

附件四、技術說明表

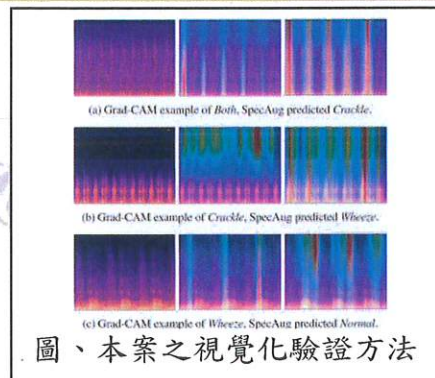


基於頻譜圖的格狀伽瑪校正數據增強

提案人：黃沛銓 臨床助理教授

單位：國立臺灣大學醫學院

簡歷：(可列出相關連結，例如系所、研究室網頁)



市場及需求：

呼吸音分類技術多在辨別正常呼吸音上表現傑出，異常呼吸音的偵測則有待提升，但是異常類別的可靠辨識結果對於精確的診斷才是至關重要。異常呼吸音包含爆裂聲 (Crackle)、喘息聲 (Wheeze) 以及兩者同時出現 (Both)，而兩種聲音有著迥異的特徵：爆裂聲是不連續性、低沉、非音樂性的聲音模式；喘息聲是連續性的高音調序列。為了讓模型兼顧兩種差異很大的異常類別，以往研究多以設計複雜繁重的模型架構為解法，本案則提出一個容易實踐的資料增強技術，得以保留住各項異常呼吸音特徵，並實現有效的數據增強，最終提升異常類別的辨識結果。

技術摘要(含成果)：

本提案技術包含格狀伽馬校正 (Gamma patch-wise correction) 和混合增強 (Mixup)，模型輸入是經過格狀伽馬校正之圖片的兩兩混合。伽馬校正是針對所有像素的指數轉換，相當於重新縮放圖片的對比度。當伽馬設為大於 1 的值時，圖片中偏暗的部分會變得更暗、較亮的區塊則會變得更加明顯，進而產生亮度調整和抑制雜訊的效果，其中圖片中存在的訊息並不會完全消失。「格狀伽馬校正」會在圖片中隨機挑選若干區塊，並對這些小區塊做伽馬校正，而伽馬會被隨機定義為較大的值以達成前述效果。

混合增強將兩張經過格狀伽馬校正的圖片依一定比例混合，並將他們的標記以相同比例相加。混合的圖片會形成新的數據增強資料，而相加的標記代表著混合的圖片屬於各項類別的機率，其中比例是依 Beta 分布隨機取自 0 到 1 的區間。

優勢：

1. 本案的技術保留住不連續性的低頻爆裂聲和連續性的高頻喘息聲在頻譜圖上之呈現，以重新縮放對比度和亮度的方法避免隨機遮罩造成的標記上之混淆，同時達到抑制雜訊的效果。
2. 現有技術並未在異常呼吸音分類達到傑出的表現。本案在維持正常呼吸音判別的分數下，大幅提升了異常類別的成績。
3. 現有技術多針對複雜繁重的模型架構，本案則提出了容易實踐的數據增強方法，能夠套用在各種以頻譜圖為輸入資料的模型。
4. 透過視覺化模型在頻譜圖上的預測，本技術訓練模型對於呼吸音分類的判斷依據，和醫學上對呼吸音性質的定義相吻合。

本資料僅供國立臺灣大學專利/技術申請使用，嚴禁使用全部或部分內容於其他用途。若有疑問請與我們聯繫，我們將盡力協助您。

競爭產品：

此技術目前市面上尚無相關產品應用，但於未來具應用潛力，以下是潛在應用廠商

1. 中華民國

- (1) 聿信醫療器材科技：為解決全球新冠肺炎肆虐之際，對於遠端呼吸監測與呼吸功能評估的需求，聿信醫療持續專注人工智慧連續呼吸監測相關應用技術的精進與擴充。

2. 美國

- (1) 3M 公司：可結合 3M 公司之呼吸監測器具
(2) ResMed 公司：可結合 ResMed 公司之睡眠監測系統

專利現況：

- (1) 本技術有相關類似專利（中華民國專利申請號: TWI818558B，中華人民共和國專利申請號: CN116580702A，CN113241062A），但此前案並未具體揭露「隨機遮擋的態樣」以及「格狀伽馬校正」等技術特徵。

