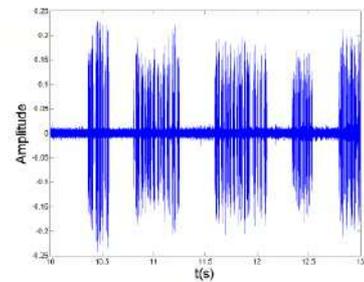
2001/5/5 陰 右 BL 左 R 抱卵期



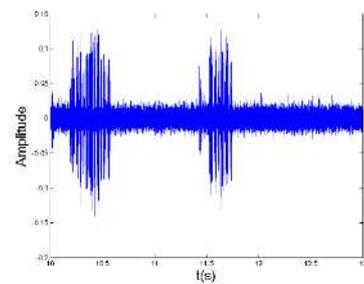
2001/7/2 晴 右 BL 左 R 下第 2 蛋
(抱卵期)



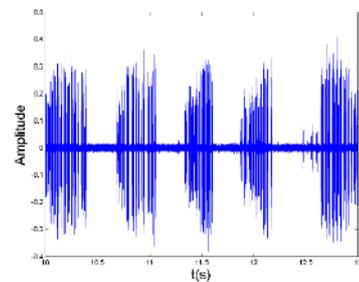
2001/7/4 雲 右 BL 左 R 抱卵期

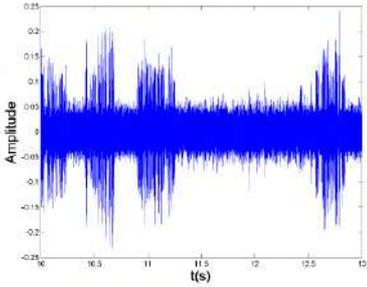
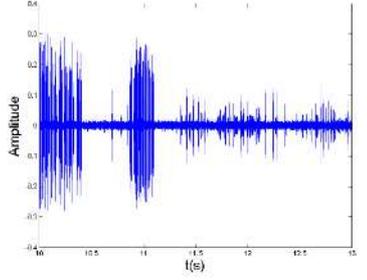
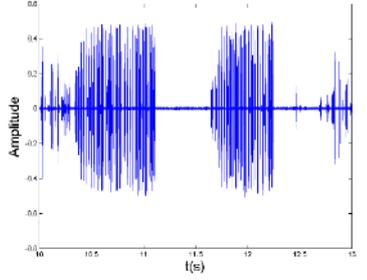
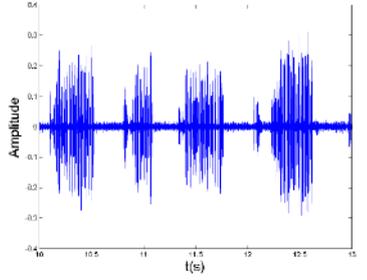
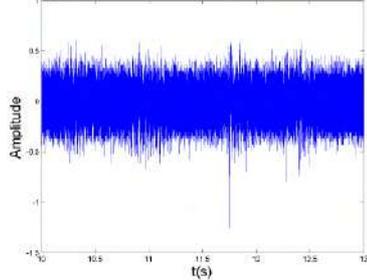


2001/7/7 晴 右 BL 左 R 抱卵期

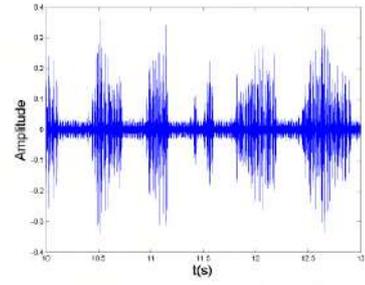


2001/7/13 晴 右 BL 左 R 育雛期

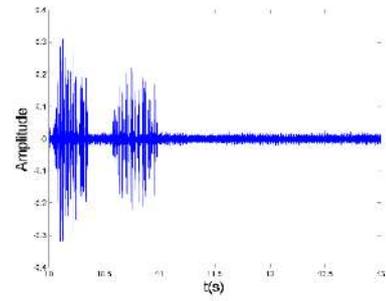


2001/7/17	晴	右 BL 左 R	育雛期	
2001/7/20	陰	右 BL 左 R	育雛期	
2001/5/3	陰	右 DG.LBL 左 BL	築巢期	
2001/5/9	陰雨	右 DG.LBL 左 BL	下第 2 蛋 (抱卵期)	
2001/7/1	晴	右 OB 左 W	抱卵期	

