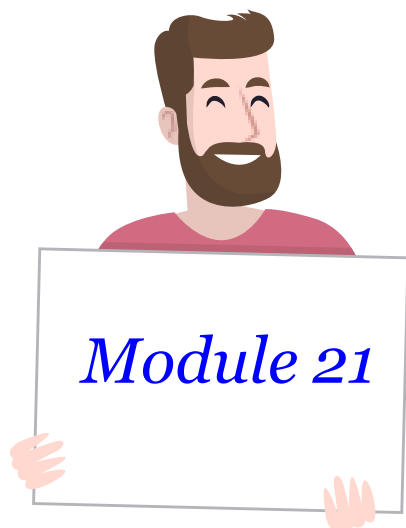




RNN進階應用



designed by  freepik

Estimated time:
45 min.

學習目標

- 21-1: 機器翻譯
- 21-2: 圖像生成描述
- 21-3: 語音辨識



21-1: 機器翻譯

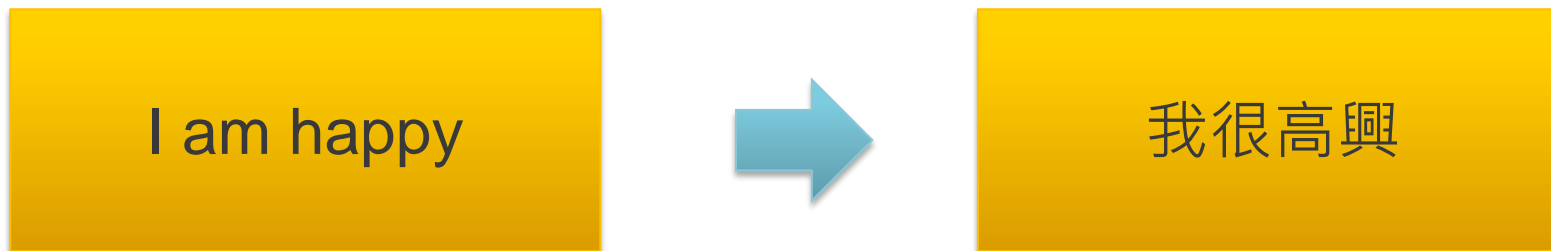
- 機器翻譯介紹
- 機器翻譯原理



designed by freepik

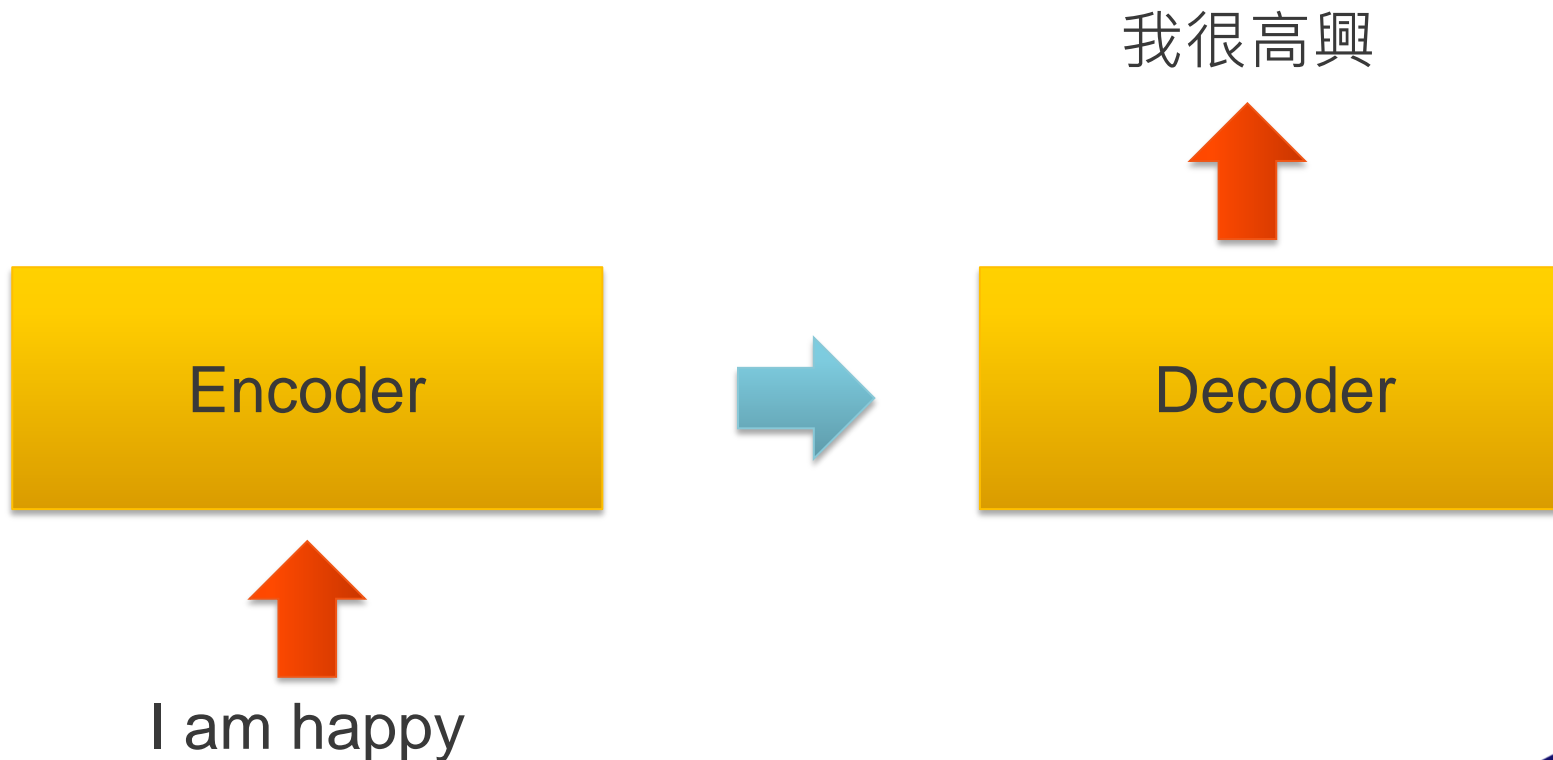
機器翻譯介紹

- 機器翻譯主要是將某一種語言的句子翻譯成另一種語言
 - 有時候輸入的句子長度會跟輸出的不一定一樣
 - 例如”I am happy”有三個字，”我很高興”有四個字



