

音樂曲風辨識

期別：台北第五期 (AT082)

組別：23

成員：

亓瑋漢(038)

吳志偉(043)

黃俊銘(096)

劉佳豪(107)



問題與目標

- 每一首歌都含有一些特定的元素可以讓聽眾知道這首歌的風格，有一些歌曲乍聽之下就能輕易的辨識其為何種曲風，而有些歌曲則是讓人難以辨識，因此需要設計一個系統來輔助聽眾辨識音樂的曲風。
- 初步目標：辨視目標為音樂曲風辨識(jazz, classical, hip-hop, rock, and blues等)
- 進階目標：音樂曲風轉換(鋼琴 ↔ 吉他)。



曲風辨識

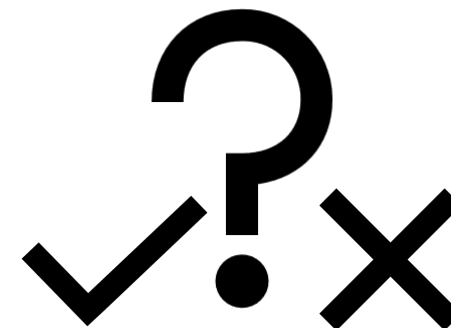
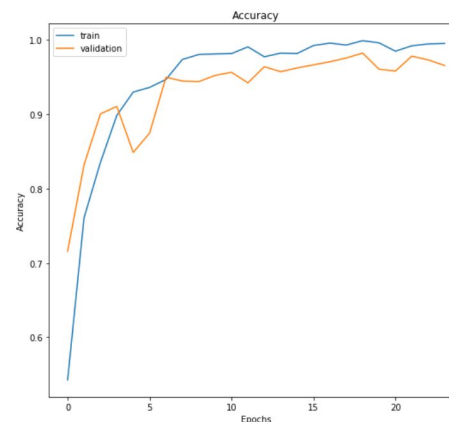
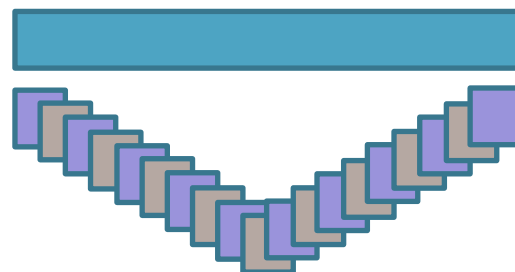
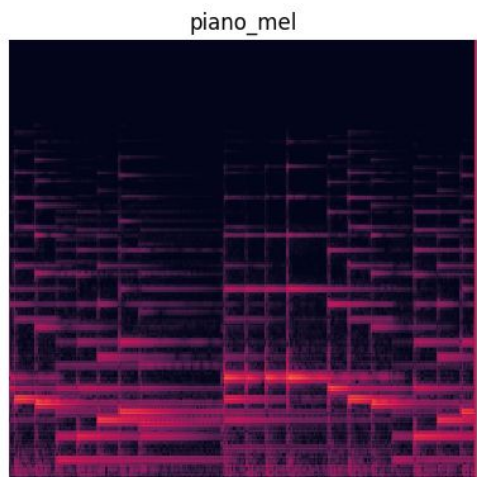
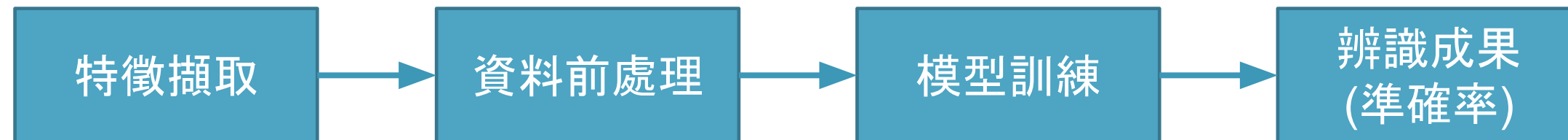
資料集 – GTZAN

Spec

- 10 genres
- 100 tracks for each genre
- 30 seconds
- 22050Hz 16-bit audio
- .au format



流程



重要特徵說明

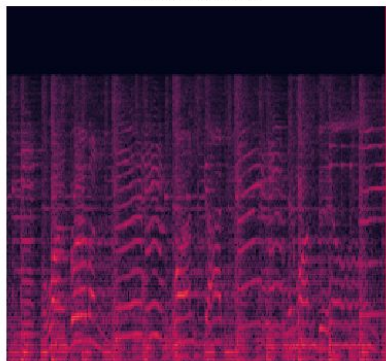
Mel-spectrogram:

