

# GSCD 之單字 corpus 數量與 accuracy 分析及 SpecAugment 效益

---

Student : 陳憲億

Advisor : Chingwei Yeh and Tay-Jyi Lin

# Outline

Action item：設計實驗證明 SpecAugment 方法是有成效的

- 目前進度

- 使用 GSCD 1000 筆 yes 音檔做訓練，正確率為 94.43% (p.3)
- 使用 GSCD 50 筆，增量到 1000 筆 (不包含原本 50 筆) 做訓練，正確率為 80.72% (p.3)
- 調整 SpecAugment 遮罩參數 (頻率與時間遮罩 1~3組)，產出結果一致性太高，猜測這樣的結果對於訓練結果會不好，但由第二點實驗結果看起來還不錯，推論為還需更多數據來驗證 (p.4、p.5)

- 未來規劃

- 繼續調整 SpecAugment 參數，令其產出結果達到與 GSCD 模型相似的多樣性
- 找出 GSCD 減少訓練模型的正確率曲線，並找出正確率劇烈下跌的點做增量 (示意圖在 p.6)
- 增加 Keyword 驗證是否會影響正確率 (目前只使用一個 keyword：yes)

# 增量的 1000 筆資料(不包含原本 50 筆) 做訓練

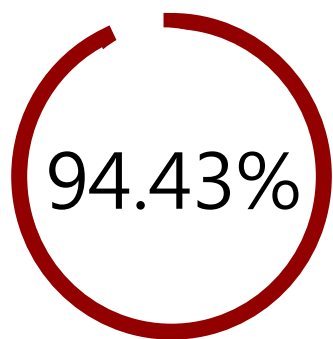
yes 音檔

參數如下

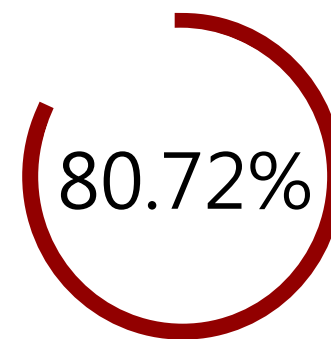
學 習 率 : 0.0005, 0.0001, 0.00002

訓練次數 : 1000, 1000, 1000

關 鍵 字 : yes



GSCD 1000 筆  
(Google Speech Commands Dataset)



GSCD 50 筆  
增量到 1000 筆 (不包含原本 50 筆)

使用 github 資源

<https://github.com/ARM-software/ML-KWS-for-MCU>、<https://github.com/DemisEom/SpecAugment>