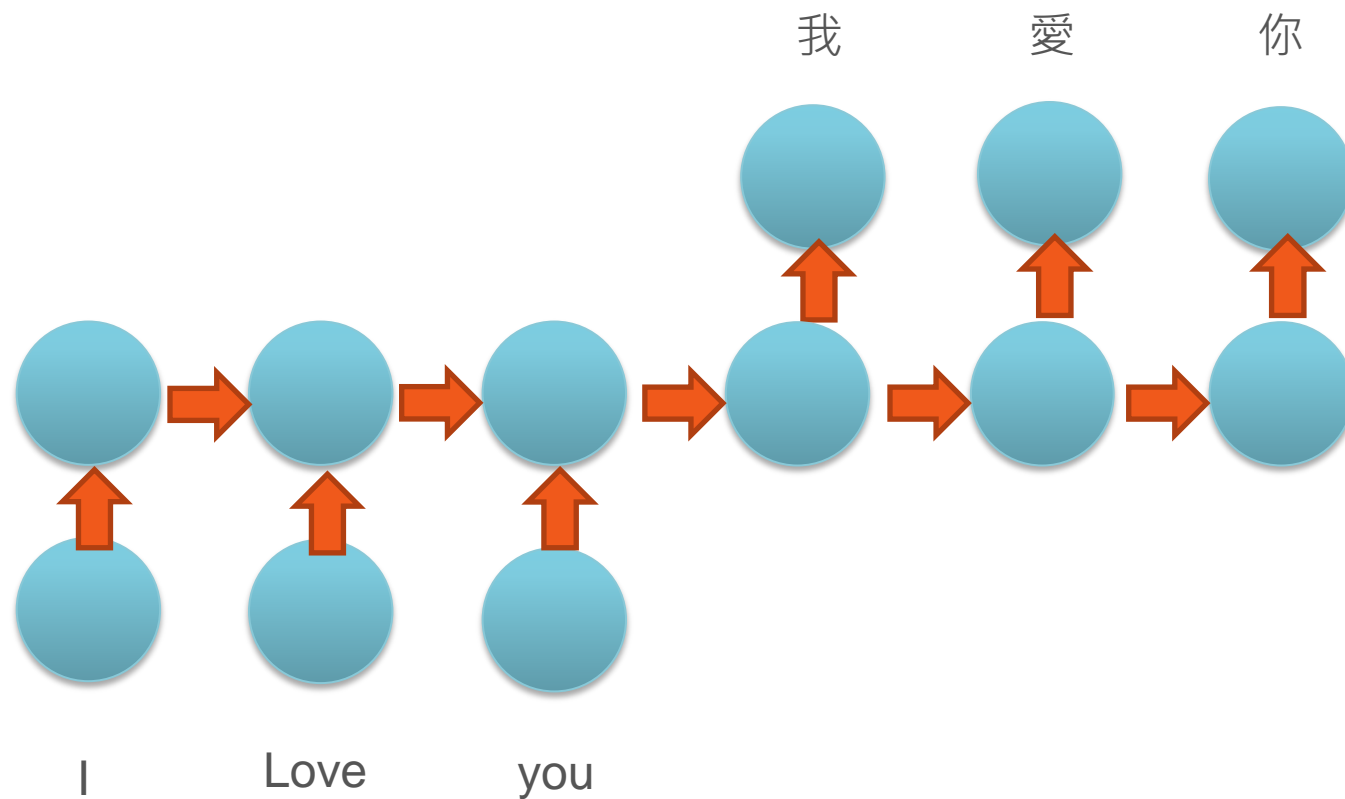
機器翻譯原理

- 機器翻譯常見的做法是定義一個encoder以及decoder
 - encoder的功用是將原始語言的句子做特徵抽取
 - decoder的功用是將得來原始語言句子的特徵解碼成另一個語言