- Mel頻率分析是基於人類對聽覺的感知，人耳就像一個濾波器一樣，只關注特定的頻率，而壓根就直接無視它不想感知的某些頻率訊號。
=> 透過此特徵模擬人類聽到的聲音，並有效壓縮音檔資料量

GTZAN-Melspectrogram

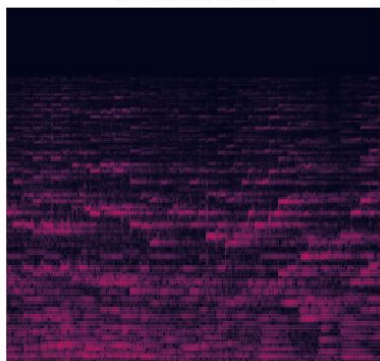
Blues

blues.00000.au



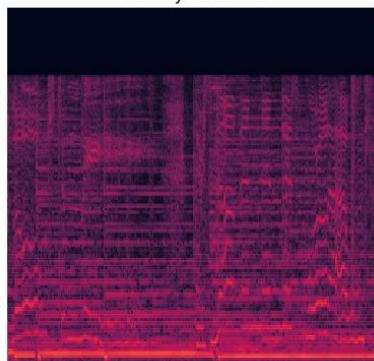
Classical

classical.00000.au



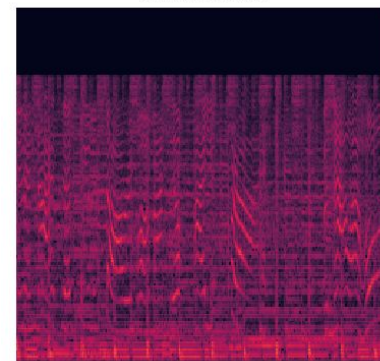
Country

country.00000.au



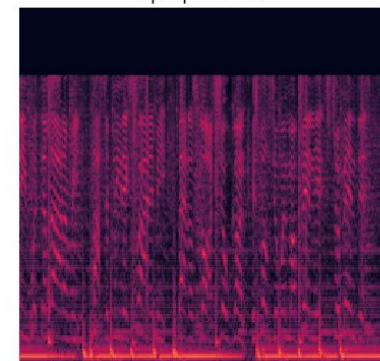
Disco

disco.00001.au



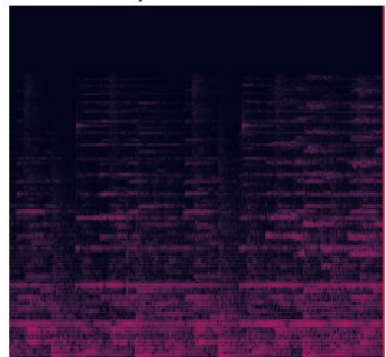
Hip-Hop

hiphop.00001.au