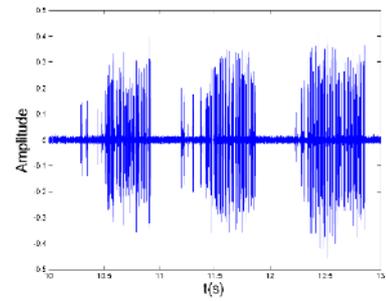
2001/7/2 晴 右OB 左W 抱卵期



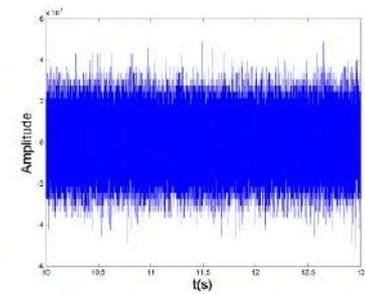
2001/7/4 雲 右OB 左W 抱卵期



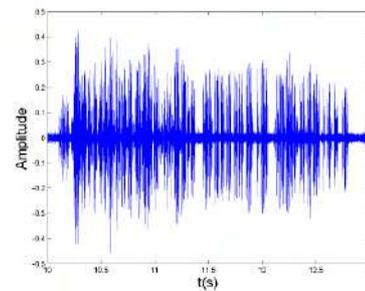
2001/7/7 晴 右OB 左W 抱卵期



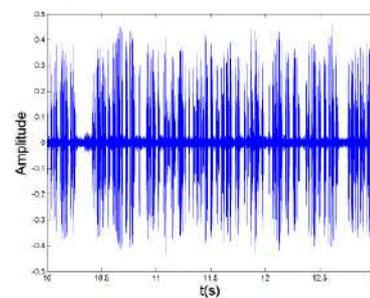
2010/4/22 陰時雨 右OG 築前期



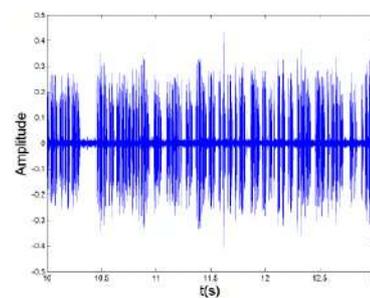
2010/4/27 霧 右OG 築巢期



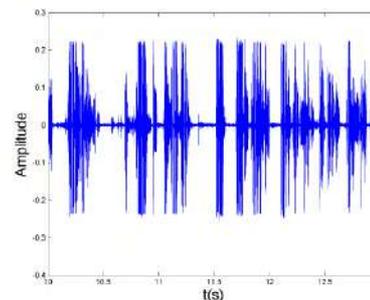
2010/5/17 雲 右 OG 抱卵期



1999/5/19 陰 右 OG 抱卵期

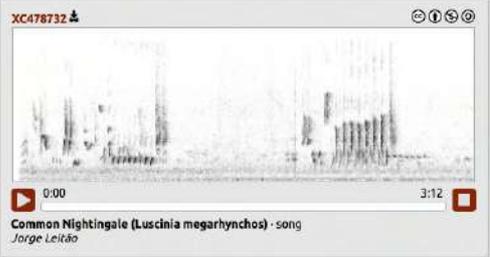
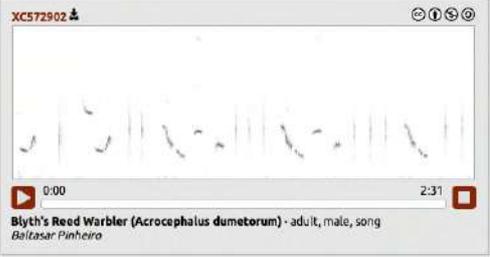
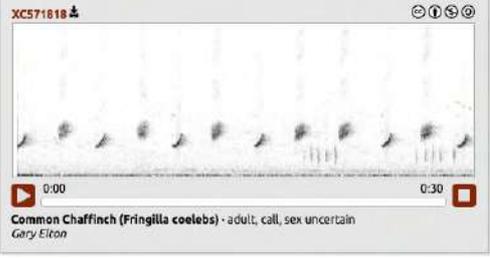
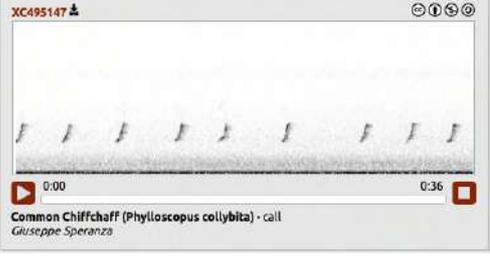
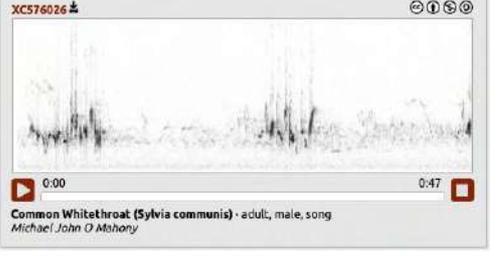
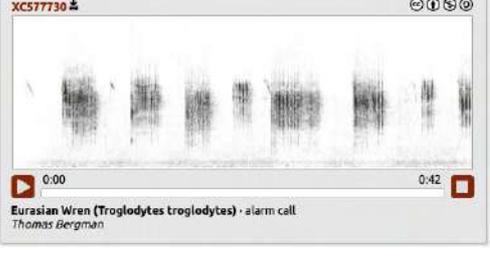


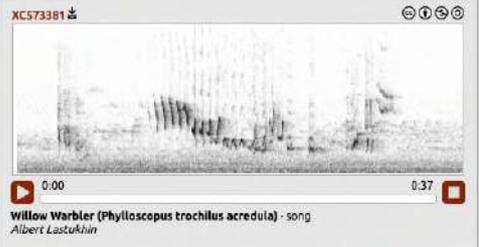
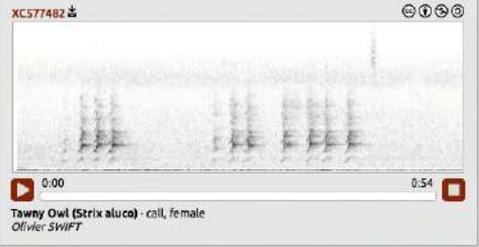
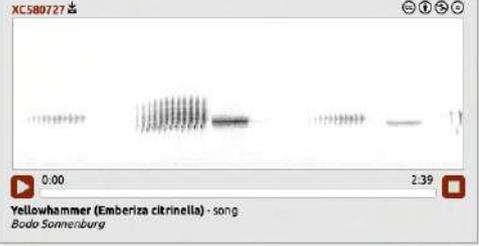
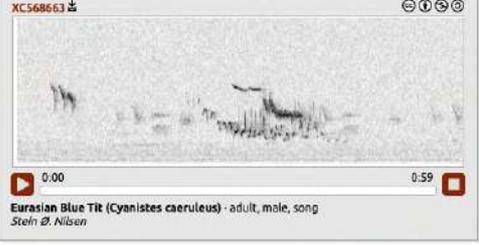
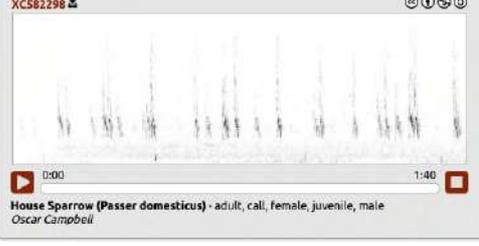
1999/5/31 陰 右 OG 育雛期



附錄表 3、CX 資料集部分原始資料預覽

俗名	學名	XC 資料集 ID	頻譜圖前 10 秒預覽 (截圖)
烏鶇	<i>Turdus merula</i>	339	<p>XC563419 Common Blackbird (<i>Turdus merula</i>) - call, juvenile Juho Könönen</p>
歌鶇	<i>Turdus philomelos</i>	90	<p>XC540810 Song Thrush (<i>Turdus philomelos</i>) - song Alan Verneau</p>
黑頂林鶯	<i>Sylvia atricapilla</i>	220	<p>XC560331 Eurasian Blackcap (<i>Sylvia atricapilla</i>) - adult, male, mimicry/imitation, song François Grimal</p>
濕地葦鶯	<i>Acrocephalus palustris</i>	497	<p>XC566103 Marsh Warbler (<i>Acrocephalus palustris</i>) - song Eetu Pajakkä</p>
歐亞歌鵲	<i>Erithacus rubecula</i>	203	<p>XC576979 European Robin (<i>Erithacus rubecula</i>) - alarm call, call Uku Paal</p>
大山雀	<i>Parus major</i>	450	<p>XC536914 Great Tit (<i>Parus major</i>) - adult, male, song Elias A. Ryberg</p>

