Taiwan Sign Language Translation

Team: 王蕣婷 林于湘 林� 洪睿謙 陳郁婷 張育禔 楊畯棋 賴藍平
Presenter: 洪睿謙 陳郁婷
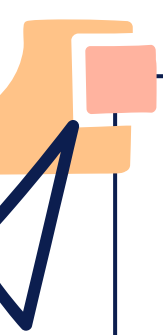
久/相見/ 無
=好久不見

Contents

01

Introduction

Why we started this project?

02

Data & Methods

How to make this happen?

03

Experiments

What have we done?
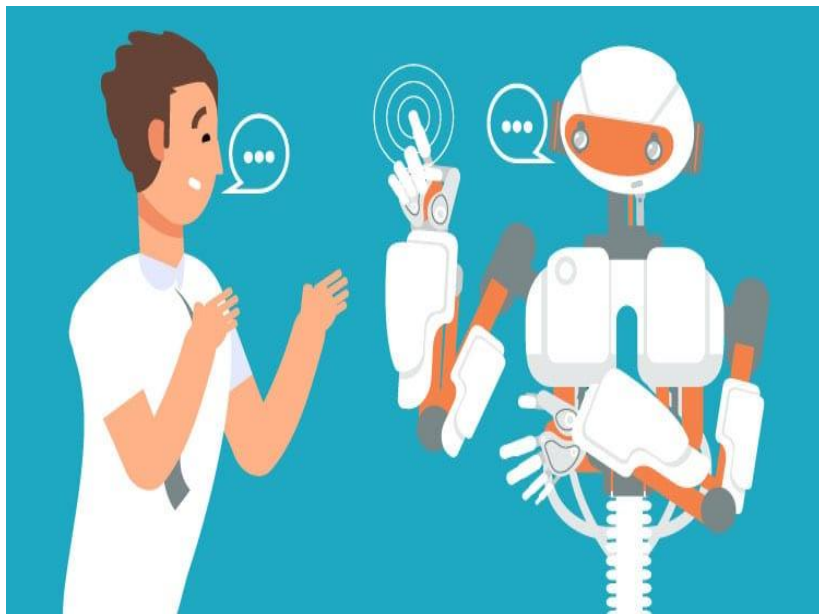
04

Conclusion & Prospect
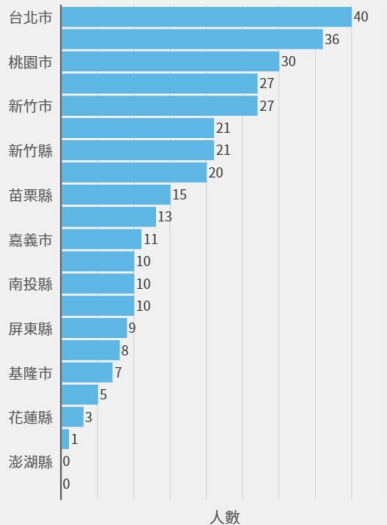
What's next?

01

INTRODUCTION

Motivation

Goal

Motivation

雅婷逐字稿

Why not...

Motivation

探究台灣手語翻譯服務現況，思考解決方案。期望改善使用者可能面臨的困境並提升生活品質與便利性。

人力缺口

培訓不易

專業門檻

場域限制

品質維護

城鄉落差

收入不穩

制度需求

提供社福服務之手譯員，共324名：所有提供社會福利類手語翻譯類人員，包含乙級、丙級及其他未取得證照但符合規定者。

	人數
台北市	40
	36
桃園市	30
	27
新竹市	27
	21
新竹縣	21
	20
苗栗縣	15
	13
嘉義市	11
	10
南投縣	10
	10
屏東縣	9
	8
基隆市	7
	5
花蓮縣	3
	1
澎湖縣	0
	0