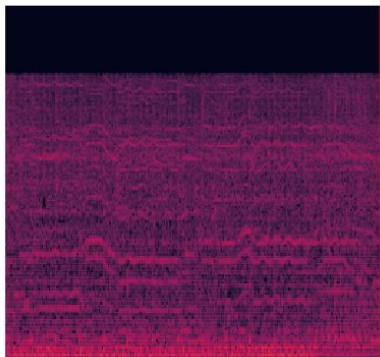
Jazz

jazz.00001.au



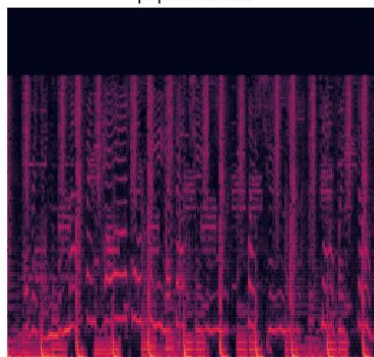
Metal

metal.00001.au



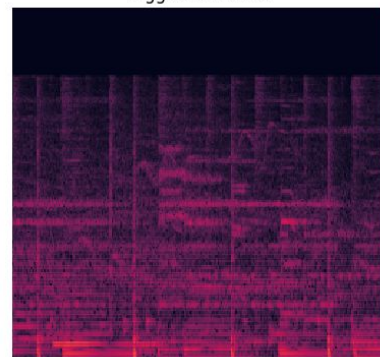
Pop

pop.00001.au



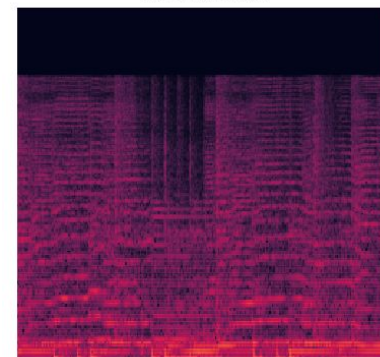
Reggae

reggae.00001.au

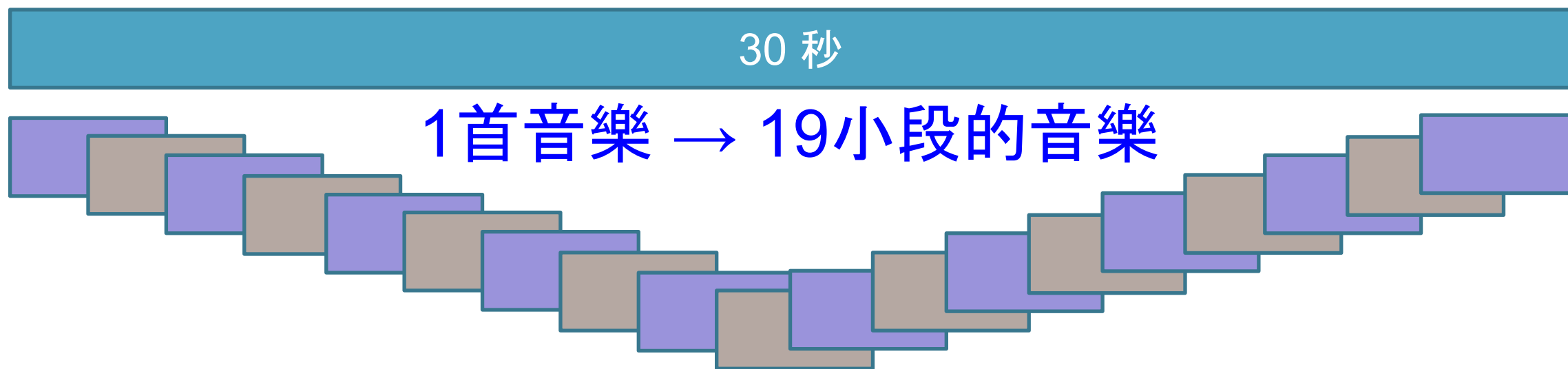


Rock

rock.00001.au



資料前處理



我們將一個30秒的音訊檔，以3秒做分割，並將下一個分割與上一個分割重疊1.5秒，藉以增加資料量

模型選擇及原因 and 資料選用

- 模型選擇: CNN相關的模型
- 原因: 可將音訊檔轉換成圖片檔, 再使用擅長圖片辨識的CNN Models。
- 模型: 傳統CNN(6 Layers)、VGG16、Xception