新疆歌鵲	<i>Luscinia megarhynchos</i>	623	 <p>XC478732</p> <p>Common Nightingale (<i>Luscinia megarhynchos</i>) - song Jorge Leitão</p>
布萊氏葦鶯	<i>Acrocephalus dumetorum</i>	191	 <p>XC572902</p> <p>Blyth's Reed Warbler (<i>Acrocephalus dumetorum</i>) - adult, male, song Baltasar Pinheiro</p>
蒼頭燕雀	<i>Fringilla coelebs</i>	117	 <p>XC571818</p> <p>Common Chaffinch (<i>Fringilla coelebs</i>) - adult, call, sex uncertain Gary Elton</p>
巖渣柳鶯	<i>Phylloscopus collybita</i>	204	 <p>XC495147</p> <p>Common Chiffchaff (<i>Phylloscopus collybita</i>) - call Giuseppe Speranza</p>
灰白喉林鶯	<i>Sylvia communis</i>	207	 <p>XC576026</p> <p>Common Whitethroat (<i>Sylvia communis</i>) - adult, male, song Michael John O'Mahony</p>
鷓鴣	<i>Troglodytes troglodytes</i>	315	 <p>XC577730</p> <p>Eurasian Wren (<i>Troglodytes troglodytes</i>) - alarm call Thomas Bergman</p>

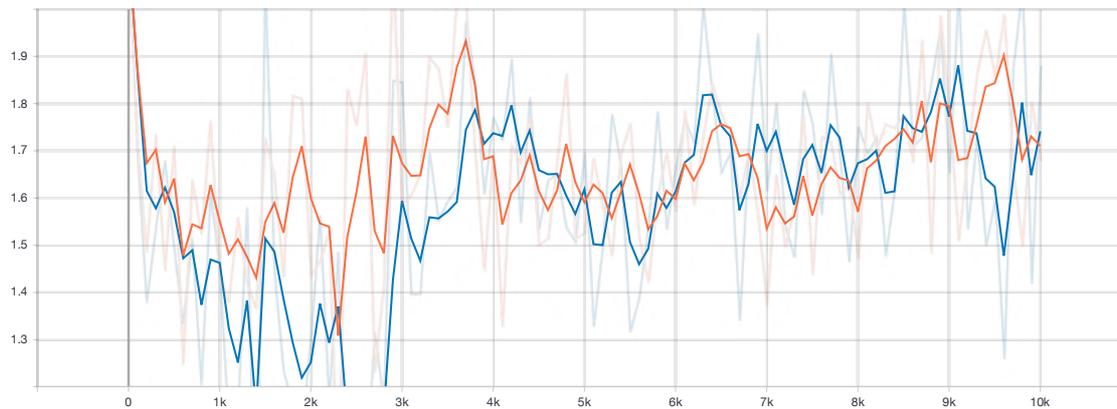
歐亞柳鶯	<i>Phylloscopus trochilus</i>	104	 <p>XCS73381</p> <p>Willow Warbler (<i>Phylloscopus trochilus acredula</i>) - song Albert Lastukhin</p>
灰林鴞	<i>Strix aluco</i>	24	 <p>XCS77482</p> <p>Tawny Owl (<i>Strix aluco</i>) - call, female Olivier SWIFT</p>
黃鸝	<i>Emberiza citrinella</i>	219	 <p>XCS80727</p> <p>Yellowhammer (<i>Emberiza citrinella</i>) - song Bodo Sonnenburg</p>
歐亞藍山雀	<i>Cyanistes caeruleus</i>	1075	 <p>XCS68663</p> <p>Eurasian Blue Tit (<i>Cyanistes caeruleus</i>) - adult, male, song Stein Ø. Nilsen</p>
家麻雀	<i>Passer domesticus</i>	243	 <p>XCS82298</p> <p>House Sparrow (<i>Passer domesticus</i>) - adult, call, female, juvenile, male Oscar Campbell</p>
黍鷓	<i>Emberiza calandra</i>	219	 <p>XCS72991</p> <p>Corn Bunting (<i>Emberiza calandra</i>) - song Stephan Risch</p>

附錄表 4a、Conv1D-Based Model 不同模型之表現 (Loss)

Model (橫軸：Epoch，縱軸：Loss (紅：Validation、藍：Training))

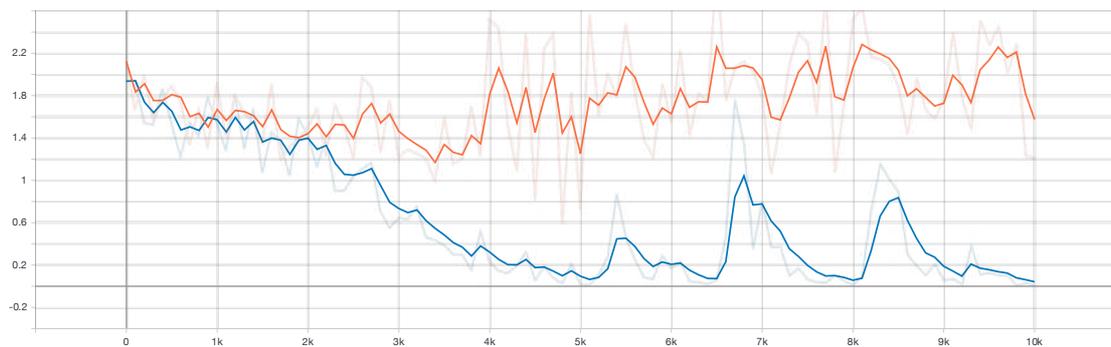
LSTM

(n=64)

