

國立台灣海洋大學資訊工程學系
專題報告

Sign Language Interaction Tutorial System
- Openpose

Instructor:

張欽圳 教授

Submitted by:

00957209 林心怡 (00957209@mail.ntou.edu.tw)

00957202 吳秉宸 (00957202@mail.ntou.edu.tw)

00853029 張正德 (00853029@mail.ntou.edu.tw)

DEPT. OF

NATIONAL TAIWAN OCEAN UNIVERSITY

MAY 2022

ABSTRACT

Keywords - Lorem Ipsum , Lorem Ipsum

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Contents

Abstract	i
CHAPTER 1: 緒論	1
1.1 研究背景與目標	1
1.2 現有相關研究概況	1
1.3 系統概述與流程圖	2
1.4 章節概述	2
CHAPTER 2: 技術介紹與方法設計	3
2.1 技術介紹	3
2.2 方法設計	4
CHAPTER 3: 系統實驗	5
3.1 實驗設計	5
3.2 效能評估與成果	6
CHAPTER 4: 結論與討論	7
CHAPTER 5: 工作分配	8
Appendices	8
References	9

Chapter 1

緒論

1.1 研究背景與目標

手語是聾啞人士很重要的溝通方式，生活中難免碰到一些聾啞人士，想要與他們交流需先學習手語，但網路上現有的資源幾乎都是用影片學習的方式，沒有可以自我檢視的系統，因此本專題設計一套主要提供手語教學、測驗及偵錯的功能的系統，提供給想要入門手語的初學者一個學習的平台。

1.2 現有相關研究概況

對於手語辨認技術，有的採用Dynamic time warping(DTW)的方式來辨識手語，其方法是比對兩個時間序列的差異並利用K-Nearest Neighbors演算法判斷辨識結果，時間複雜度為兩個時間序列的長度的乘積—— $O(mn)$ ，當時間序列長度越長所需時間越長，而在現今的手語辨認技術上，已經可以利用深度學習辨識出動態的手語片段，但無偵測錯誤的功能。

1.3 系統概述與流程圖

本系統主要提供手語教學、測驗及偵錯功能，其中測驗及偵錯為本專題的核心功能，可細分為三個階段：（一）辨識與擷取支幹特徵、（二）由神經網路進行分類及（三）利用神經網路其中一層的梯度資訊找出決策的依據進行偵錯。

系統的第一階段，選擇使用Mediapipe擷取支幹特徵，做適當的影片裁減、支幹點座標的正規化。

系統的第二階段，嘗試設計CNN、LSTM與Transformer三種不同的神經網路架構，經過實驗，最終採用最符合本系統的——CNN架構。

系統的第三階段，選擇使用Grad-cam技術，擷取神經網路其中一層的梯度資訊，找出各項類別的關鍵偵數，推斷出該時間序列的錯誤偵數。

其中，本系統的流程圖如下：



1.4 章節概述

Chapter 2

技術介紹與方法設計

2.1 技術介紹

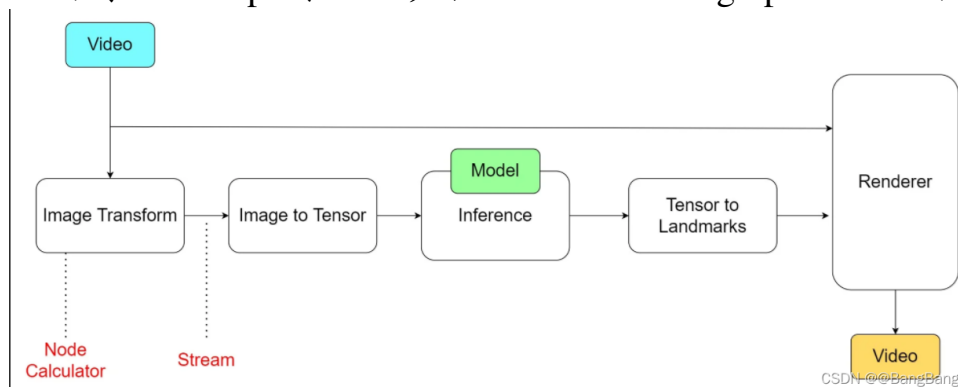
本系統將架構分為三部分:

