國立清華大學資訊工程系 110 學年度 2 學期專題報告

專題名稱	擴展視訊會議功能之機器學習模型研究與實作				
參加競賽或計畫	☐ 參加對外競賽	☐ 參與其他計畫	■ 無參加對外競賽或任何計畫		
學號	107062214	107062313	107062314		
姓名	陳伯瑾	黃寶萱	陳柏均		

摘要

2019 年年底，新型冠狀病毒爆發後快速的蔓延至世界各國，企業、學校等紛紛實施在家工作、遠距教學，線上通訊顯然已成為疫情下的趨勢，人們對視訊會議軟體的使用需求快速上升。本研究的目標在於提升視訊會議品質，以現有的視訊會議軟體為核心概念，利用機器學習模型，希望可以對影音信息做優化處理。

本研究所涵蓋的面向可分為四大部分：Image Instance Segmentation、Super Resolution、Speech to Text、Speech Denoising。首先 Image Instance Segmentation 部分，利用 PointRend model 實作，讓使用者在會議過程中可以即時更換會議背景，提升使用者的隱私保證性；接著 Super Resolution 部分，利用 Real-ESRGAN model 提高會議錄影檔中模糊影像的解析度，使得會後回顧會議內容時，可以有更清晰的影像畫面；至於 Speech to Text 部分，以 Google 的 API 進行實作，同時與一個由中國開發之 Chinese ASRT(Auto Speech Recognition Tool)進行一定的比較，該 ASRT 主要將語音轉成 spectrogram 再以 VGG 的方式進行訓練；最後關於 Speech Denoising 部分，以 U-net 為 Model，針對會議記錄的音訊做降噪處理，讓使用者在會後回顧時，可以更清楚的識別會議中的談話內容。

中華民國 110 年 11 月

目錄

壹. 專題研究動機與目的

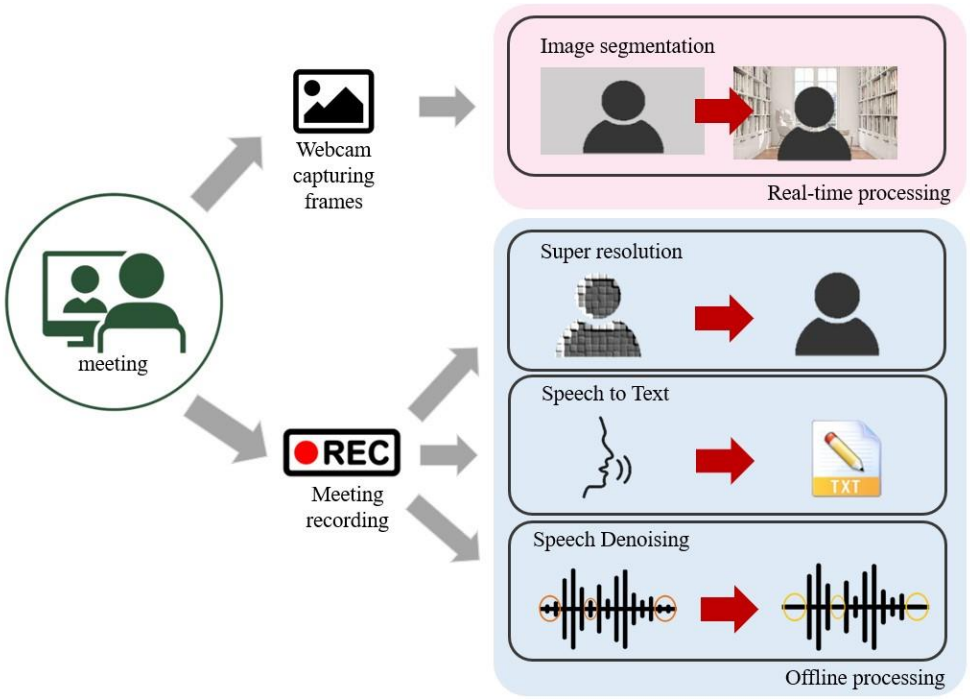
近兩年受到 COVID-19 疫情的影響，全球各產業皆受到嚴重的衝擊，減少外出改變了人與人之間的互動模式，WFH(Work from Home)更是疫情下的常態，隨著對遠距通訊的使用需求提高，提升視訊會議系統的品質與功能已成為現在的趨勢。

本研究以現有視訊會議軟體的架構為設計主軸，目標是利用機器學習模型與深度學習模型對影音進行不同的處理：更換會議背景(instance segmentation)、提高影像解析度(super resolution)、會議內容文字記錄(speech to text)、去除影片背景雜音(speech denoising)，並將我們的線上會議系統以網站的方式呈現，提供更完善的視訊會議品質。

貳. 研究方法與步驟

廣泛閱讀 Instance Segmentation、Super Resolution、Speech to Text 以及 Speech Denoising 相關論文文獻， 了解 State-of-the-Art models 的架構與設計原理，並實際應用這些模型架構，利用網路上提供的 pre-trained weights 判斷該模型是否達到我們期望的目標結果，若不符合預期，則利用符合我們需求的 dataset 重新 training 或是微調模型參數設定，最終達到我們要的功能呈現。實現機器模型應用的同時，我們也同步逐次建立起通訊系統的 prototype：建立單一 user 與 server 之間的影像傳送、加入第二位 user、兩個 user 之間相互連線達到影音傳送的視訊通話、網頁 UI 功能與結合不同的機器學習模型功能，最終對網頁進行排版美化。

參. 設計原理



(圖 1) 系統架構圖

在我們的會議系統中，我們主要實作了 4 項功能:

1. Image Instance Segmentation : 實作 real-time 即時影像分割處理，以利後續做背景模糊或是換背景的功能。
2. Super Resolution: 提高影像解析度，使在會議系統中所記錄下來的影像能夠更清晰，讓使用者體驗更好。
3. Speech to Text: 將錄好之會議影片的聲音部分提取出來，再進行中文 Speech to text 的任務。
4. Speech Denoising: 針對提取出來的會議音訊進行降噪處理。

本會議系統，資訊傳遞將分為 Server side 以及 Client side 進行簡單說明。

一、**Server side**

可再分為 Real-time image processing 和 off-line processing 兩部分。

(一)Real-time image processing：

透過 Http Request 將 webcam 拍攝到的即時影像傳回 server，並利用 PointRend module 中的模型對影像作 instance segmentation 達到人像與背景的切割，經過後續處理後，實作背景模糊、更換背景等功能，同時也會計算影像亮度值，自動調整影像亮度，再將處理完的影像傳回 client side，顯示在網頁上。

(二)Off-line processing：

主要應用在會議錄製的影音檔後製處理，包含 Super Resolution 提高影片解析度、Speech to Text 自動產生會議對話內容的文字檔和 Speech Denoising 降低影片背景雜音三個項目。

整理流程上，會先將會議錄製的影音檔中的音訊部分提取出來並存在資料夾中，供 Speech to text 和 Speech Denoising 使用，而餘下的影像則會做 super resolution 處理，最後再將 Denoising audio 和經過 super resolution 的影像做結合。

1. Super Resolution

   為了更真實的模擬 low resolution 的影像，因此先將原本的錄影檔透過 Gaussian Blur function 降低影片的解析度，再利用 Real-ESRGAN model 針對模糊後的影片做 super resolution 的處理，達到提高解析度的功能。

2. Speech to Text

   獲得音訊檔案後，會對音訊檔進行分割，而分割的方法是用 silence 的間隔時間，將分割後的音檔放到一個叫做 splitAudio 的資料夾。分割音檔的目的在於：目前我們在會議系統中所使用的是 Google 所提供的 API，而上述進行分割的做法不但能解決 Google 無法免費處理大量語音資料的問題，同時也能讓使用者在看 speech to text 結果的時候會比較清楚。

   而產生的 Text 會輸出到一個 txt 檔中保存下來，並在使用者於網頁上看資料時，將其呈現出來。

3. Speech Denoising

   獲得音訊檔案後，藉由將該音訊檔傳入進行 Denoising 的 Model，完成降噪的功能。Denoising Model 會產生的經過降噪處理的音訊檔，而該檔案會再結合加入完成 super resolution 的影像中。