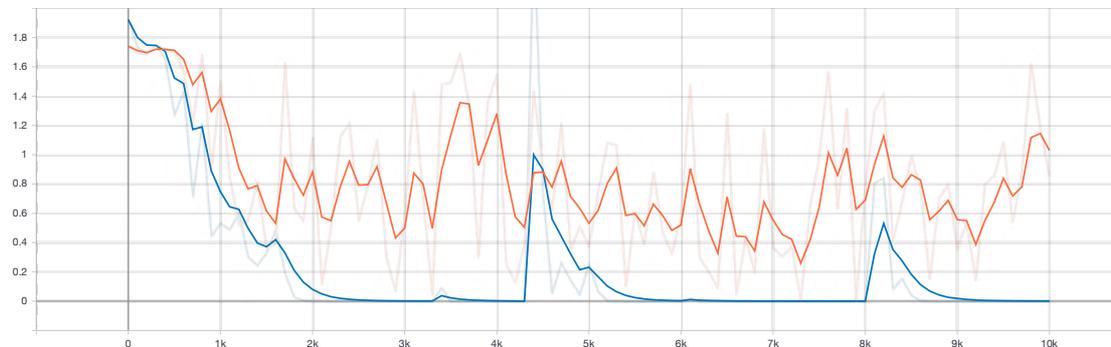


LSTM

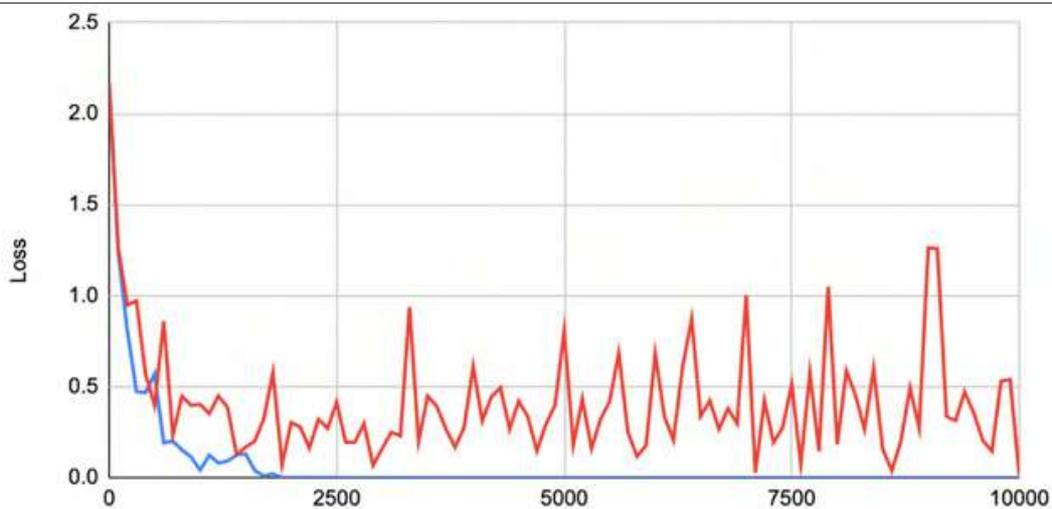
(n=128)



GRU



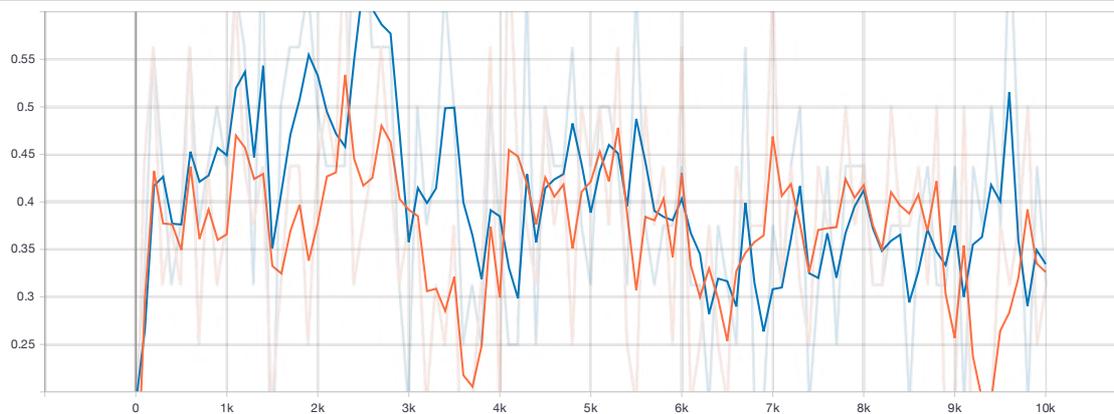
Attention



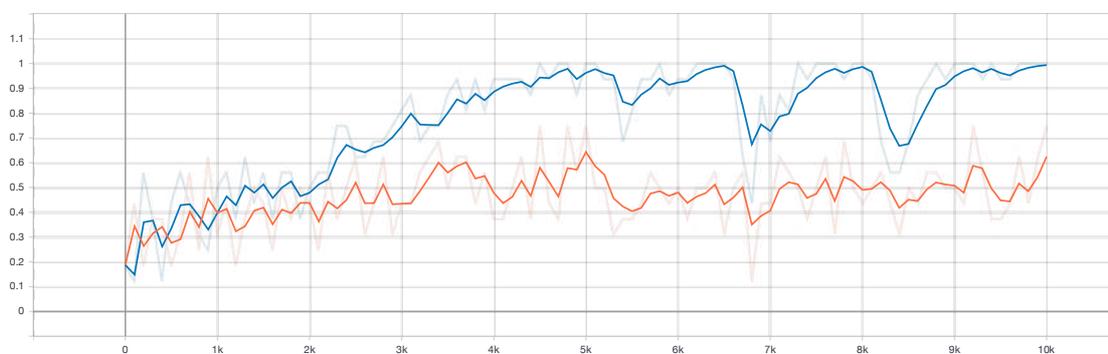
附錄表 4b、Conv1D-Based Model 不同模型之表現 (Accuracy)

Model (橫軸：Epoch，縱軸：Accuracy(紅：Validation、藍：Training))

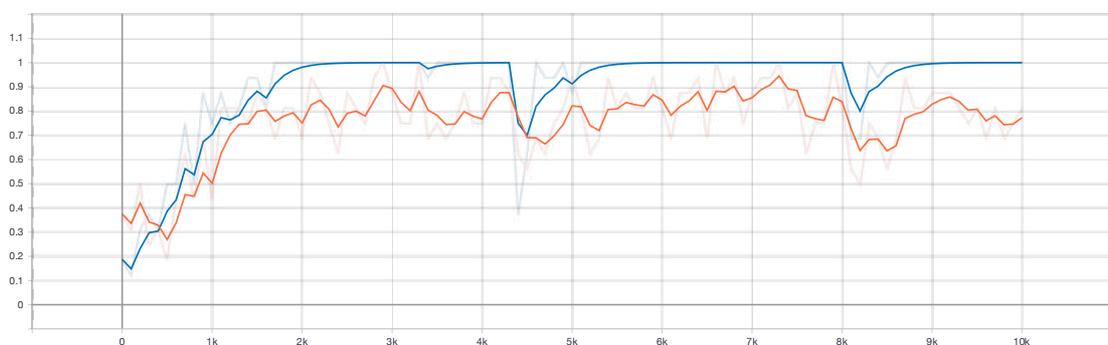
LSTM
(n=64)