使用此專案做訓練與作為對照組

使用此專案將 GSCD 增量

# SpacAugment 三項重要參數

## 時間扭曲

Time warping

時間扭曲參數  $W$

預設為：80

這次訓練使用預設值

## 頻率遮罩

Frequency masking

頻率遮罩參數  $F$

預設為：100

數量預設為：1

這次訓練數量

## 時間遮罩

Time masking

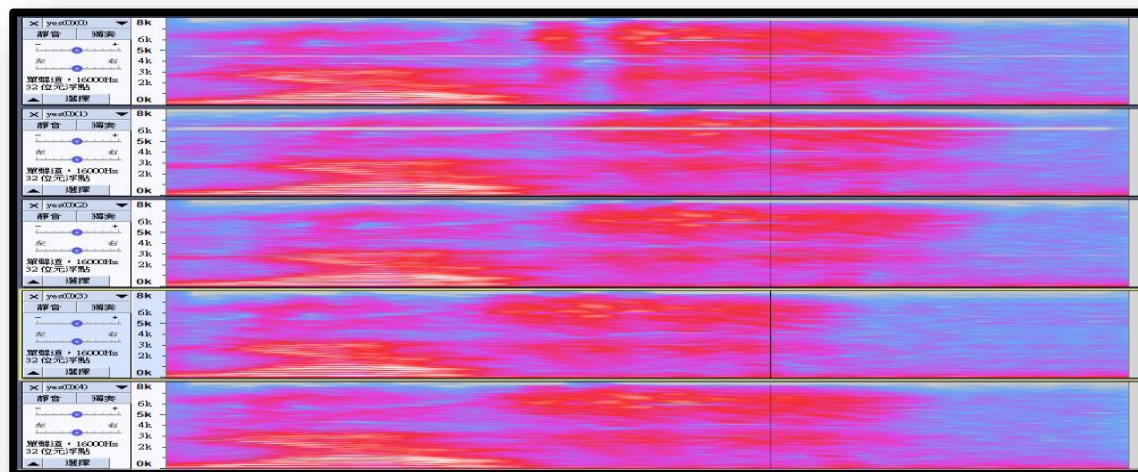
時間遮罩參數  $T$

預設為：27

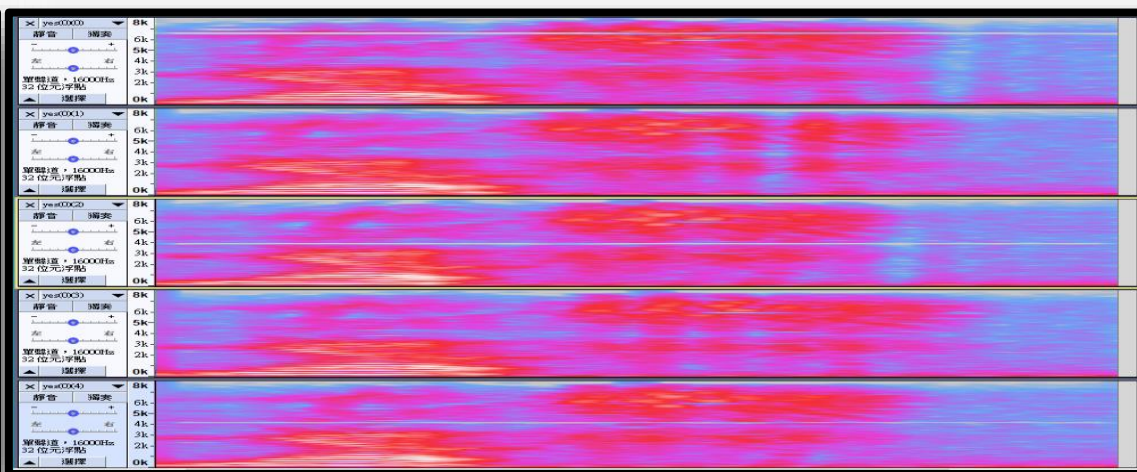
數量預設為：1

# SpacAugment 參數調整與 GSCD 比較

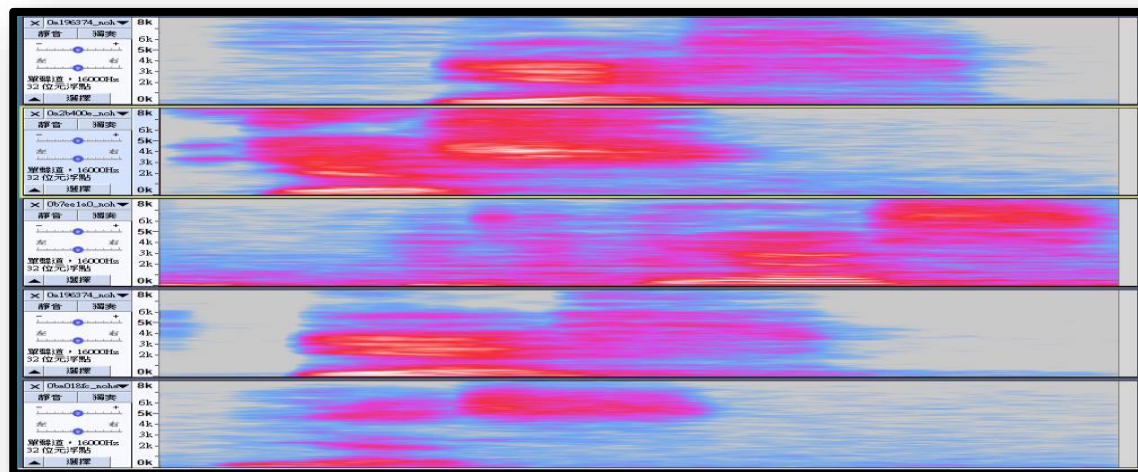
頻率與時間遮罩 1組



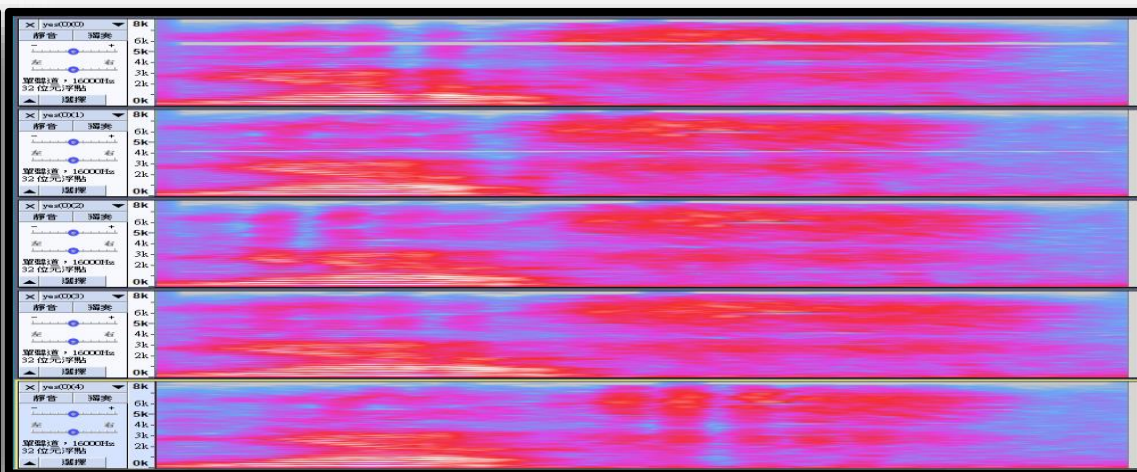
頻率與時間遮罩 2組



GSCD 隨機抽 5 句的頻譜圖



頻率與時間遮罩 3組



# GSCD 減少訓練模型的正确率假想曲線

假設

找出這個下降幅度最誇張的點，代表此資料模型不足，再利用 SpecAugment 做增量，看正確率是否會上升