Goal

Ordinary People   Interpreter   Deaf People
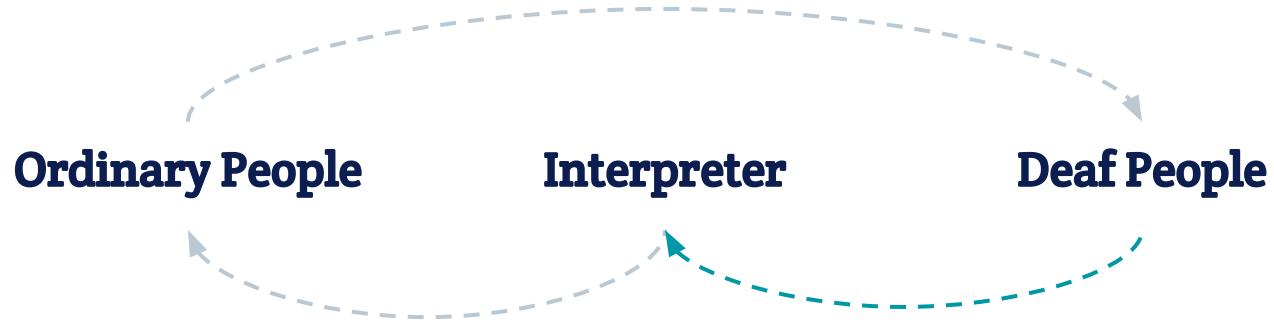
TSL, Translation

Goal

- 手語隱含文化獨特性

- 語言系統與時俱進

- Hand gesture recognition

- Hand pose estimation & tracking

新技術＋AI ＝持續性的研究與學習價值

Taiwan Sign Language
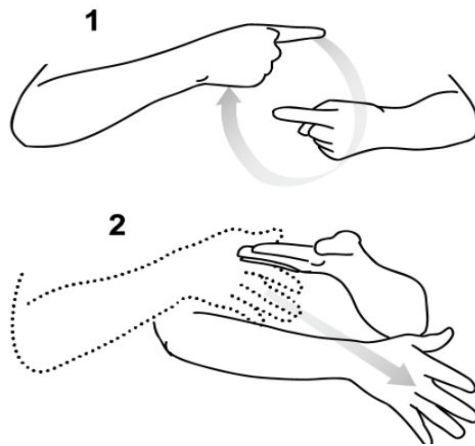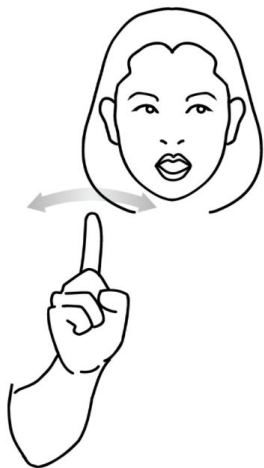
Mediapipe Holistic

02

Data & Methods

Data Collection

手語是多重且連續性的資料, 包含:

1. 手形
2. 位置
3. 動作型態
4. 方向
5. 表情

Data Collection

來源	參考 Youtube 手語教學，自行拍攝
內容	涵蓋六大主題常用語 「年齡、問候、家人、時間、住家、描述」
筆數	共 55 個字詞，可組合至少 25 個短句
特徵	保留手形、位置（胸部以上）、動作型態及方向

有 買 久 一 你 誰 租 棒 零 我 無

今天 昨天 明天 台北 桃園 手語 朋友
什麼 將近 父母 一共 家裡 他們 吃飯
房子 銀行 認識 見她 比較 是嗎 一樣
生日 天氣 上課 孩子 相見 運動 年齡
名字

星期一 星期二 星期三 星期四 星期五
星期六 星期天 捷運站 高鐵到 我問你
還沒有 完了嗎

我們兩個 幾月幾號 會不會呢

「你叫什麼名字」
你　名字　什麼

Data Collection

我問你	我問你	我問你	…	我問你
手語	手語	手語	…	手語
會不會呢	會不會呢	會不會呢	…	會不會呢

我問你/手語/會不會呢

＝ 請問你會不會打手語

**55 words**
60 frames (2 sec * 30 fps)

Hands & Pose Tracking Solution

Hands

- Palm detection model: TFLite model, TF.js model
- Hand landmark model: TFLite model, TFLite model (sparse), TF.js model
- Model card, Model card (sparse)