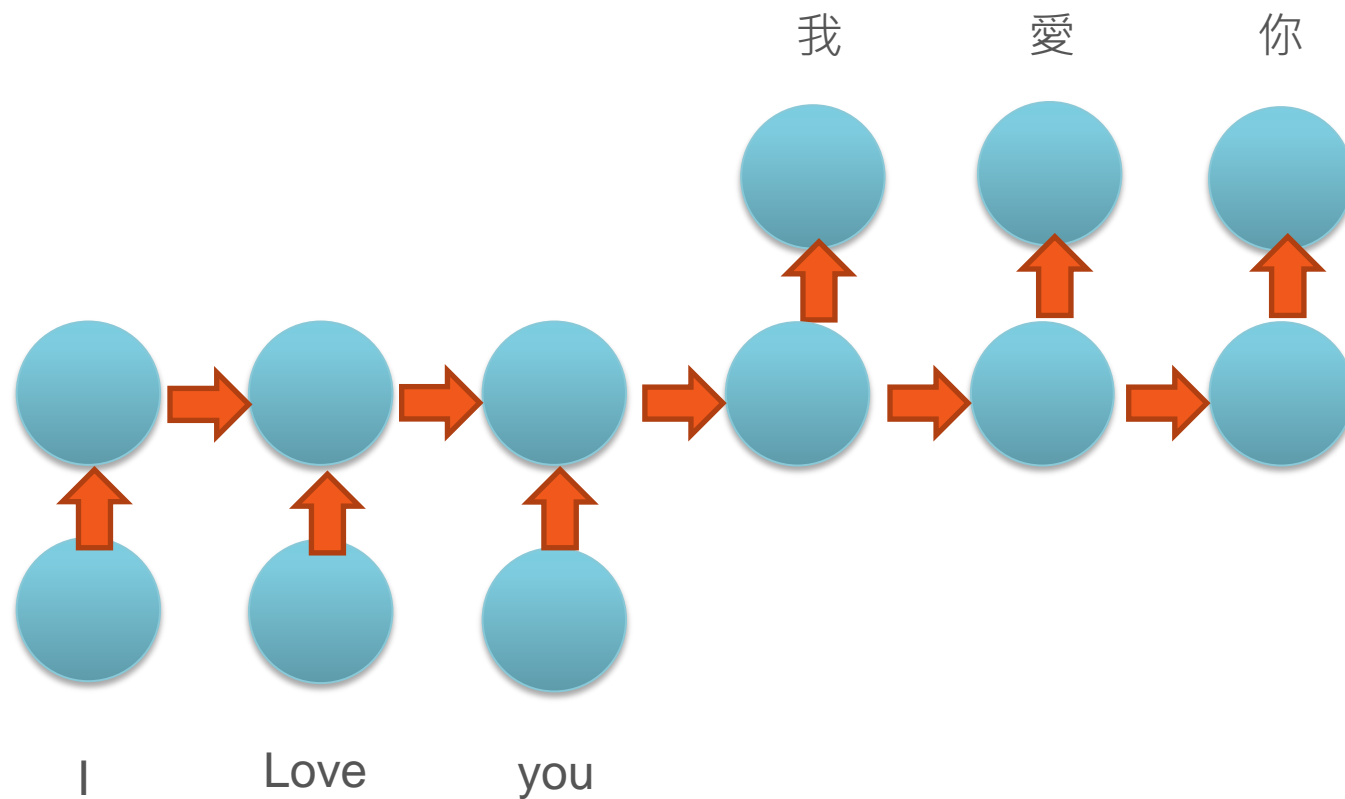
機器翻譯原理

- 我們常常使用RNN來當作encoder以及decoder，示意圖如下



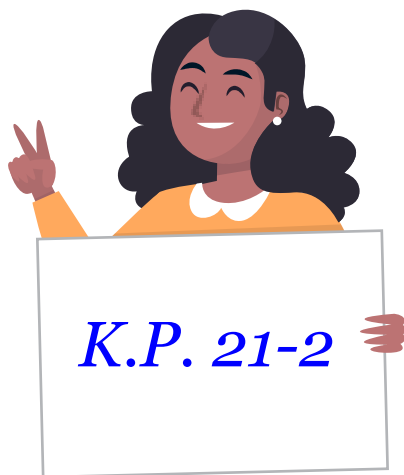
機器翻譯原理

- 此類的模型又叫做**Sequence-to-Sequence**模型