二、Client side

可分為 user 之間通訊連線、網頁使用者介面設計兩部分。

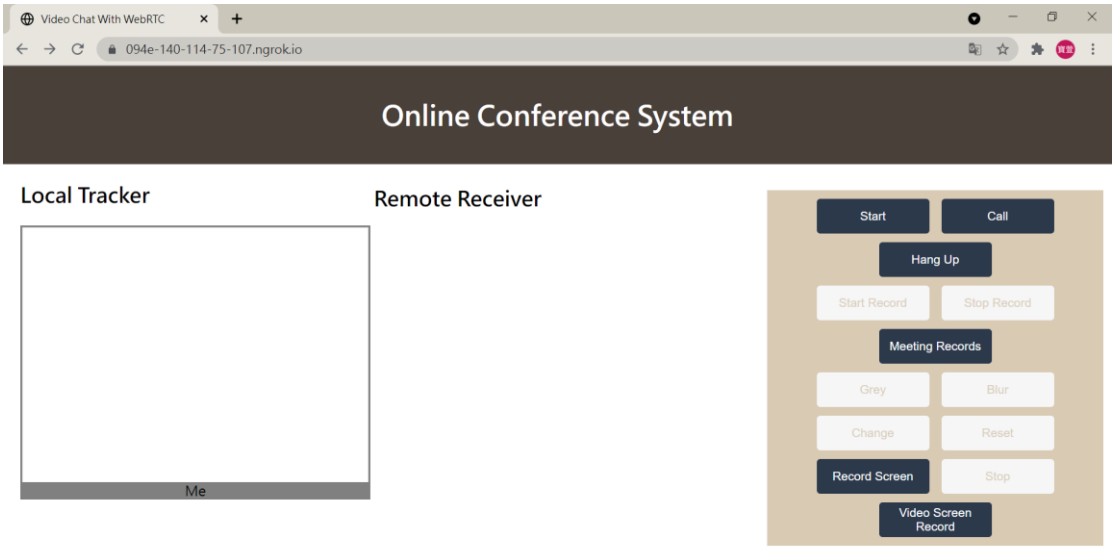
(一)User 通訊連線：

利用 WebRTC 架設兩個 user 之間的影像與音訊傳送，達到即時視訊的功能，並個別連結每個 user 和 server 之間的資料傳送。

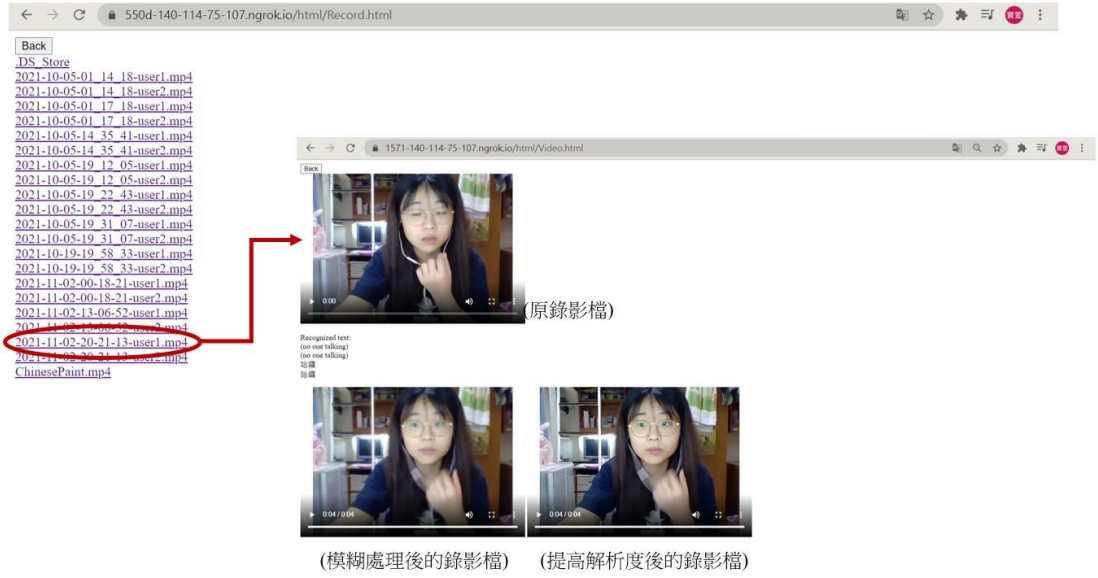
(二)網頁使用者介面設計：

設計使用者與系統互動的網頁平台，如下方圖 2，包括開始視訊會議、會議錄製、更換背景等功能。另一方面，提供會議錄製功

能，並會將錄製的影音資料經過 backend offline processing 後，呈現在另一網頁上，讓使用者可以點擊回顧過去的會議內容(見圖 3)。



(圖 2) 視訊會議主畫面



(圖 3) 會議紀錄畫面

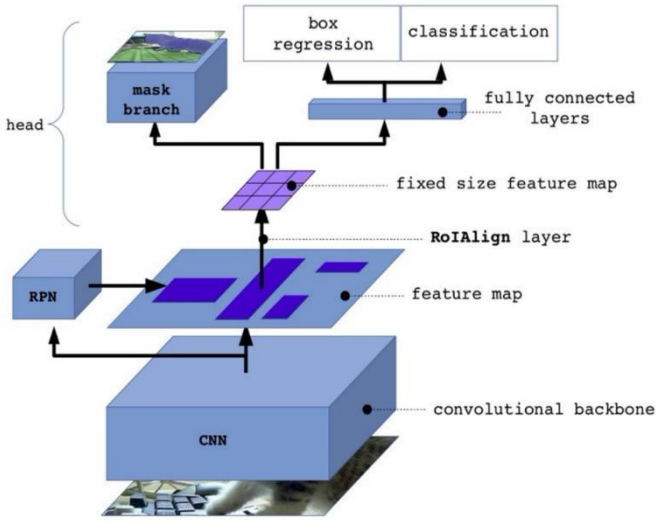
肆. 系統實現與實驗

以下主要分為 Image Instance Segmentation、Super Resolution、Speech to Text 及 Speech Denoising 四大面向進行介紹。

一、Image Instance Segmentation

本實驗最初採用 Mask R-CNN (Mask Region-based CNN)模型實作，同時可以做到 object detection 與 instance segmentation。Mask R-CNN model 的架構可以分成 5 大部分，如下圖 4：

- convolutional backbone：提取 input image 的 feature map
- RPN(Region Proposal Network)：框出所有大小不同的 RoI(Region of Interest)
- RoIAlign：調整每個 RoI 的大小
- Mask branch：預測每個 object 的 mask
- Bounding-box Recognition branch：判斷每個 object 的 class 與 bounding box。



(圖 4) Mask R-CNN model

根據 model output (如附圖 5)的結果，保留每個 object 分割出 mask 的部分並做後續背景模糊的處理，可以得到如下方圖 6 的結果。

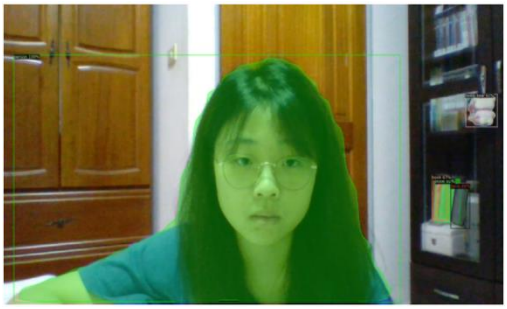
(圖 5) Mask R-CNN inference

(圖 6) The result of processing the background blur according to the mask

然而，實際應用在視訊會議人像背景模糊處理時，得到的效果卻不盡理想，受限於 mask 切割物體邊緣的精準度，如下圖 7，而會議畫面又以人像上半身為影像主體，需要更準確分割出人像與背景，因此後來嘗試 re-train Mask R-CNN model，設計以上半身人像為影像主體的照片為 dataset，並指針對 mask branch 做 training，最終得到的結果卻仍未有效改善 instance segmentation 的效能。