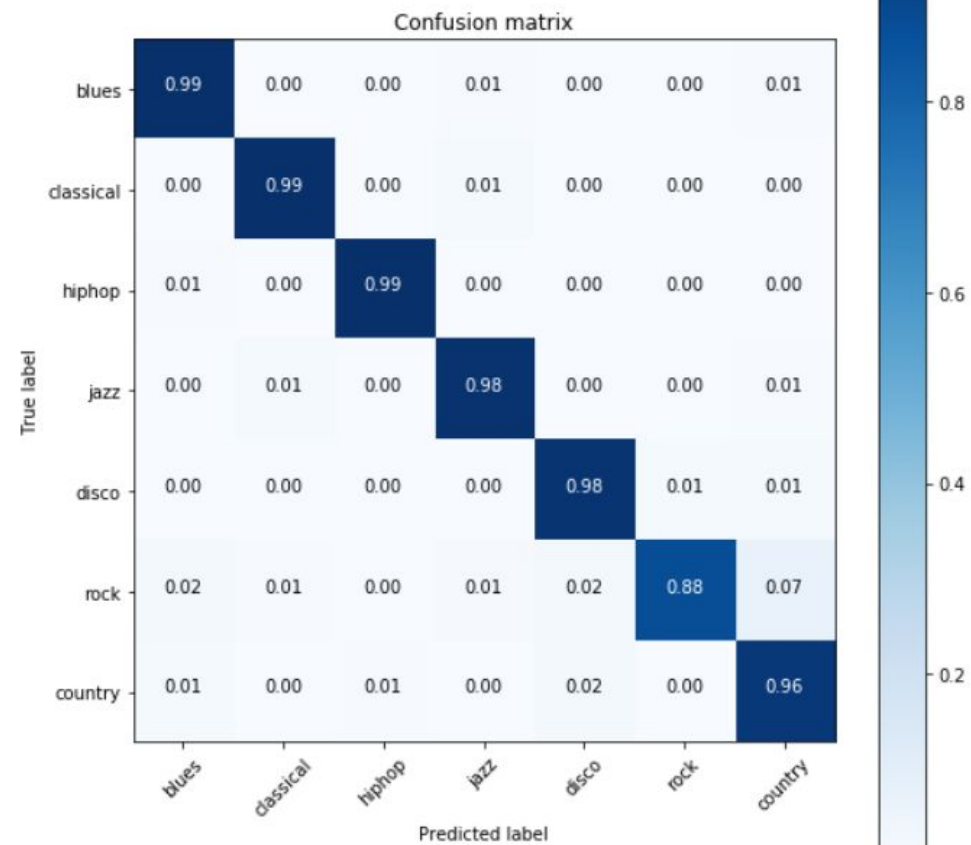
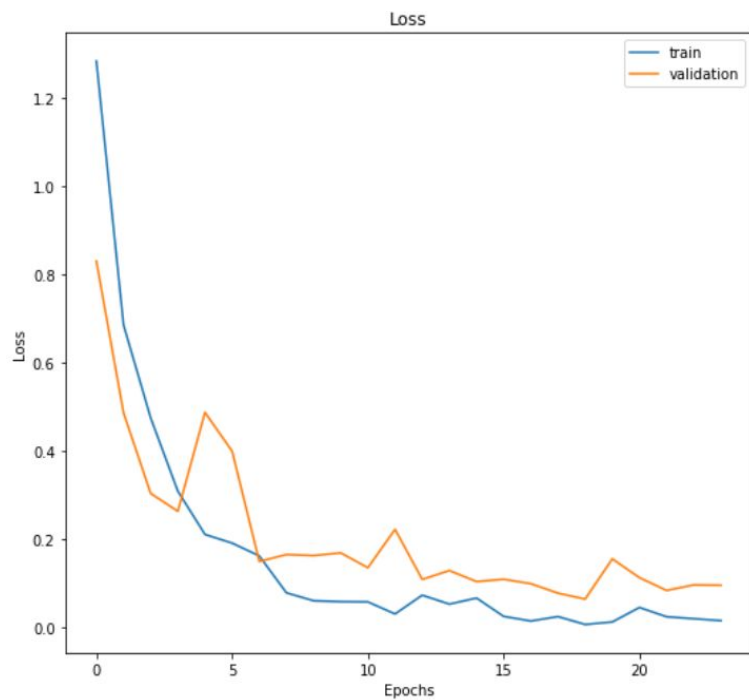
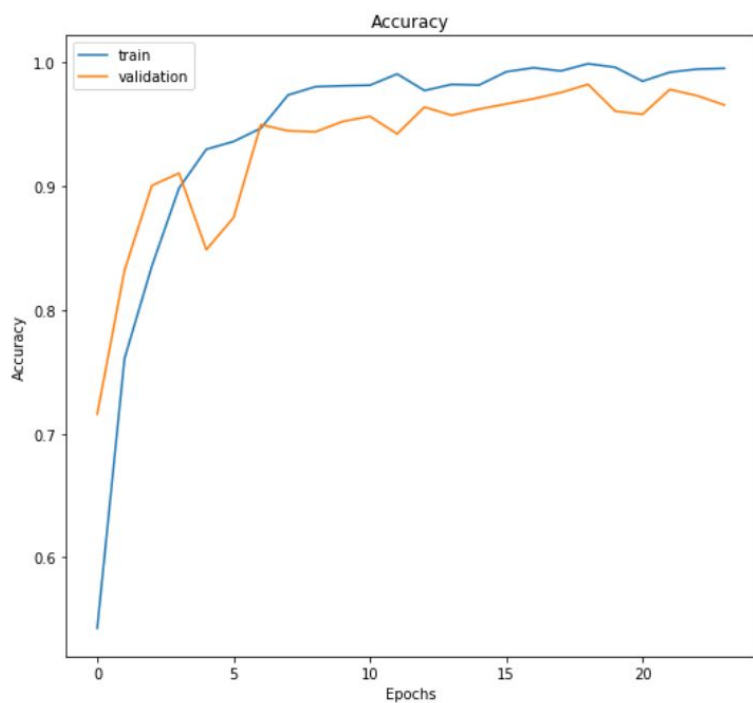
資料選用:

- 7 genres of GTZAN: blues, classical, disco, hip-hop, jazz, rock, country.
- Training Set: genre.00000 ~ genre.00089
訓練資料量 = $90 \times 19 \times 7 = 11,970$; Shape (128, 129, 3); 10% for Validation
- Testing Set: genre.00090 ~ genre.00099

模型比較

Model	Parameters	Calculation Time	Batch Size	Epochs	early stop	Acc.
CNN (6 Layers)	6,271,623	490s	128	50	5	0.871
Xception	20,875,823	648s	128	30	5	0.843
VGG16	16,813,895	364s	128	30	5	0.871