Pose

- Pose detection model: TFLite model
- Pose landmark model: TFLite model (lite), TFLite model, TFLite model (heavy)
- Model card

Holistic
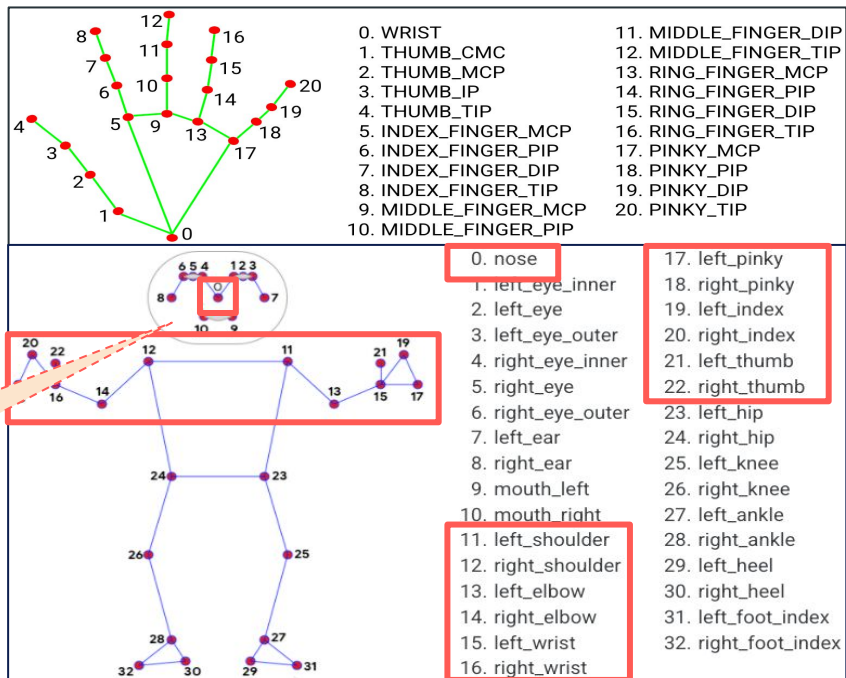
- Hand recrop model: TFLite model

MediaPipe
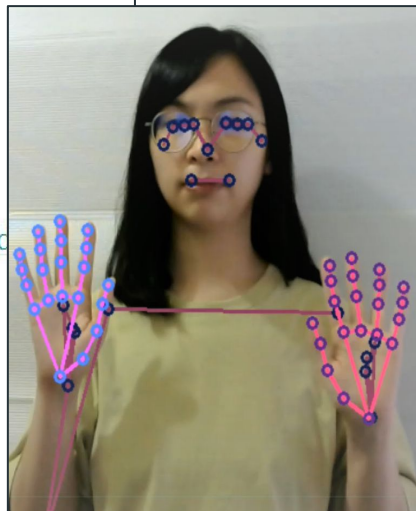
Handedness: [classification {
  index: 0
  score: 0.99946386
  label: "Left"
}
, classification {
  index: 1
  score: 0.99999803
  label: "Right"
}
]
hand_landmarks: landmark {
  x: 0.5342746
  y: 0.57851875
  z: -4.123216e-05
}
landmark {
  x: 0.51628554
  y: 0.5465713

