

唇語辨識與雙向傳輸系統

LipNet Connect : A Lip Reading and Bidirectional Transmission System

組員：電機四 李季蓓

導言

LipNet Connect 的設計初衷是為了解決在吵雜環境中溝通不便的問題，或是為聽障者提供更便利的交流方式。無論是透過鏡頭捕捉影像，或是無聲無字幕的影片，皆能提供即時的唇語分析。

這是一個基於深度學習的唇語辨識系統，結合影像傳輸技術，並透過ZeroTier內網穿透，提供穩定且安全的資料傳輸，實現雙向影像傳輸與即時字幕預測。使用者僅需操作四個按鈕，即可啟動伺服器、建立連線、開始預測並傳輸字幕，輕鬆完成本地與遠端的無縫交流。

系統架構

影像捕捉與字幕生成

- 使用本地鏡頭獲取影像作為輸入。
- 透過唇語辨識模型，將影像轉化為文字預測結果。

ZeroTier內網穿透

- 提供端到端加密，確保數據傳輸安全。
- 透過虛擬內網將不同地點的設備連接為同一局域網，簡化配置並實現跨網路點對點連接。

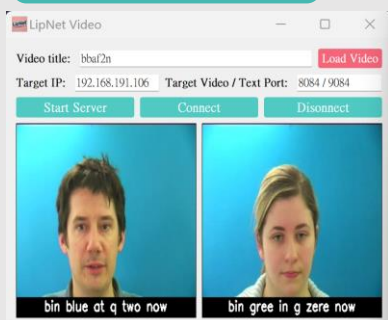
雙向傳輸

- 利用UDP協議傳輸影像與字幕，確保低延遲通信。
- 系統將自動抓取本地及遠端虛擬IP，讓操作更加直觀。

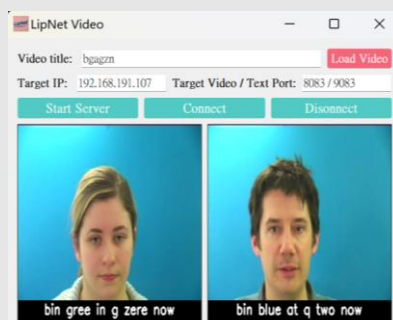
視窗顯示

- 左側顯示本地影像與字幕，右側顯示遠端影像與字幕。
- 設計四個按鈕 (Start Server / Connect / Disconnect / Start Prediction)，並顯示提示字樣，方便使用者操作。

影片實驗結果



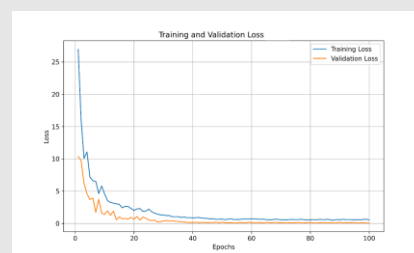
圖[2] 影片分析:本地視窗畫面



圖[3] 影片分析:遠端視窗畫面

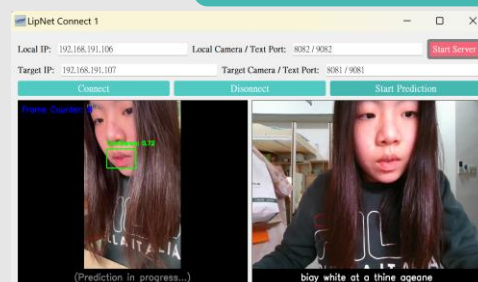
優化部分

- 提升模型訓練速度
- 降低模型訓練Loss值
- 加速鏡頭即時預測速度

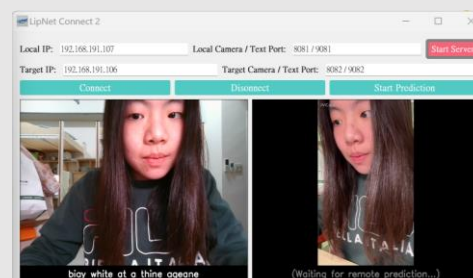


圖[1] 訓練與驗證損失曲線

即時影像結果



圖[4] 即時影像分析:本地視窗畫面



圖[5] 即時影像分析:遠端視窗畫面

結論

在系統開發過程中，深入了解了唇語辨識技術的原理與影像傳輸的技巧，並體會到解決技術挑戰的重要性。LipNet Connect 系統成功結合了深度學習與影像傳輸技術，實現了即時字幕預測與影像傳輸。透過ZeroTier的內網穿透功能，克服了遠端傳輸中不同區域網的挑戰，並提升了系統的穩定性與效能。