VGG16 訓練結果





音色轉換

進階目標



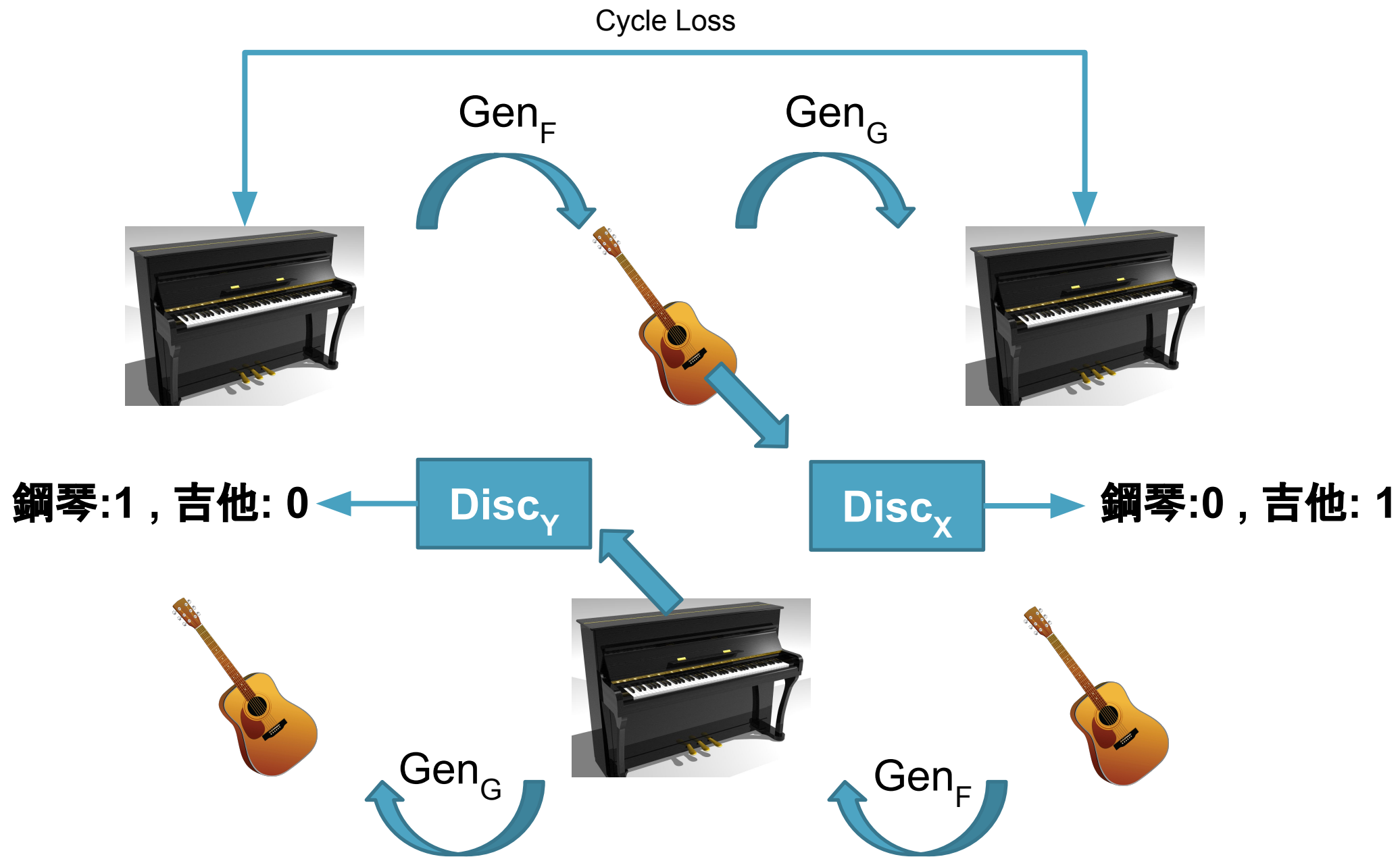
←
Musical Timbre
Transfer
→



資料前處理

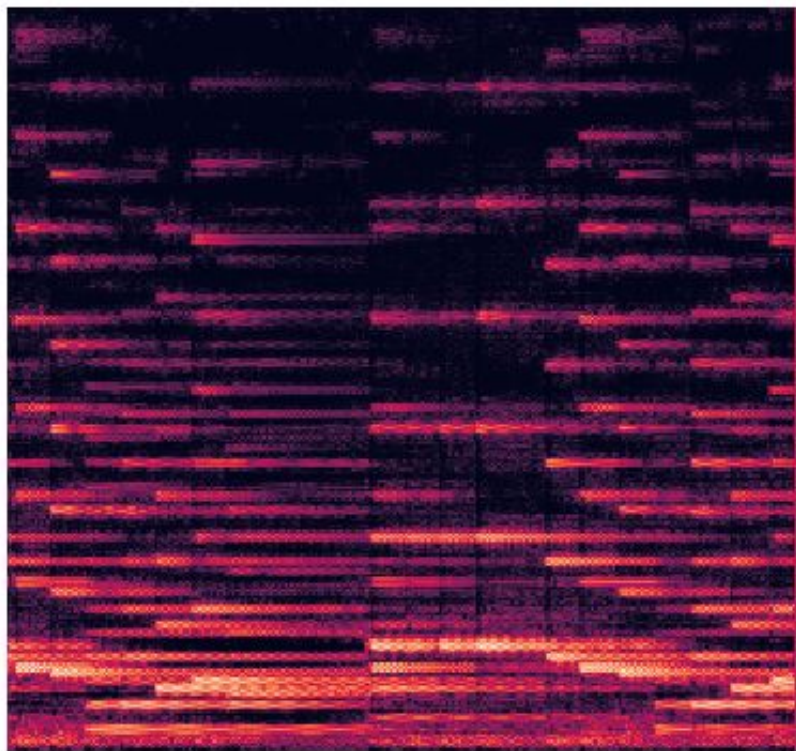
1. 由於資料集中的音樂大都為許多樂曲的集合，尚有些音樂有人聲的部分，樂曲與樂曲間的空白與人聲會影響程式學習，故使用Audacity移除空白與人聲的部分，並將移除的地方平滑化後輸出成wav檔。
2. 分別將處理後的wav檔(piano與guitar各5首約一小時的音訊檔)，以5.9秒的長度切割所有的音訊檔，分別有piano檔案3,887段落，guitar檔案3,275段落