1. MediaPipe:

是一款由Google 開發並開源的數據流處理機器學習應用開發框架。它是一個基於圖的數據處理管線，用於構建使用了多種形式的數據源，如影片、音頻、傳感器數據以及任何時間序列數據。

MediaPipe提供人體支幹和手部的關鍵點，又因為MediaPipe的輕量、以及性能良好等優點，為本專題選擇的工具之一。

下圖為MediaPipe對於一影片做hand solution graph的流程圖



2. 神經網路:

- Transforer:
- Convolution:
- LSTM:

3. Grad-Cam

模型最後一層捲積層生成的每張特徵圖在經過GlobalAveragePooling後變成一個像素包含了整個特徵圖的訊息，CAM的技術是利用反向思考，因為經過GAP後的像素陣列會乘以權重 w ，權重 w 值越大代表該像素的影響越大，

改良了CAM最後一層必須是GlobalAveragePooling的限制。

Grad-CAM 的思想即是「不論模型在卷積層後使用的是何種神經網路，不用修改模型就可以實現CAM」，最後不論是全連接層、RNN、LSTM 或是更複雜的網路模型，都可以藉由Grad-CAM 取得神經網路的分類關注區域熱力圖(Heatmap)。

Grad-CAM關鍵是能夠透過Back Propagation計算在CAM中使用的權重

2.2 方法設計

本專題利用三種神經網路架構(Transformer、CNN、LSTM)實踐辨識手語，並分別應用Grad-Cam演算法於時間序列資料達到錯誤偵測的功能。探討這三種不同的神經網路模型對於二維時間序列資料分類的準確度，與分析對於三種模型結合Grad-Cam的效果，研究出最適合本專題應用的神經網路架構。

Chapter 3

系統實驗

3.1 實驗設計

1. 手勢類別辨識利用混淆矩陣針對辨識結果

$$\text{Accuracy} = (\text{TP} + \text{TN}) / \text{Tot.N}$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

score值介於0-1，1是最好，0是最差。

-	實際正確	實際錯誤
預測正確	TP	FP
預測錯誤	FN	TN

2. 特徵擷取工具

3. 錯誤偵測精準度判斷

3.2 效能評估與成果

Chapter 4

結論與討論

Chapter 5

工作分配

學號	姓名	負責項目說明
00853029	張正德	Transformer、Grad-cam
00957202	吳秉宸	Convolution、Grad-cam
00957209	林心怡	LSTM、資料處理

Appendix

Bibliography

- [1] 黃致巽, ”基於動態時間扭曲之人體姿態辨識Dynamic Time Warping Based Recongition Of Human Body Gestures”,碩士論文, 國立臺灣師範大學資訊工程研究所, 2013/7
- [2] Feature Extraction with Video Summarization of Dynamic Gestures for Peruvian Sign Language Recognition, Andre Neyra-Gutierrez Department of Computer Science Universidad Peruana de Ciencias Aplicadas Lima, Per
- [3] Sign Language Action Recognition System Based on Deep Learning, Chaoqin Chu, Qinkun Xiao, Jielei Xiao, Chuanhai Gao, 2021
- [4] Arabic Sign Language Recognition System Using 2D Hands and Body Skeleton Data, Mohamed A. Bencherif, 2021/4
- [5] Isolated Sign Recognition from RGB Video using Pose Flow and Self-Attention, Mathieu De Coster, Mieke Van Herreweghe, Joni Dambre, 2021

PAPER ACCEPTANCE PROOF

REGISTRATION PROOF

SCOPUS INDEXED CONFERENCE PROOF