Data Preprocessing

- **Remove unused landmarks (face, pose 1~10, 23~32) and visibility**
- **Keep 55 landmarks (right & left hands, pose 0, 11~22)**
- **Each has xyz axis**
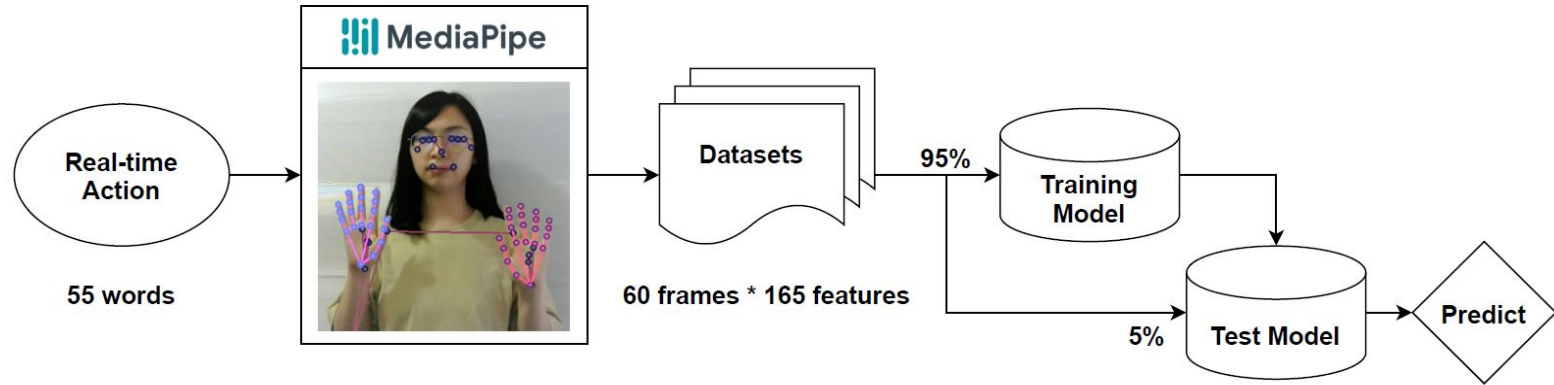- **55 * 3 = 165 features**

以鼻子為中心
轉換相對座標

0. WRIST	11. MIDDLE_FINGER_DIP
1. THUMB_CMC	12. MIDDLE_FINGER_TIP
2. THUMB_MCP	13. RING_FINGER_MCP
3. THUMB_IP	14. RING_FINGER_PIP
4. THUMB_TIP	15. RING_FINGER_DIP
5. INDEX_FINGER_MCP	16. RING_FINGER_TIP
6. INDEX_FINGER_PIP	17. PINKY_MCP
7. INDEX_FINGER_DIP	18. PINKY_PIP
8. INDEX_FINGER_TIP	19. PINKY_DIP
9. MIDDLE_FINGER_MCP	20. PINKY_TIP
10. MIDDLE_FINGER_PIP	

0. nose	17. left_pinky
1. left_eye_inner	18. right_pinky
2. left_eye	19. left_index
3. left_eye_outer	20. right_index
4. right_eye_inner	21. left_thumb
5. right_eye	22. right_thumb
6. right_eye_outer	23. left_hip
7. left_ear	24. right_hip
8. right_ear	25. left_knee
9. mouth_left	26. right_knee
10. mouth_right	27. left_ankle
11. left_shoulder	28. right_ankle
12. right_shoulder	29. left_heel
13. left_elbow	30. right_heel
14. right_elbow	31. left_foot_index
15. left_wrist	32. right_foot_index
16. right_wrist	