訓練架構

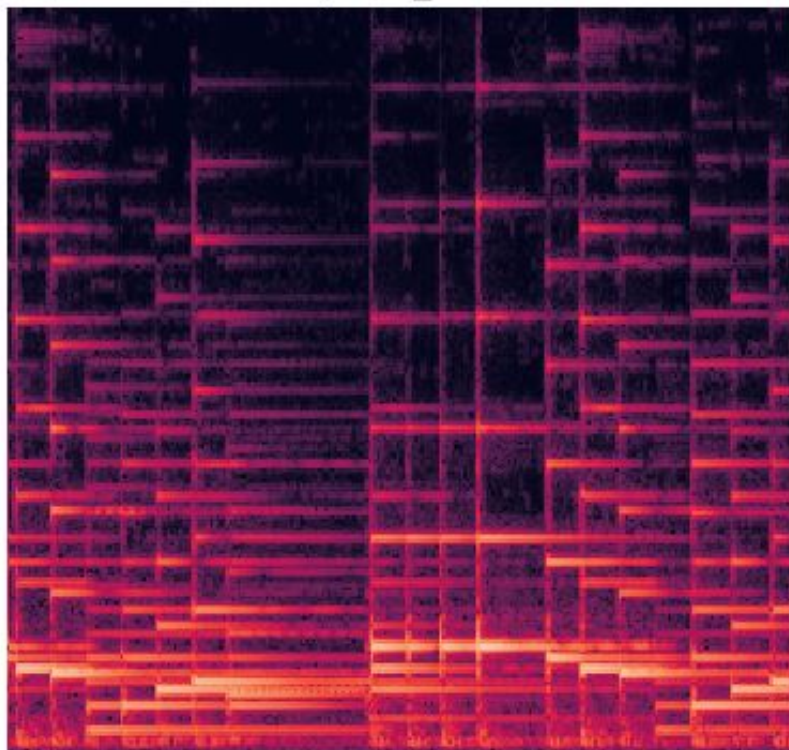


HPSS分解

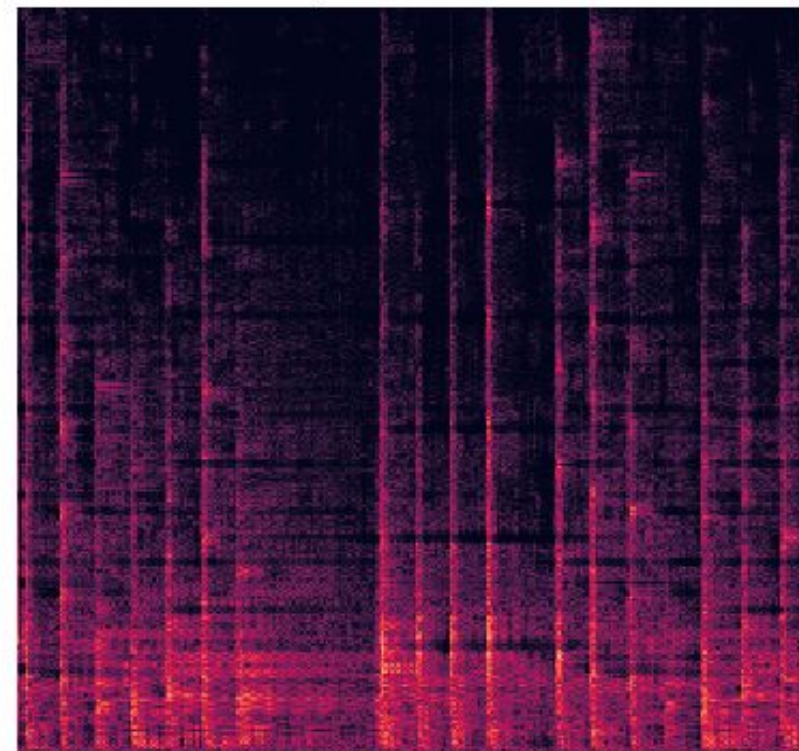
harmonic



piano_stft



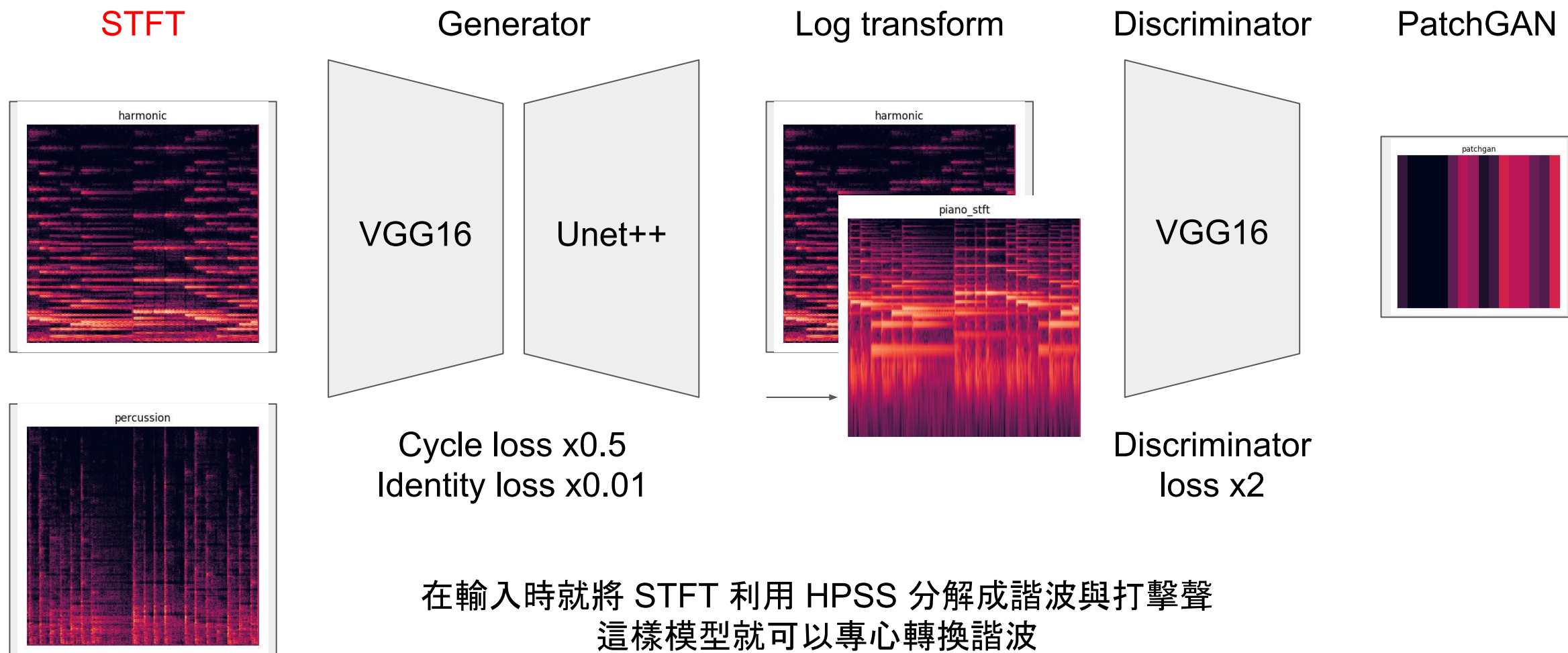
percussion



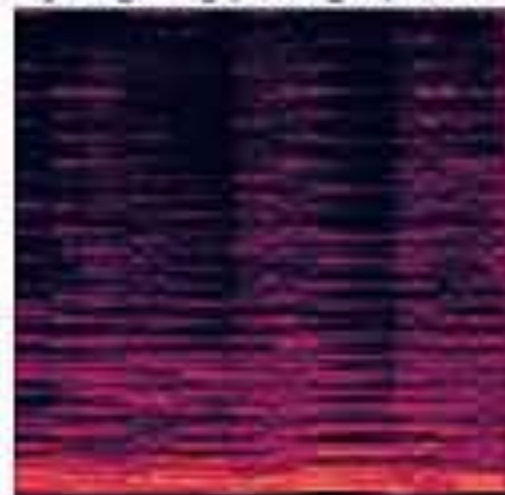
諧波：
根音與其泛音列，是音色構成的