(2)	專利號碼	CN116580702A
	專利名稱	基於人工智慧的語音辨識方法、裝置、電腦設備及介質
	專利權人	平安科技（深圳）有限公司
	技術摘要	<u>本發明通過對梅爾頻譜矩陣進行隨機行遮擋和列遮擋，得到缺失了不同語義資訊的第一語音增強矩陣和第二語音增強矩陣；根據編碼器提取第一語音幀特徵和第二語音幀特徵，計算度量損失對編碼器進行自監督訓練；得到語音融合特徵並輸入至解碼器中得到映射字元，結合預設字元計算預測損失來對語音辨識模型進行有監督訓練，根據零字元和非零字元的數量對預測損失和度量損失加權相加，得到目標損失來訓練編碼器和解碼器，將自監督和有監督的訓練方式相結合，提高了語音辨識模型的識別準確率，極大地提高了醫療技術領域中資訊錄入的即時性、便利性和準確性。</u> 具體地，在本實施例中，<u>對各個梅爾頻譜矩陣的行進行隨機遮擋，得到第一語音增強矩陣，對各個梅爾頻譜矩陣的列進行隨機遮擋，得到第二語音增強矩陣，則與梅爾頻譜矩陣相比，對梅爾頻譜矩陣進行了部分遮擋後得到的第一語音增強矩陣和第二語音增強矩陣缺失了部分語義資訊，且第一語音增強矩陣和第二語音增強矩陣缺失的語義資訊不同。</u> 因此，可以將缺失了不同語義資訊的第一語音增強矩陣和第二語音增強矩陣作為對語音辨識模型進行自監督訓練的基礎，同時，通過對梅爾頻譜矩陣進行隨機行遮擋和列遮擋，可以得到大量的自監督訓練樣本。
	比對分析	此前案提到「將音檔轉為梅爾頻譜圖，且進行隨機遮擋」的技術特徵，然而，此前案並未具體揭露所謂隨機遮擋的態樣以及格狀伽馬校正等技術特徵。

2	專利號碼	CN113241062A
	專利名稱	語音訓練資料集的增強方法、裝置、設備及存儲介質
	專利權人	平安科技（深圳）有限公司
	技術摘要	本發明提供了一種語音訓練資料集的增強方法、裝置、設備及存儲介質，其中，方法包括：<u>通過提取各個語音訓練資料對應的梅爾頻譜圖並進行圖元點重排處理，得到圖元點重排處理後的暫時梅爾頻譜圖，為每個所述暫時梅爾頻譜設定擦除區域面積，設定擦除區域的形狀參數，更</u>

本資料僅供國立臺灣大學專利/技術申請使用，嚴禁使用全部或部分內容於其他用途。若有疑問請與我們聯繫，我們將盡力協助您。

		改參數或者隨機擦除係數，得到多個延伸梅爾頻譜圖，將各所述延伸梅爾頻譜圖轉化為對應的目標語音訓練資料，從而完成對語音訓練資料的補充。發明的有益效果：解決了語音訓練資料較少，導致語音模型在訓練過程中容易出現過擬合的問題，並且增加了語音模型的魯棒性，避免語音模型陷入過擬合，大大提高了語音模型的應用範圍。
	比對分析	此前案提到「將音檔轉為梅爾頻譜圖，且進行遮擋」、「對頻譜圖進行影像處理」的技術特徵，然而，此前案並未具體揭露所謂隨機遮擋的態樣以及格狀伽馬校正等技術特徵。

3	專利號碼	TWI818558B
	專利名稱	用於病理語音識別之系統及方法以及電腦可讀儲存介質
	專利權人	國立陽明交通大學
	技術摘要	本揭露提供一種用於病理語音識別之系統及方法以及電腦可讀儲存介質。所述用於病理語音識別之方法係包括：採集語音信號； <u>利用梅爾頻率倒頻譜係數(Mel Frequency Cepstral Coefficients,MFCC)演算法處理該語音信號以獲得 MFCC 頻譜圖</u> ；自該 MFCC 頻譜圖提取多個特徵；以及藉由深度學習模型根據該語音信號的該 MFCC 頻譜圖的該數個特徵， <u>預測該語音信號的病理狀態</u> ，其中，語音信號的病理狀態包含正常語音、單側聲帶麻痺、內收肌痙攣性發聲障礙、聲帶萎縮、及器質性聲帶病變。
	比對分析	此前案提到「將音檔轉為梅爾頻譜圖」的技術特徵，但其並未揭露本提案格狀遮擋及格狀伽馬校正等技術特徵。

聯絡方式(請不用填):

臺大產學合作總中心

Tel: 02-3366-9945, E-mail: ordiac@ntu.edu.tw