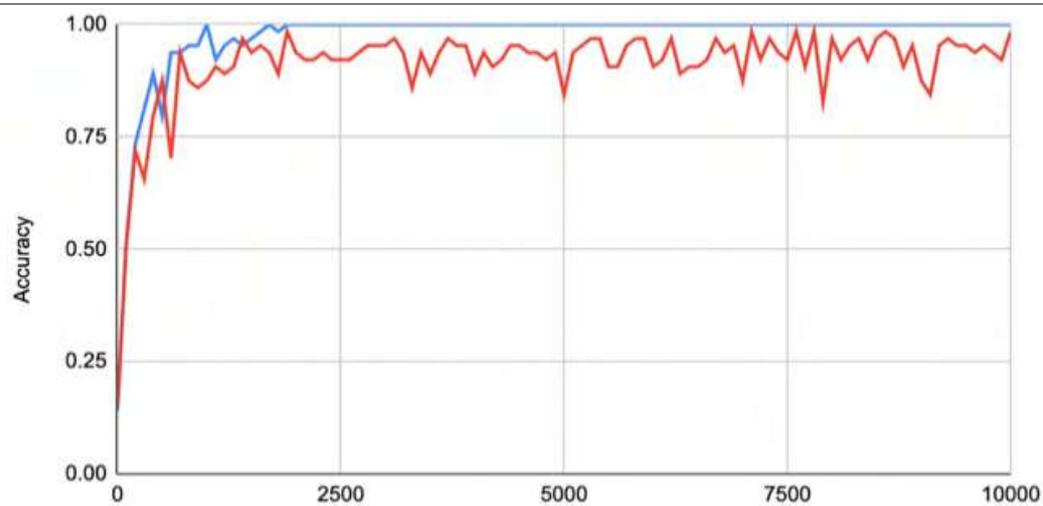
LSTM
(n=128)



GRU



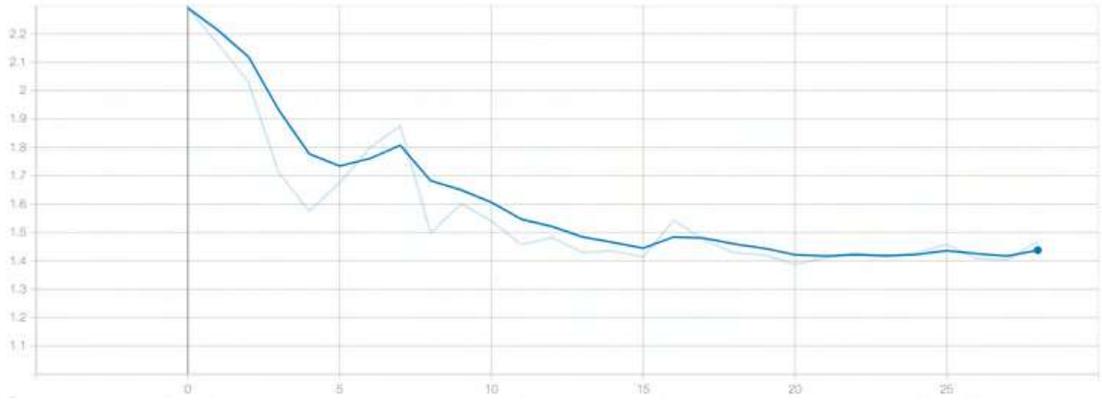
Attention



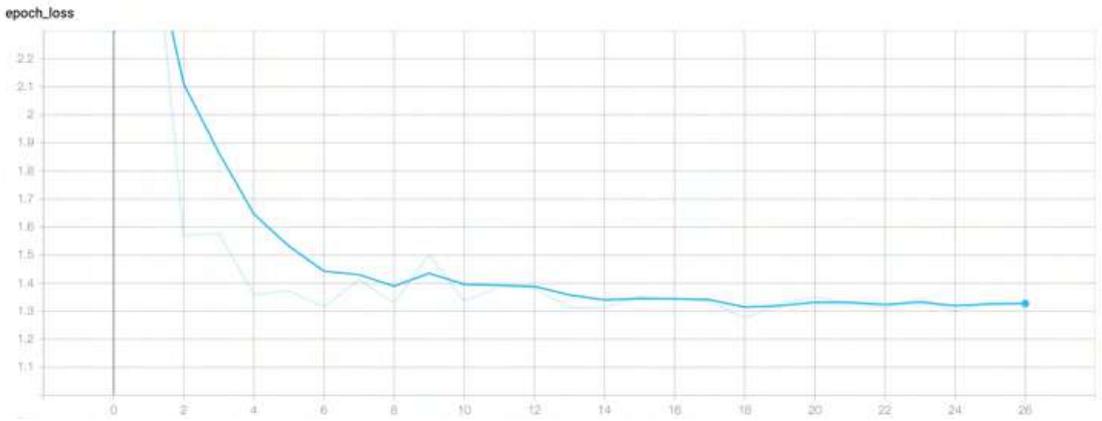
附錄表 5a、FFN 枚舉 Encoder Layer 及 Attention head 個數之結果 (Loss)

(層, 頭) (橫軸：Epoch，縱軸：Validation Loss)，實線是虛線經移動平均的結果

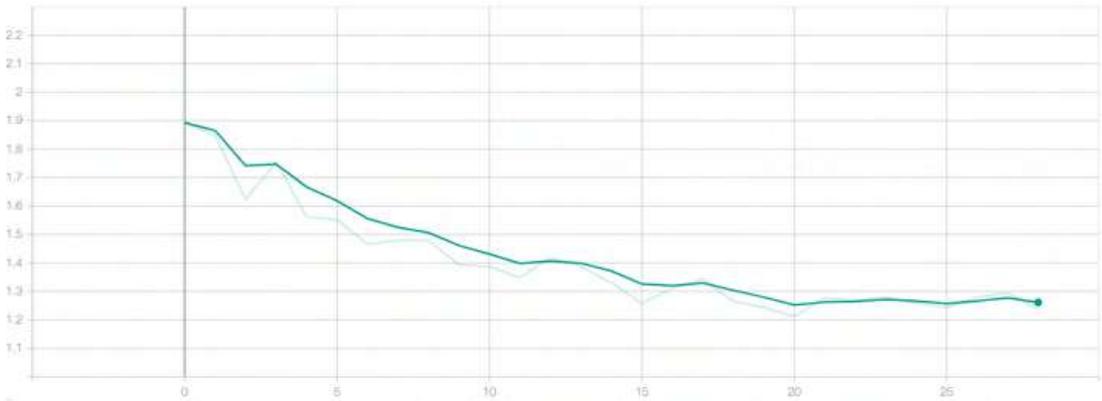
(4, 2)