(圖 7) The performance of instance segmentation is not accurate enough

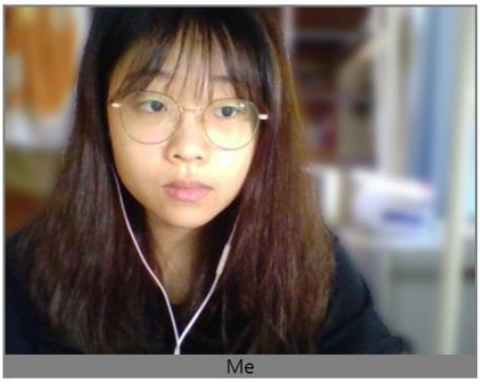
而後續，在持續增加 data 進行 training 的過程中，也同時查找相關論文，尋求他解，最後找到由 Mask R-CNN 延伸出的 instance segmentation model, PointRend (Point-based Rendering) neural network module，保留 Mask R-CNN 的架構，並優化 instance segmentation(mask branch)的效能，在每次 iteration 由小影像逐步恢復到原影像大小的過程中，PointRend module 會針對不確定的 points (通常為人像與背景的交界處)重新預測那些 points 屬於哪個 class，因而能得到更精準的分割，如下方圖 8、圖 9 的比較。



(圖 8) Mask R-CNN

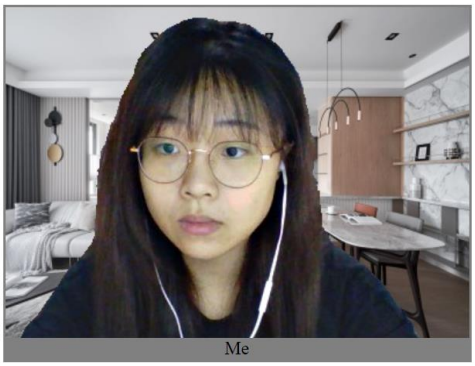
(圖 9) PointRend inference result

因此最後以 PointRend 實作 real-time 即時處理影像，呈現的結果如圖 10、圖 11，並在後續針對人像與背景的交界處做模糊的處理，讓前景與背景能更自然的合在一起，如下圖 12、13 的比較。

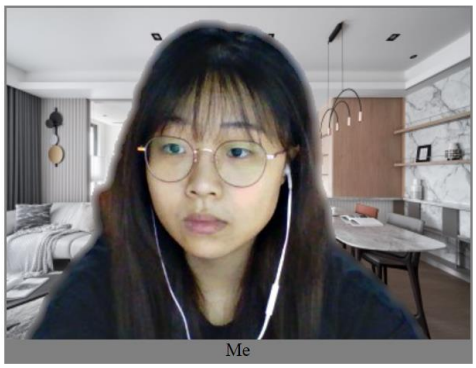


(圖 10) The result of background blurred

(圖 11) The result of changing the background



(圖 12) Without edge blurring



(圖 13) Do edge blurring process

二、Super Resolution

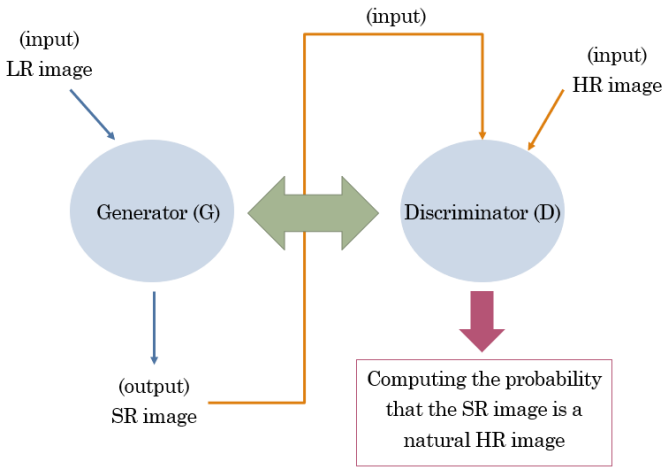
本實驗利用 Real-ESRGAN 模型實作提高影像解析度的功能，Real-ESRGAN 全名為 Real Enhanced Super-Resolution Generative Adversarial Networks，由 ESRGAN model 經過修改後所延伸出的模型，能更好的模擬真實世界中 low resolution 影像。利用 low resolution data training 後得到較好的 model performance，針對影像中的細節、紋理都皆有更好的高解析度還原，如下圖 14、15、16 的比較。

(圖 14) 原錄影檔

(圖 15)模糊處理

(圖 16) Super resolution

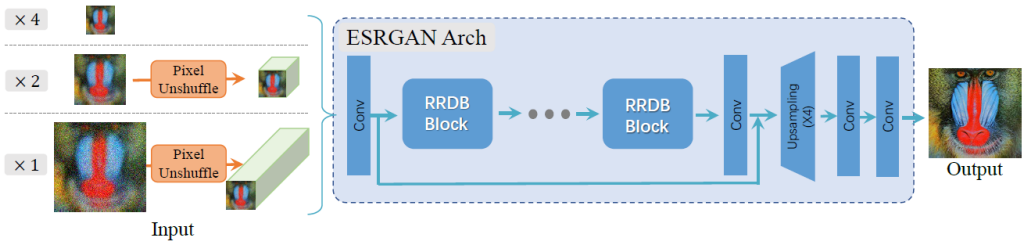
Real-ESRGAN 架構可以分成 Generator、Discriminator 以及 Dataset Construction 三部分(如下圖 17)：