 - 有興趣的同學可以再去深入研究



21-2: 圖像生成描述

- 圖像生成描述介紹
- 圖像生成描述原理



designed by freepik

圖像生成描述介紹

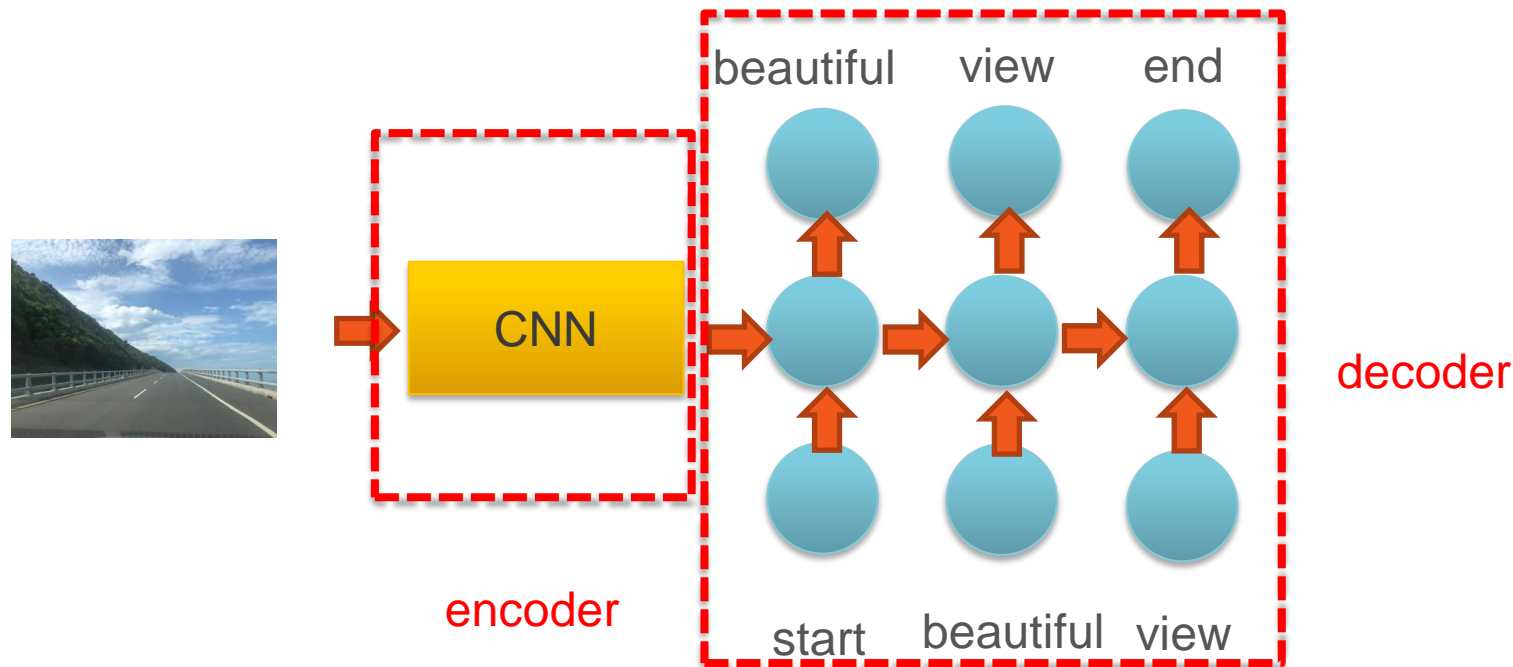
- 圖像生成描述是一種深度學習常見的應用
 - 給予電腦一張照片，電腦會基於這張照片產生一段描述