來源，看起來是橫線

打擊聲：
脈衝波，頻率廣但時間短，看起來
是垂直線

STFT 分解成諧波與打擊聲



guitar_harm_epoch06_step02490



piano_harm_epoch06_step02490



disc_fake_x_epoch06_step02490



disc_fake_y_epoch06_step02490



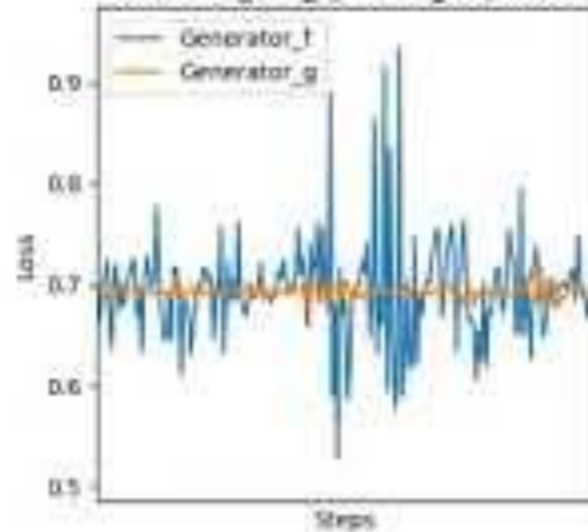