(圖 17) GAN 架構

(一) Generator (下圖 18)：

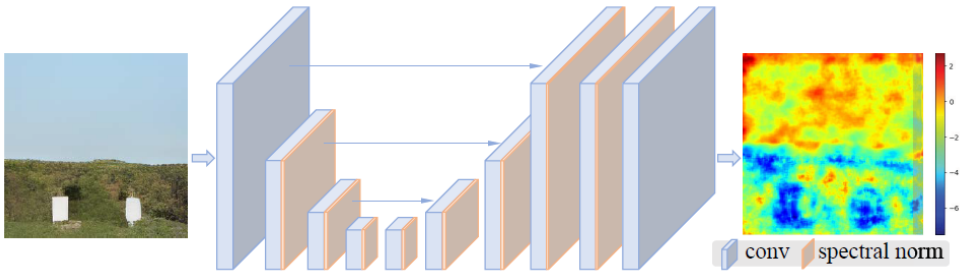
Generator network 架構與 ESRGAN 中 generator network 大致相同，主要由數層 convolutional layers、activation layers(PReLU)所組成，並利用 Dense Block 的架構組成更深的 network，讓 model 有較好得 performance。



(圖 18) Generator架構圖

(二) Discriminator (下圖 19)：

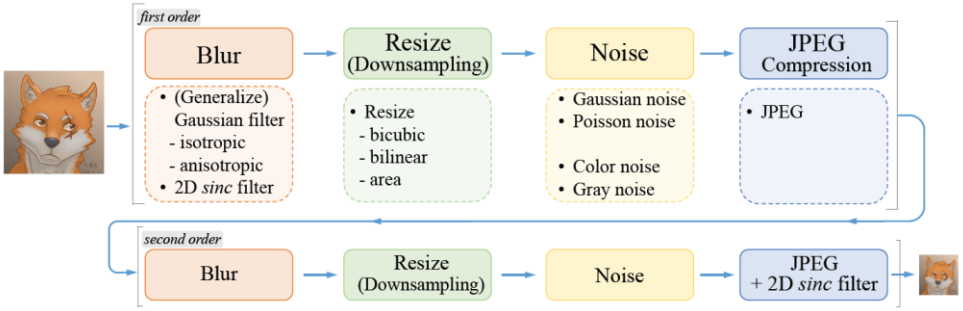
Discriminator 的架構是由 U-net 結合 Spectral normalization layer，由於在 training Real-ESRGAN model 的過程中使用了更複雜的 data set，因此需要更完善的 discriminator 對 generator 生成的圖像進行判別。在 ESRGAN 中的 discriminator 是以圖像的整體角度判別真偽，而 Real-ESRGAN 使用 U-Net 架構可以用更細微的像素角度，對每個像素進行真假判斷，因此能有更好的成效。



(圖 19) Discriminator架構圖

(三) Data Construction (下圖 20)：

為了更好的模擬真實世界的影像，Real-ESRGAN 採用 second-order degradation model 將原本 high resolution 影像轉為 low resolution，並且經過更複雜的處理過程，其中包含：Blur、Resize、Noise、JPEG compression 等等，讓合成出的影像模擬真實影像中的雜訊，並加入 2D *sinc* filter 模擬 ringing 與 overshoot 的情況。

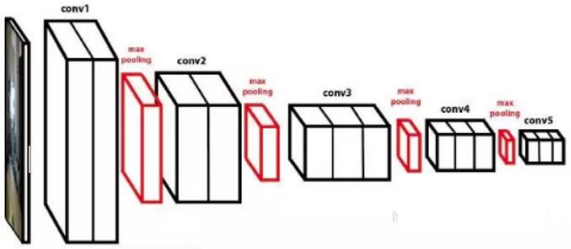


(圖 20) Data construction流程

三、Speech to Text

本實驗在 speech to text 方面，主要目標為中文的 speech to text 處理，值得注意的是中文的同音字很多，這也是其比英文的 speech to text 較困難許多的一主要原因，也因此網路上的資源數量遠不如英文部分。後來參考了一位中國人所開發之 Chinese ASRT(Auto Speech Recognition Tool)，其主要會把 input 的 audio 轉換為 spectrogram 的形式，並且使用在訓練圖形資料常見的 VGG 去進行訓練。訓練的資料主要來自 openSLR 上提供的免費中文語音資料。由於模型最後是先以英文拼音的形式先輸出再轉成中文字，而中文總共有 1423 種發音，因而模型的最後一層大小為 1424(1423 拼音 + 1 blank)。另外，由於該 model 的輸出為簡體字，因此需額外做處理，將簡體字轉為繁體字。

關於 Model 架構(見下圖 21)，其是由十層左右的 convolution layers，搭配 maxpooling 及 dropout 等組成。Activation function 選用 ReLU，loss function 則使用 CTC (Connectionist Temporal Classification) loss function。



(圖 21) 模型示意圖
(參考至 Chinese ASRT 原網站)

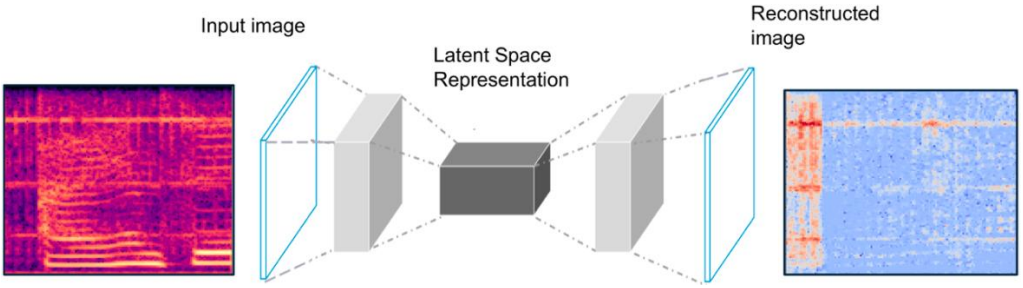
四、Speech Denoising

本實驗在 Speech Denoising 方面，找到兩份 model 分別實作了 U-net 以及 DSTN-LSTM 兩種架構。實驗部分，我們自行錄製多種不同的噪音，包含常見的車聲、雨聲、人說話聲、鬧鈴聲及機器聲等，並自行錄製多個 audio files (參雜上述一種噪音)，看看兩份 model 在 denoising 上的處理效果。更進一步，我們也直接取用架設好的系統網頁，實際進行錄製功能，並將錄製的影音的音訊提出，同樣傳入兩 model 中，比較兩者效果。而在經過多個實驗後，雖然部分表現上，DSTN-LSTM 效果較好，但在整體上，U-net 的表現略顯穩定一些，因此我們的架構主要以 U-net 作為 Denoising 的處理 Model。