Workflow

Real-time Action

55 words

MediaPipe

Datasets

60 frames * 165 features

95%

Training Model

5%

Test Model

Predict

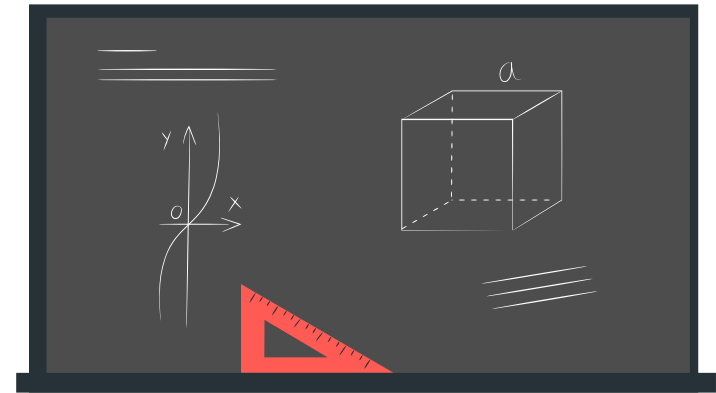**Filming** - - - - - - - - - - - → **Array** - - → **AI Model** - - → **Words**
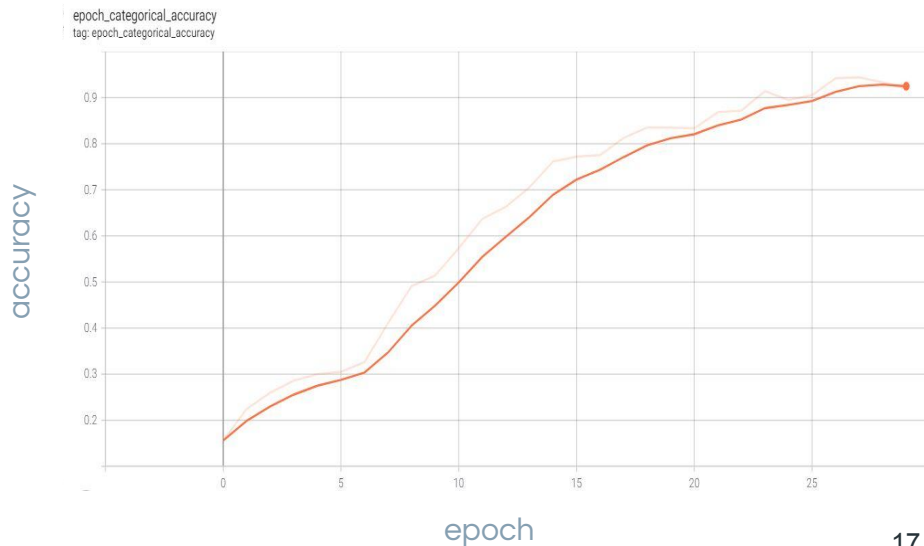
5 ppl

03

EXPERIMENTS

Our Journey Begins Here...

1. Small datasets to test models (10 words)
2. Began with famous RNN model - LSTM
3. Turned to GRU model to increase speed

	LSTM	GRU
Train accuracy	0.7333	0.9193
Train loss	0.7887	0.2065
Test accuracy	0.7333	0.9000
Test loss	0.8419	0.1724

GRU

epoch_categorical_accuracy
tag: epoch_categorical_accuracy

accuracy

epoch

17

Sticking with GRU...

4. Bigger datasets based on GRU model

	GRU-10	GRU-55
Epoch	30	75
Train accuracy	0.9193	0.9573
Train loss	0.2065	0.1321
Test accuracy	0.9000	0.8909
Test loss	0.1724	0.4429
Real-time test	60% (6/10)	42% (23/55)