A man is sitting on the ground

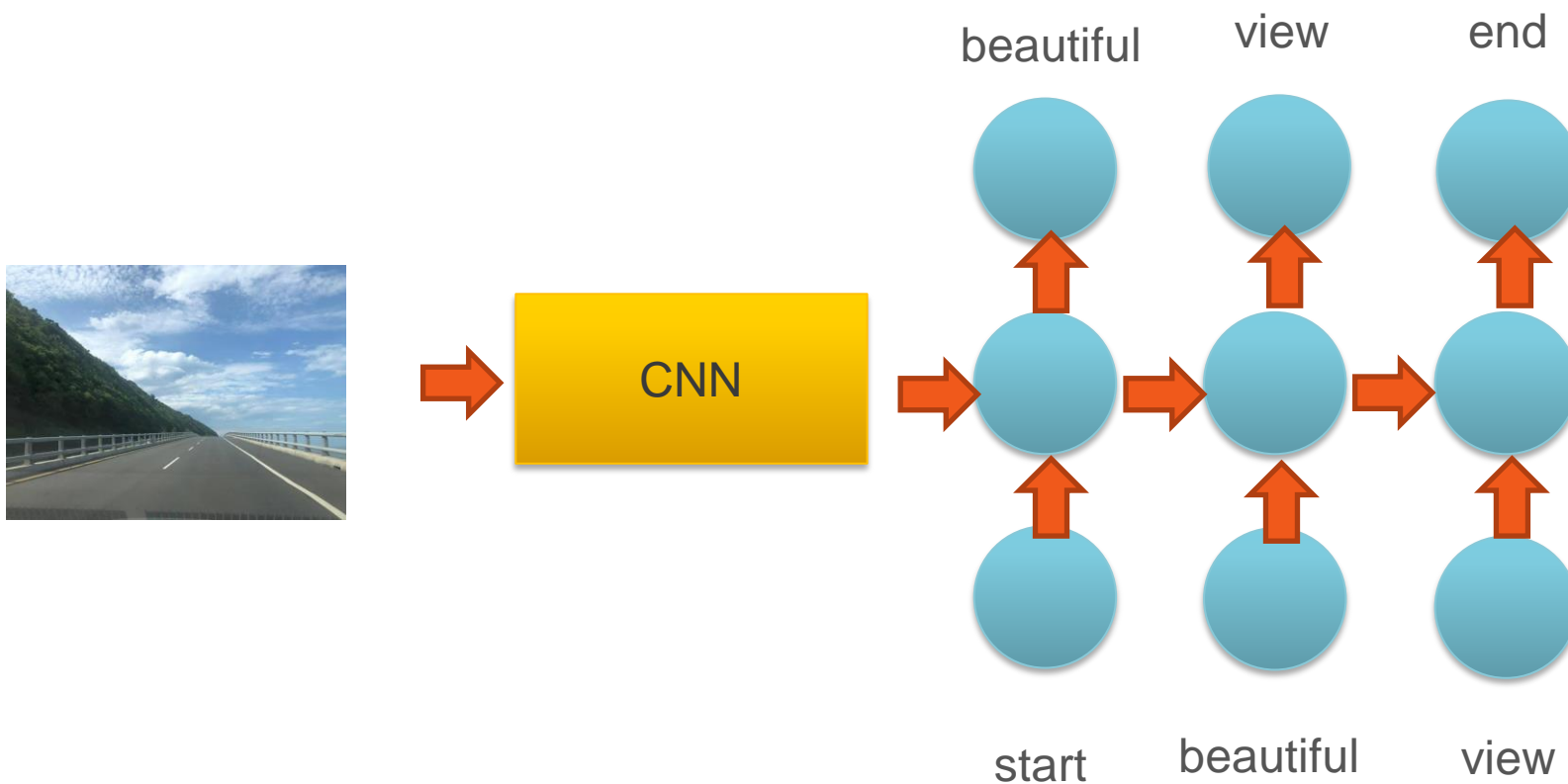
圖像生成描述原理

- 一般來說，圖像生成描述的網路通常包含encoder以及decoder兩個部分
 - Encoder專門把照片做特徵抽取
 - Decoder專門把照片的特徵轉換成描述



圖像生成描述原理

- **CNN**在圖像特徵抽取上比較厲害，所以常被拿來當**encoder**
- **RNN**在生成文字序列比較厲害，所以常被拿來當**decoder**

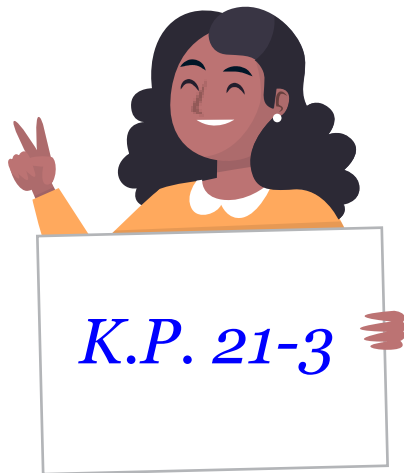