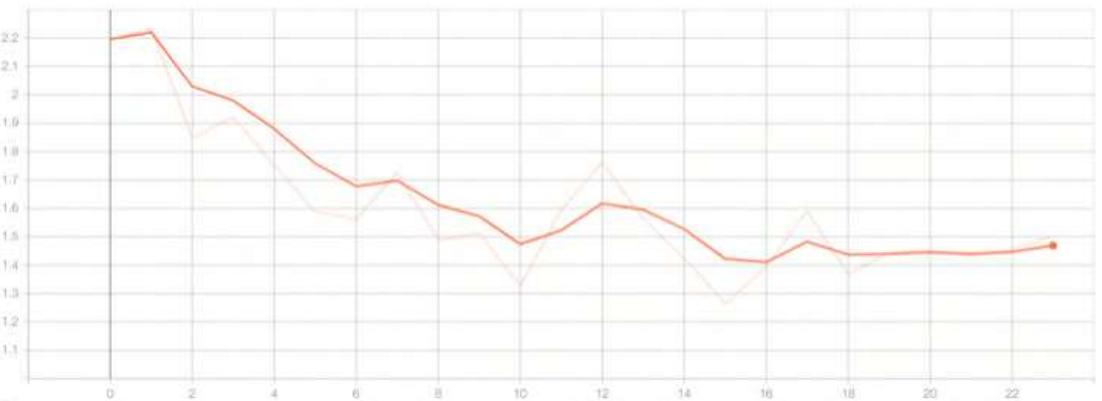
(4, 4)



(4, 8)



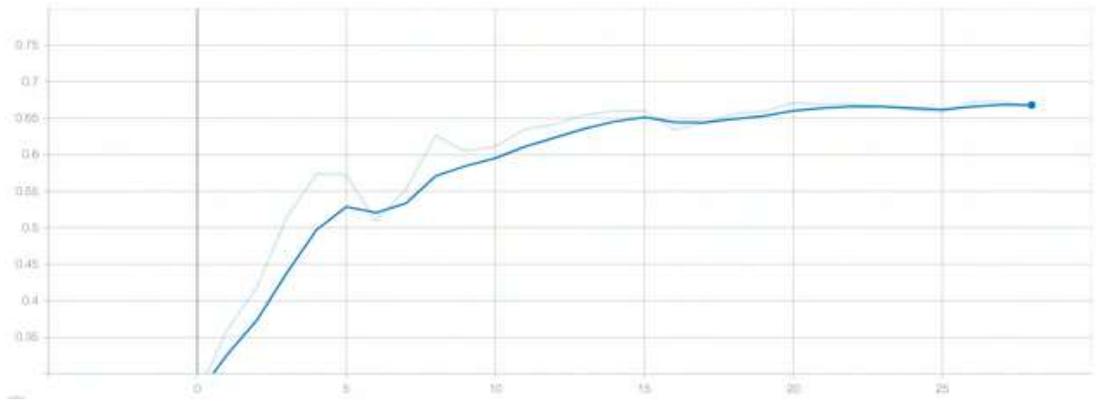
(8, 4)



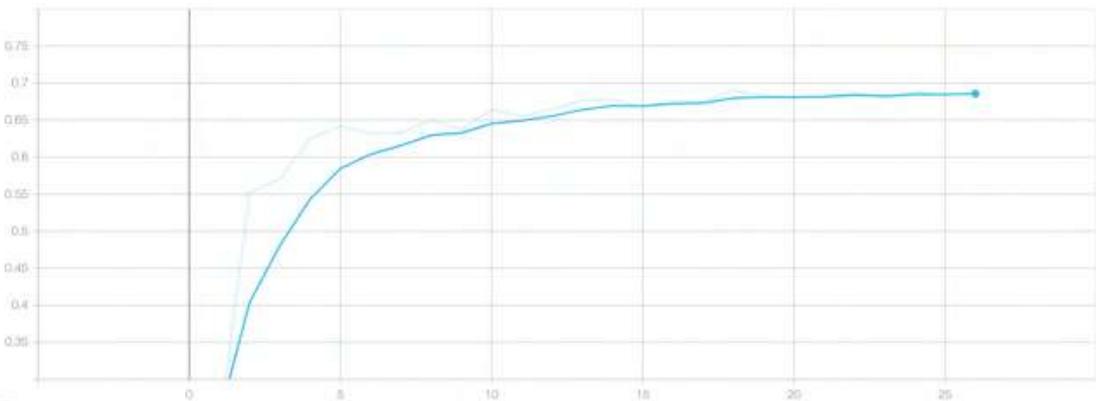
附錄表 5b、FFN 枚舉 Encoder Layer 及 Attention head 個數之結果 (Accuracy)

(層, 頭) (橫軸: Epoch, 縱軸: Validation Accuracy), 實線是虛線經移動平均的結果

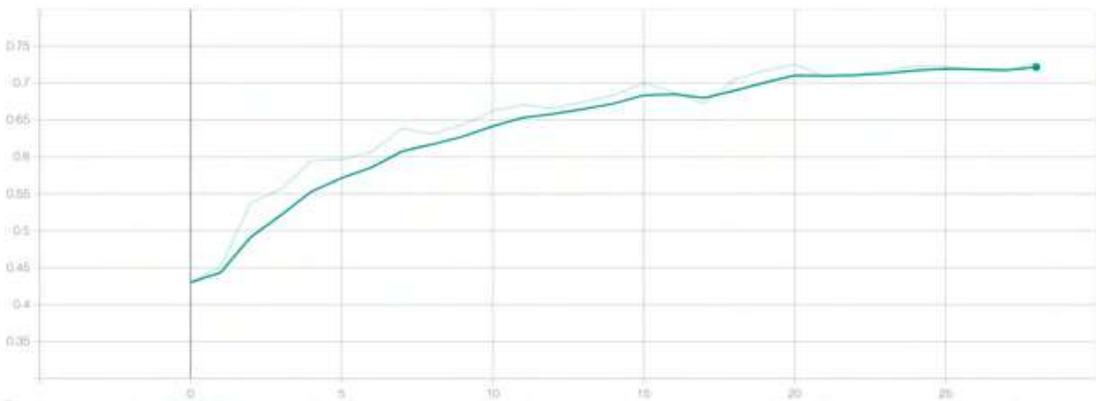
(4, 2)



