

基於 log-Mel spectrogram 與卷積神經網路的吉他音色辨識系統

指導教授：潘仁義教授

學生：張智捷

摘要

吉他音色深受效果器影響，而以人耳判斷音檔中的效果器往往耗時且主觀。近年深度學習在音訊辨識領域迅速成熟，但多專注在語音辨識、曲風分類等功能，因此本研究使用EGFxSet以3450 筆吉他音檔為資料，使用對數梅爾頻譜(log-Mel)作為特徵，並以卷積神經網路(CNN)進行五類效果器分類。希望建立一套能自動辨識吉他效果器的模型，協助音色分析與音樂製作。

研究方法

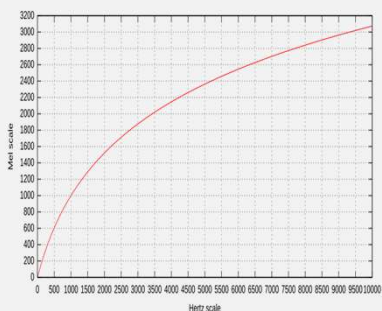


圖一、研究流程圖

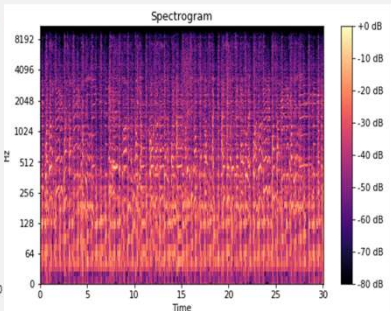
1. 音訊輸入：收集 EGFxSet 中的 5 秒吉他音訊，涵蓋五類效果器：Clean、Digital Delay、Hall Reverb、RAT、TubeScreamer。

2. Log-Mel 特徵擷取：對音訊進行短時傅立葉轉換 (STFT)，將頻率軸映射至 Mel scale 並取對數，形成 log-Mel spectrogram 作為模型輸入。此特徵能有效保留音色紋理與效果器能量分布，是音訊分類的標準方法。

3. CNN 模型分類：將 log-Mel 輸入 CNN 進行五類效果器分類。log-Mel spectrogram 與圖片結構相似，使 CNN 能辨識效果器特徵，如 Delay 週期 脈衝、Reverb 尾音擴散、Distortion 高頻能量集中等音色特徵。

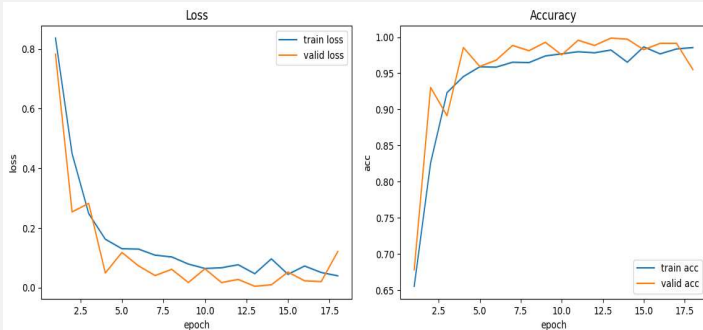


圖二、Mel scale 是一種模擬「人耳如何感受音高」的頻率刻度。



圖三、梅爾對數頻譜

研究結果



圖四、訓練集和驗證集損失

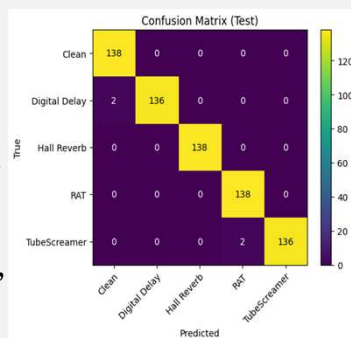
圖五、訓練集和驗證集準確率

模型在訓練過程中快速收斂且保持高度穩定：Loss 在約 10 個 epoch 後趨於平穩、訓練與驗證一致；Accuracy 則穩定維持在 97 - 99%，代表模型無過擬合並具良好泛化能力，整體訓練效果優異。

Class	Precision	Recall	F1-score	Support
Clean	0.9857	1	0.9928	138
Digital Delay	1	0.9855	0.9927	138
Hall Reverb	1	1	1	138
RAT	0.9857	1	0.9928	138
TubeScreamer	1	0.9855	0.9927	138

圖六、各種類效果器分類準確率

混淆矩陣中五類效果器皆落在對角線，僅出現少量誤判，代表模型能清楚區分不同音色特徵。分類報告中各類的 Precision、Recall、F1-score 皆高於 0.98，多數達到 1.0，顯示模型在所有效果器類別上都具有一致且高度可靠的辨識能力。



圖七、各種類效果混淆矩陣

結論

本研究成功建構一套基於 log-Mel + CNN 的吉他效果器辨識系統，能以近乎完美的準確率判斷吉他音訊中的效果器種類。此成果顯示深度學習技術已能有效應用於音色分析，並具備即時性與實用性，可應用於音樂製作、教學與音色管理。未來希望能擴展更多效果器類別，且進一步預測效果器參數。