disc_real_x_epoch06_step02490



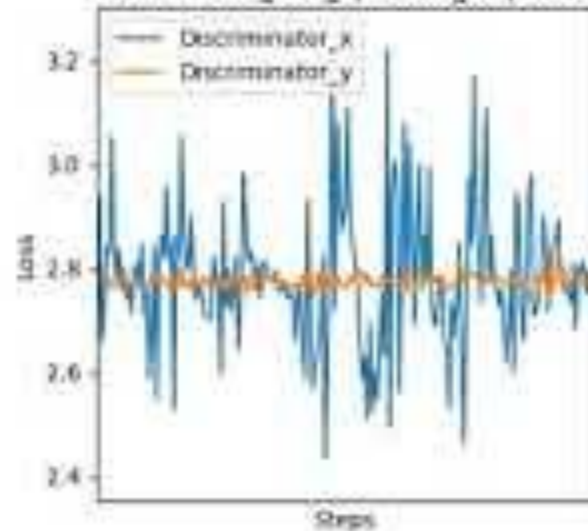
disc_real_y_epoch06_step02490



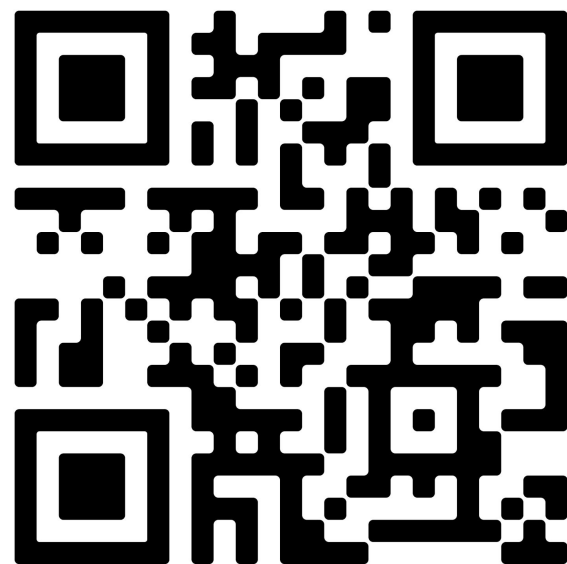
Generator_loss_epoch06_step02490



Discriminator_loss_epoch06_step02490



聽聽看



 Play Now



Thank you