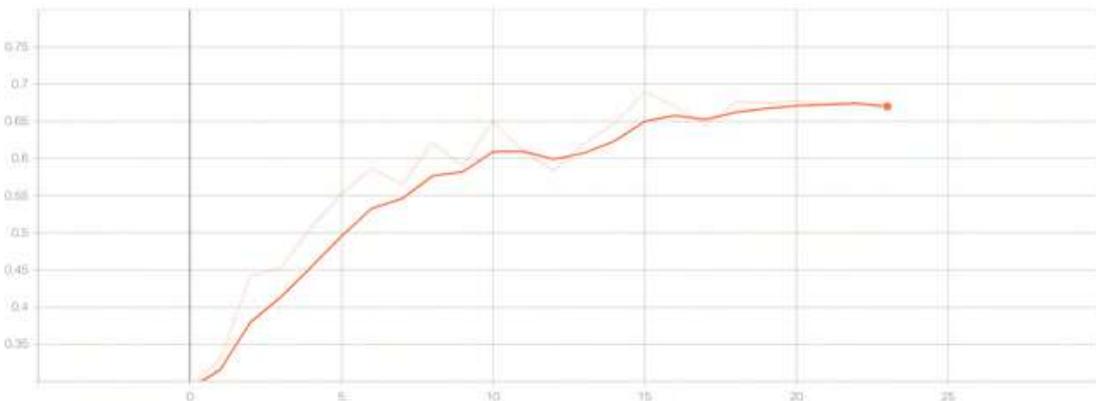
(4, 4)



(4, 8)



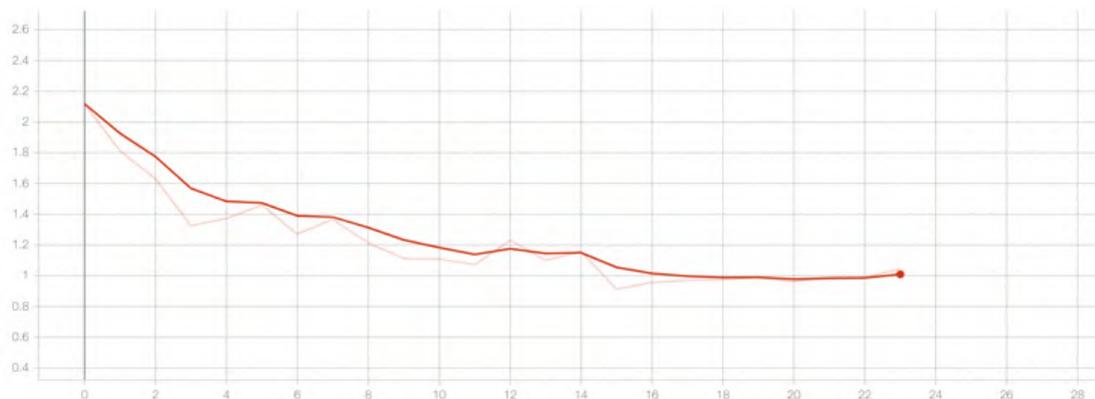
(8, 4)



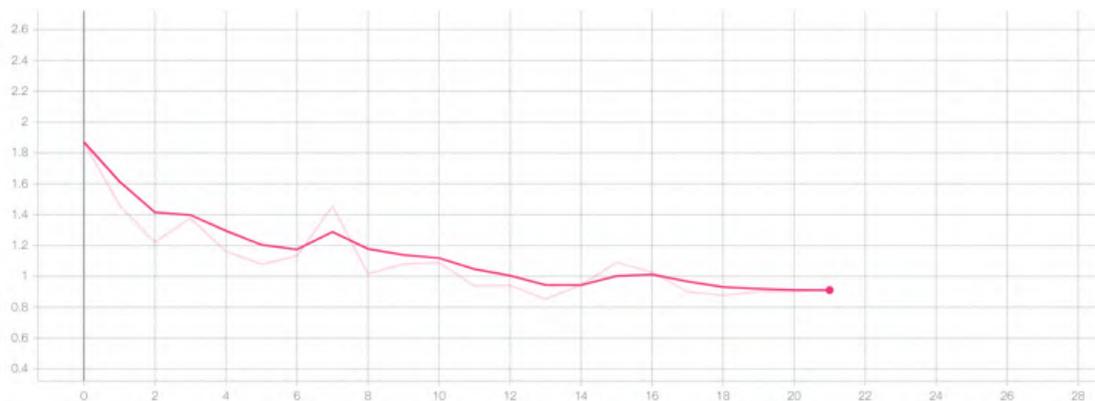
附錄表 6a、Conv1D 枚舉 Encoder Layer 及 Attention head 個數之結果 (Loss)

(層, 頭) (橫軸: Epoch, 縱軸: Validation Loss), 實線是虛線經移動平均的結果

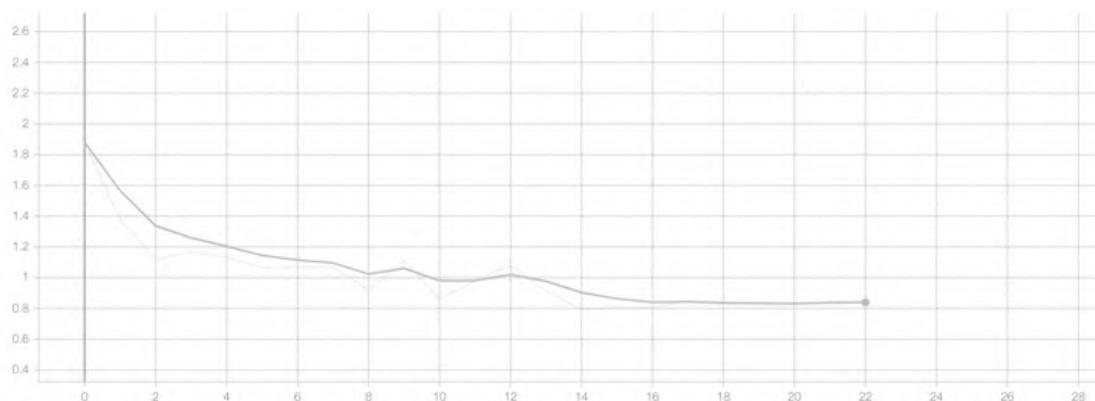
(4, 2)