(一) 關於 U-net Model 的主要架構，可見圖 22、圖 23。

關於 Denoising 部分(見圖 22)，使用的是 U-net，主要目標為預測傳入的 audio spectrum 中，其 noise 部分的 spectrum，架構上分為前半段的 input data 降維，和後半段的 data 升維兩部分。實作在我們系統上的 U-net 總共有 10 層，每層皆由 4~6 不等的 convolution layers 組成，第 1~4 層搭配 Max Pooling (pool size 為 2*2) 進行降維，第 6~9 層則會先和第 1~4 層的 output data 做 concatenate 進行 data 升維，再傳入 Convolution 進行 feature extraction 和 neural weight 的訓練，而中間的第五層為過度，使用 4 個一般的 convolution layer，無搭配 Max Pooling 或 Concatenate，第 10 層則為輸出層，含一 Convolution layer。

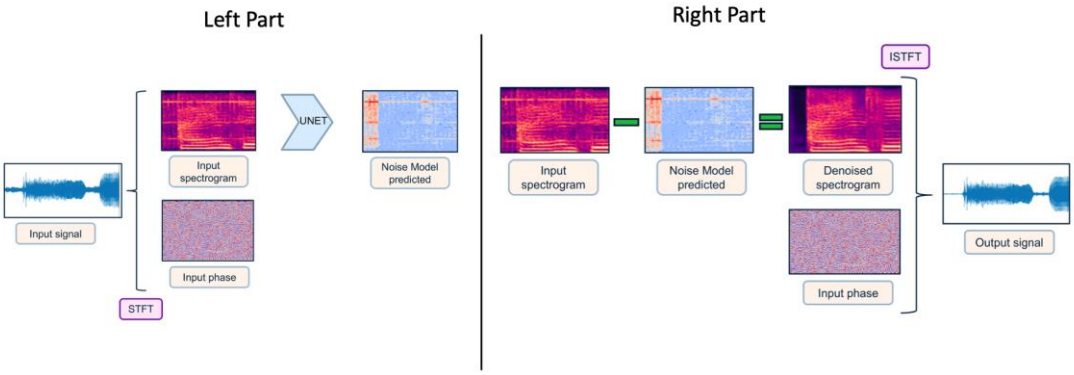
使用的 activation function 是 Leaky Relu，loss function 為"Huber loss"。Huber loss 的特點在於其為 L1 loss 及 L2 loss 的折中，會根據不同條件，使用 L1 loss 或 L2 loss，幫助 weight update。



(圖 22) Unet Model架構

關於 Project 整體架構運行上(見圖 23)，左半邊進行為 De-noising 前，針對 input audio file 做的前處理，主要會將 audio 分為 spectrum 及 phase 部分，而 Denoising 主要是針對 Audio 的 spectrum

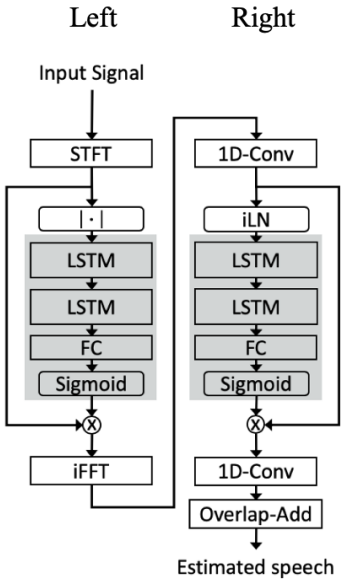
進行處理。右半邊部分則是完成 denoising 後的後處理，會將原始提出的 spectrum 減去**預測的 Noise spectrum 部分**，產生 Denoised 的 spectrum，最後再與先前提出的原始 phase 進行合成輸出。



(圖 23) Unet Project整體架構

(二) 關於 DSTN-LSTM project 的主要架構，可見圖 24。

整體架構分為兩個 separation core (左右)，每個 separation core 包含兩個 LSTM 及一個 Fully Connected layer。左半邊會以 audio 經過 STFT (short time fourier transform) 產生的 frequency domain data 進行 training，右半邊則是會先將 data 藉由 iFFT (inverse Fast fourier transform) 轉換回 time domain data，並藉由一層 Convolution layer 生成 data frame 的 feature representation 後，再傳入右半邊的 separation core 進行 training，最後經由一層 Convolution layer 將 feature representation 轉換回 time domain，並輸出 prediction。