圖像生成描述原理

- 在訓練此網路的時候，訓練集資料為每一張照片，而每一張照片對應的標籤為一句話
 - 有興趣的同學可以更深入去了解圖像生成的原理

21-3: 語音辨識

- 語音辨識介紹
- 語音辨識原理



designed by freepik

語音辨識介紹

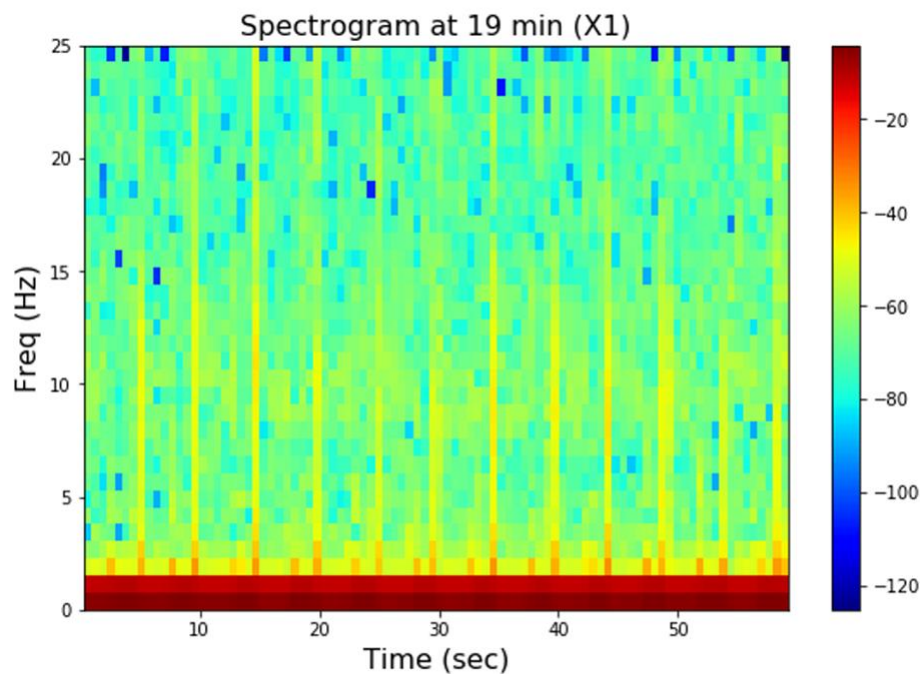
- 語音辨識是人工智慧裡常見的應用
 - 其目的是將給予電腦音檔，電腦能辨識出音檔裡面在說甚麼
 - 有人叫它ASR(Automatic Speech Recognition)、STT(Speech To Text)
 - 是一個滿難的問題的

