

基於變分自編碼器的語音轉換技術

本院覽號

公告日期

智財權狀態

05T-1100925

2021-01-10

know-how

摘要

語音轉換旨在不改變語言內容的情況下將語音從來源型式轉換為目標型式。最典型的任務是說話人語音轉換。大多數傳統方法需要平行訓練數據，即來源和目標說話者唸相同的轉錄本進行訓練。由於此類數據難以收集，非平行訓練長期以來一直是語音轉換的最終目標之一。透過變分自編碼器的自我督導訓練，可以實現非平行訓練。整合生成式對抗型網路和跨域特徵可以進一步提升輸出音質。

技術優勢

- 使用變分自編碼器可利用大量語音對模型進行預訓練
- 整合生成式對抗型網路可以提升輸出音質
- 整合跨域特徵可以提升輸出音質

應用範圍

- 一般性的人聲轉換
- 受損語音轉換成正常語音
- 角色扮演遊戲中的人聲轉換

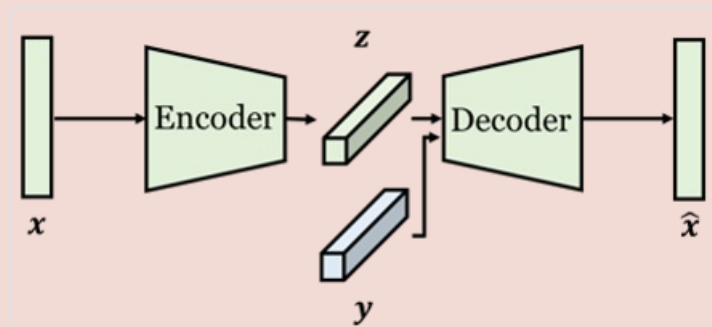


圖1.基於變分自編碼器的語音轉換模型， x 為輸入特徵， \hat{x} 為輸出特徵， z 是潛在代碼， y 是說話者代碼。

創作人

王新民、曹昱



中央研究院
ACADEMIA SINICA