(圖 24) DSTN-LSTM架構

伍. 現有相關研究概況及比較

一、Instance segmentation

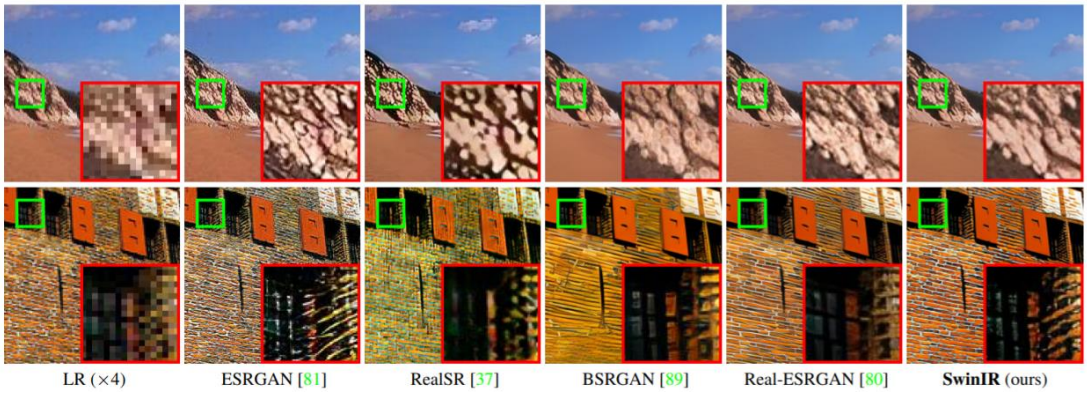
以現有的視訊會議軟體(Google Meet)更換背景的功能做比較，兩者 instance segmentation 的成效相去不遠，最大的差別在於人像與背景交界處模糊的處理方式，如下圖 25、26 比較，Google Meet 用大範圍且柔和的方式漸層式的模糊交界，而我們的系統則可以明顯看出利用模糊效果所框出的邊界，但整體上仍有良好的成效。

(圖 25) Google Meet 更換背景

(圖 26) 我們的成果

二、Super resolution

相較於今年 8 月最新發表的論文 "*SwinIR: Image Restoration Using Swin Transformer*"，Real-ESRGAN 的表現，在細節紋理處理的部分仍有提升的空間，可參考下圖 27 *SwinIR* 論文中的比較，上半部岩石的部分看起來 Real-ESRGAN 較符合真實世界的影像，但在下半部牆壁紅磚的處理則是 SwinIR 有更細微的處理，整體畫面也較清晰銳利。



(圖 27) Real-ESRGAN與SwinIR比較

三、Speech to text

在 speech to text 方面，目前做的最好的是 Google，但其只有提供 API 而無法得知其詳細做法。而網路上比較成熟的 speech to text 模型也幾乎都以英文為主，中文部分則相對較少。起初搜尋資料的過程中，較常找到以與時間序列較相關的 model 實做 speech to text，譬如 RNN 或其變形 LSTM 及 GRU 等，但最後發現若要以 Model 進行訓練的話，將語音先轉成 spectrogram，再將其以圖片當作輸入的方式訓練似乎效果會較好。

四、Speech Denoising

實作 Speech Denoising 部分，在進行搜尋的過程中，了解 U-net、DSTN-LSTM、Wavenet 及 SEGAN 等 model 皆會用來實作 denoising 的處理，另一方面，也有人提出以 speaker recognition 和 speaker separation 為基底，將 Noisy 的部分視為另一位 user，以此將真正說話者的音訊提取出來。其中我們主要比對了 U-net 和 DSTN-LSTM 兩者的效果，在我們實驗的多種噪音中，其中在鬧鈴聲的噪音上 DSTN-LSTM 沒有處理得很好(圖 28 為 U-net 結果，圖 29 為 DSTN-LSTM 結果)(在兩圖中，紅框處有腦鈴聲響起)，下圖也許示意較不顯著，但實際聽起來，鬧鈴的聲響明顯沒有被過濾掉，也是由於此

錯誤，因此我們決定使用雖然部分表現效果略遜 DSTN-LSTM，但全面表現較穩定的 U-net。

(圖 28) U-net結果

(圖 29) DSTN-LSTM結果

陸. 專題重要貢獻

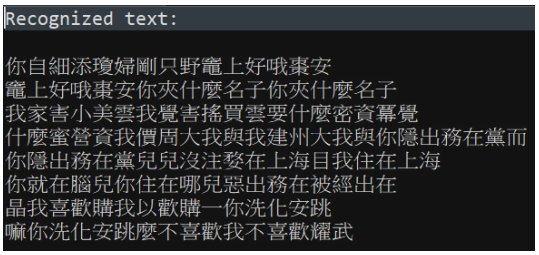
這次的專題中，使用者介面包含會議錄影的功能、網頁的排版與設計。在影像處理方面，包括 Instance Segmentation、Super Resolution 的實作與影像亮度調整。在 Speech to text 部分，主要針對 Google 及參考之 Chinese ASRT 的時間限制進行一定的改善，兩者可以容忍的語音輸入皆有限(Google 目前不免費提供對於太長語音的 speech to text，而會議系統所需處理的語音通常都是數十分鐘到數小時)。我們利用了語音跟語音之間的 silence 去進行長語音的切割再去執行 speech to text，不只解決了 input 時間限制的問題，同時也讓最後輸出的文字能夠不全連在一起，進而避免造成閱讀上的吃力。最後在 Denoising 方面，我們了解到原先應用於醫學影像的 U-net 在 Speech Denoising 上，也可以有不錯的表現。

柒. 效能評估與成果

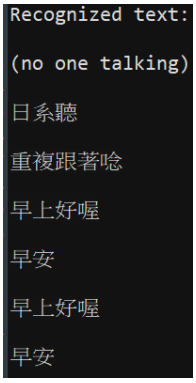
在這次的專題研究主題中，instance segmentation 的部分已達到我們期望的成果，只要在網路順暢地環境下，便能提供給使用者良好的視覺體驗。super resolution 的功能在大多數的情況下亦能有效將低解析度的影像還原成高解析度的影像，但在某些特殊的環境下，例如：使用者所在的環境光線過於明亮，這類影像經過我們的系統還原後會有過度磨平紋理的情況，因此視覺上看起來會略顯不符合真實情況，這也是我們可以繼續改進的部分。