語音辨識介紹

- 需要牽扯到非常多的學問
 - 訊號處理、語言模型、聲學模型等
 - 由於牽扯到聲音，所以麥克風的好壞也會影響辨識品質

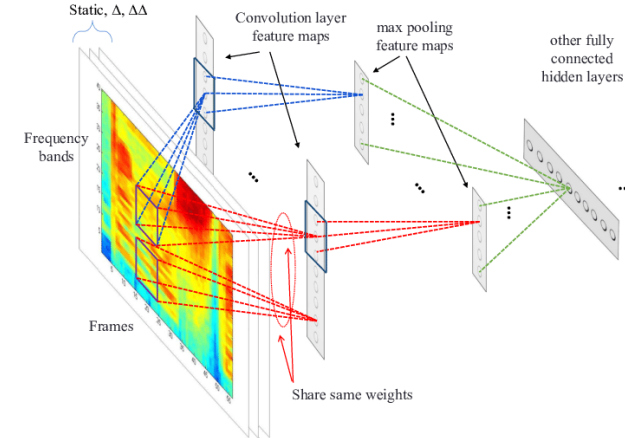
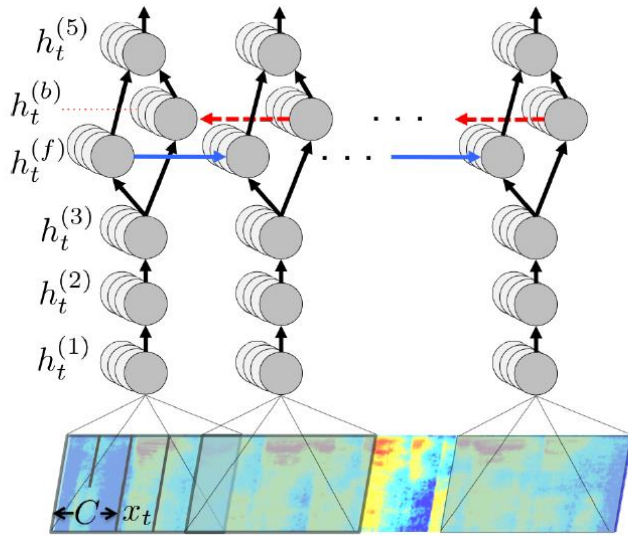
語音辨識原理

- 語音辨識的作法有非常多，以下提供幾種作法供大家參考
 - 首先先將一個聲音轉換成頻譜圖，並把頻譜圖視為一張照片



語音辨識原理

- 將頻譜圖丟到一個雙向RNN的網路裡面去預測文字序列，如左下圖
- 也有人將頻譜圖視為一張照片，丟到CNN網路裡面去預測文字序列，如右下圖
- 語音辨識的原理很複雜，如果各位同學有興趣可以再去深度了解



Source: <https://arxiv.org/pdf/1412.5567.pdf>

Source: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/CNN_ASLPTrans2-14.pdf

線上Corelab

- 題目1：設定取樣率為 16000，錄音 5 秒鐘，說「我要預約」
- 題目2：讀取音檔 record.wav，繪製波型並播放音檔
- 題目3：讀取音檔 record.wav，將音檔變大聲，並顯示、播放，將音檔變小聲，並顯示、播放

本章重點精華回顧

- 機器翻譯
- 圖像生成描述
- 語音辨識



Lab: PyAudio 使用

- **Lab01: 安裝 PyAudio 和 Matplotlib**
- **Lab02: 外接麥克風錄音**
- **Lab03: 顯示、播放、後處理聲音**

Estimated time:

20 minutes