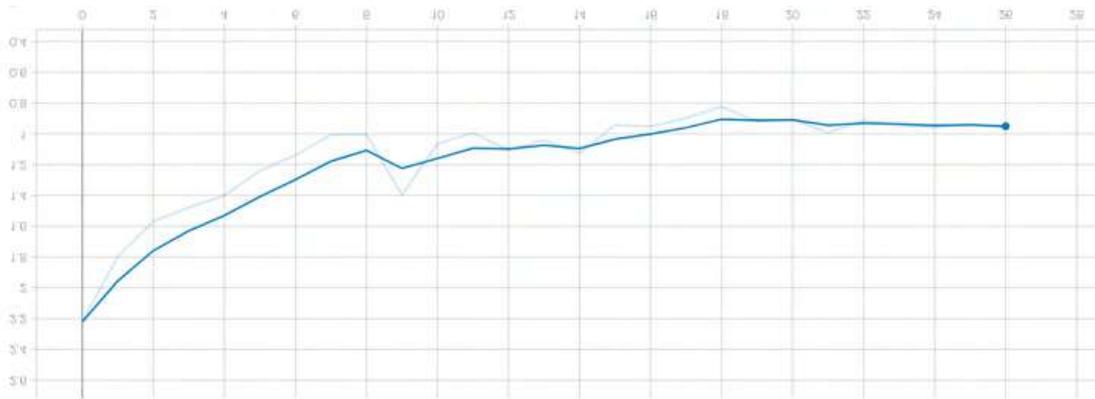
(4, 4)



(4, 8)



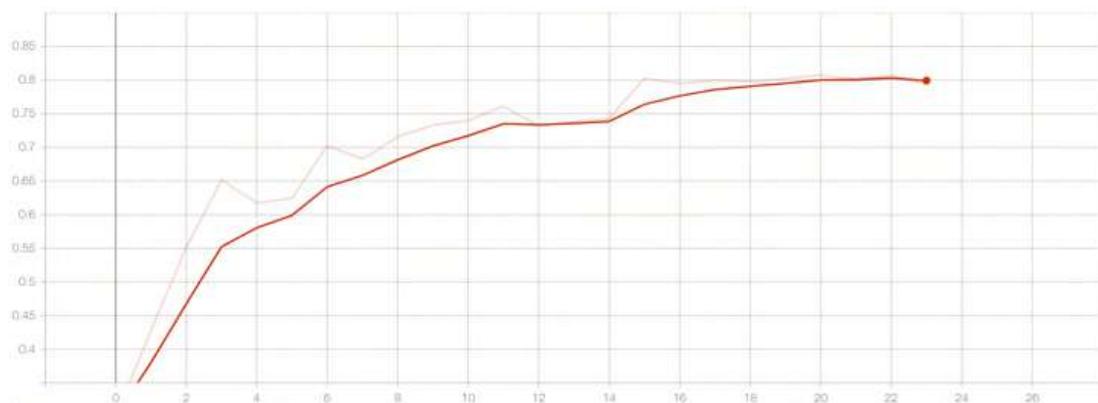
(8, 4)



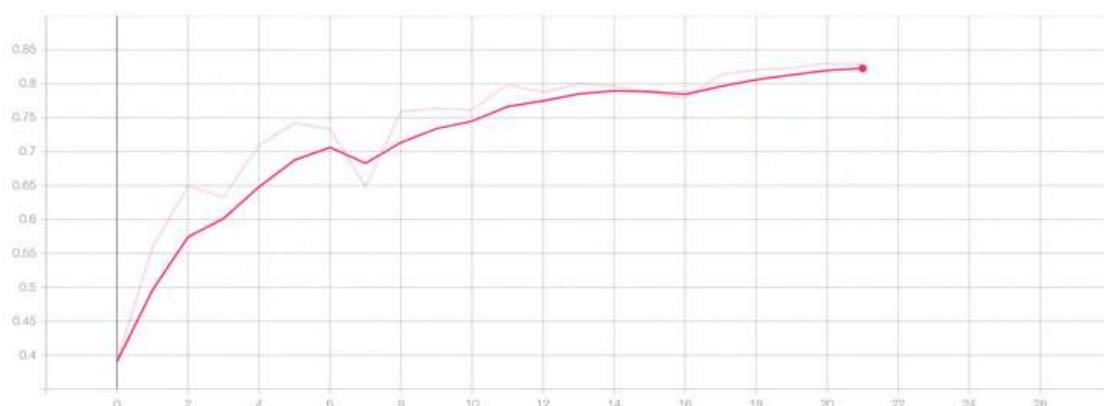
附錄表 6b、Conv1D 枚舉 Encoder Layer 及 Attention head 個數之結果 (Accuracy)

(層, 頭) (橫軸: Epoch, 縱軸: Validation Accuracy), 實線是虛線經移動平均的結果

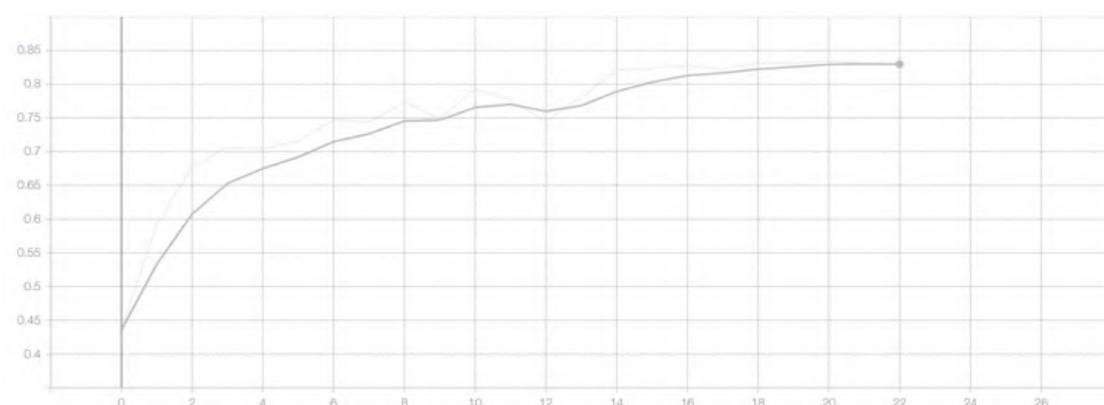
(4, 2)



(4, 4)



(4, 8)



(8, 4)

