

全卷積神經網路於直接評估指標優化的端到端語音波形增強

摘要

本發明是一種全卷積神經網路於端到端的語音波形增強模型。此模型可直接優化任意的評估指標，包括客觀理解度或品質分數。當考慮理解度分數時，自動語音識別（ASR）系統和人類受試者在噪聲條件下的字錯誤率（WER）皆可以大大降低。

技術優勢

我們的方法有很多優點列舉如下：

1. 不需要將時域波形映射到頻域以增強振幅頻譜。因此，可以避免所有相關的預處理和後處理。
2. 由於所提出的模型直接對噪聲波形進行去噪，相位信息不會被忽略。
3. 在傳統的基於幀的處理模型中觀察到的增強語音的不連續性藉由將每個語句作為一個整體來解決。
4. 語音增強的真實目標與所採用的目標函數之間的不匹配可以通過我們基於整個語句的波形優化來解決。
5. 與其他基於深度學習的模型相比，參數的數量可以大大減少。

本院覽號

32T-1070119

公告日期

應用範圍

該模型可以用於需要去噪演算法的任何設備，例如，手機，耳機和助聽器等。或是在某些對於語音理解度的提高是很重要的情況。

創作人

傅思維、曹昱