在 speech to text 部分我們將目前最好的 Google speech to text API 以及找到的 Chinese ASRT 進行一定程度的比較，明顯就可以看出 Google 在各方面的表現都遠勝此 ASRT，不論是中文 character、中文單詞亦或是整個句子的正確率，Google 都能夠比 ASRT 做更加精準的翻譯，character 的 accuracy 大致為 96%。反觀使用 VGG 進行訓練的結果，原作者表示 accuracy 可以來到 80%(作者沒提是哪個 accuracy)，但實際運作起來對於中文單一 character 的正確率似乎不到 60%，但純以發音而言是有達到 accuracy 80%以上的。因此若要再對此 model 進行改進，可能需要注重的點是如何將句中前後單詞進行關係連接的相關訓練，提高其在遇到某發音時，能夠正確去判斷應該為哪

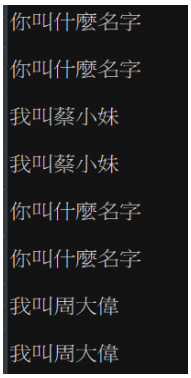
個字(ex. 同樣碰到「ㄗㄠˇ」，其要能判斷「安」字前應該為早安的「早」，而非棗子的「棗」) (見下圖 30~34)



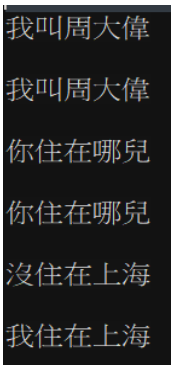
(圖 30)



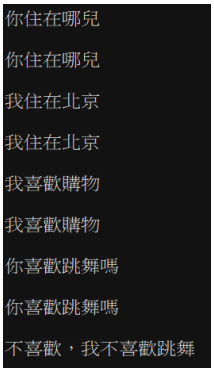
(圖 31)



(圖 32)



(圖 33)



(圖 34)

最後關於 Denoising 的部分，我們認為錄製的音訊，雖然無法做到完全消除噪音，但比對原先的音訊，噪音確實降低了不少，而 DSTN-LSTM 的 Model 雖然在部分的噪音狀況下，表現較 U-Net 好，但鬧鈴聲的噪音上，降噪效果卻出了問題，如何調整 model 參數，或是否考慮藉由其他 audio attribute 進行 training，皆是可以考慮改進方向，進而讓 DSTN-LSTM 更全面的表現。

捌. 團隊合作方式

一、陳伯瑾主要負責的部分為 Speech Denoising 部分，以及 UI 方面，包含 user 與 server 之間的資料傳遞、影音錄製功能和 Record Page 的實作等等。

二、黃寶萱主要負責影像處理的部分，包括 Instance Segmentation、Super Resolution 的實作與影像亮度調整，以及會議整體畫面錄影的功能、網頁 UI 的排版與設計。

三、陳柏均主要負責中文的 Speech to Text，以及將錄製下的影音存到 local 端的資料夾中以便 offline record post-process。

玖. 結論

在這份專題中，建築於視訊會議的基底下，我們結合不同的機器學習模型與深度學習模型技術，在影像與語音兩方面，皆為會議系統的品質帶來了提升。

更換會議背景(instance segmentation)方面，藉由 PointRend model 分割人像與背景，為使用者增加會議期間的隱私度。提高影像解析度(super resolution)則是使用 Real-ESRGAN model 實作將模糊的低解析度影像轉為高解析度，並搭配 Speech Denoising 使用 U-net 進行降噪，優化錄製的會議記錄。最後關於 Speech to Text，則先將語音進行切割，再運用 Google API 進行文字轉換處理，進而幫助使用者產生會議過程中，語音訊息的文字檔，讓使用者在回顧會議時，搭配文字訊息，可以更清晰的瞭解會議過程的內容。

拾. 參考文獻

1. Kaiming He, Georgia Gkioxari, Piotr Doll´ar, Ross Girshick：Mask R-CNN
2. Alexander Kirillov, Yuxin Wu, Kaiming He, Ross Girshick：PointRend-Image Segmentation as Rendering
3. Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan：Training Real-World Blind Super-Resolution with Pure Synthetic Data
4. Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, Xiaoou Tang：Enhanced Super-Resolution Generative Adversarial Networks
5. Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi：Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network
6. Nils L. Westhausen and Bernd T. Meyer：Dual-Signal Transformation LSTM Network for Real-Time Noise Suppression
7. 中文語音轉文字模型
   https://github.com/nl8590687/ASRT_SpeechRecognition
8. Google API 使用方法
   https://yanwei-liu.medium.com/%E4%BD%BF%E7%94%A8python%E9%80%B2%E8%A1%8C%E8%AA%9E%E9%9F%B3%E8%BE%A8%E8%AD%98-%E8%81%B2%E9%9F%B3%E8%BD%89%E6%96%87%E5%AD%97-9ab12b750ffe
9. 中文語音資料來源
   https://www.openslr.org/38/
10. Speech-enhancement model
    https://github.com/vbelz/Speech-enhancement
11. Vincent Belz : Speech-enhancement with Deep learning
12. Transformation LSTM Network
    https://github.com/breizhn/DTLN#dual-signal-transformation-lstm-network
13. WebRTC
    https://github.com/BJ0815/WebRTC_Practice/tree/sample_03