More experiments based on GRU model

GRU Model Architecture
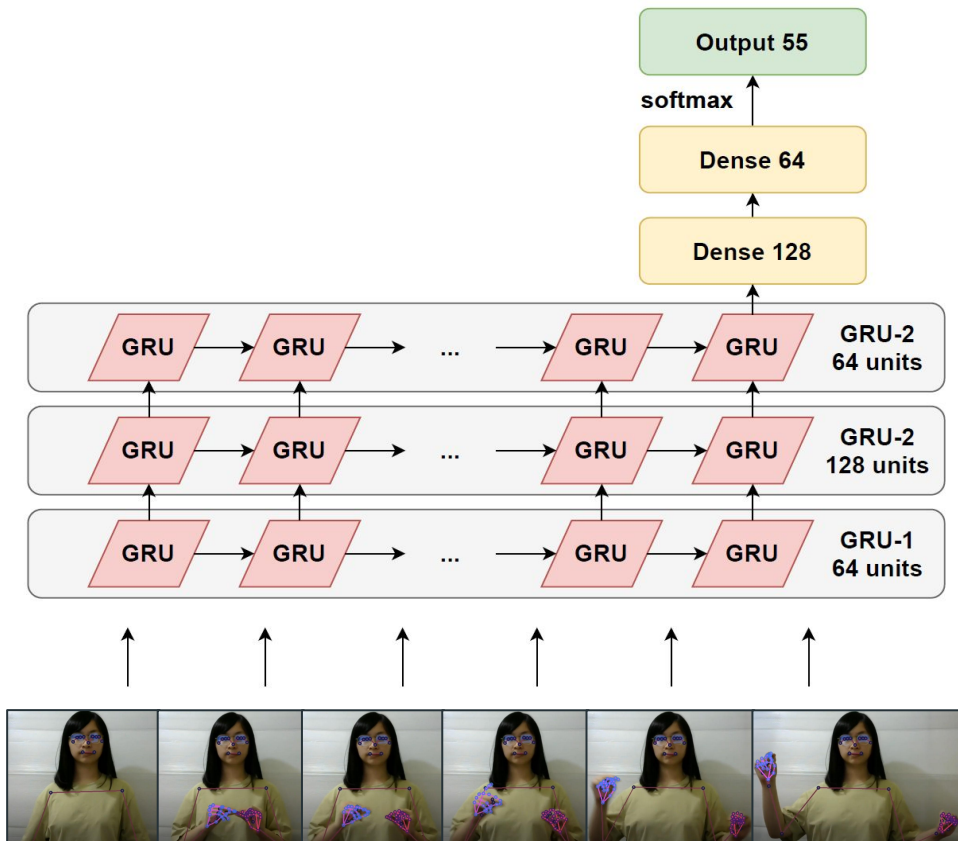
Input shape (60, 165)

Boss 1

GRU model no longer performs
as well with 55 words

What have we done?

Experiments

5. Chose ELU instead of ReLU to preserve features

	ReLU	ELU
Train accuracy	0.9420	0.9910
Train loss	0.1960	0.0420
Test accuracy	0.8850	0.9330
Test loss	0.3750	0.2030
Real-time test	55% (30/55)	67% (37/55)

```
Model: "sequential_1"

Layer (type)              Output Shape              Param #
=================================================================
gru_3 (GRU)               (None, 60, 128)           113280

gru_4 (GRU)               (None, 60, 256)           296448

gru_5 (GRU)               (None, 128)               148224

dense_3 (Dense)           (None, 256)               33024

dense_4 (Dense)           (None, 128)               32896

dense_5 (Dense)           (None, 55)                7095
=================================================================
Total params: 630,967
Trainable params: 630,967
Non-trainable params: 0
```

Discovered a BUG in Coordinate Transformation

BREAKING NEWS

Boss 2

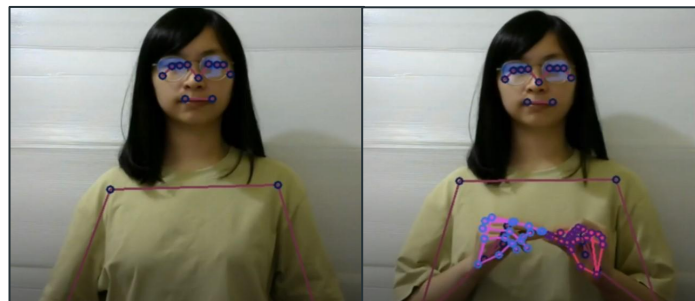**Hand coordinates was not transformed into relative coordinate !!**

What's wrong?

	ELU Old data	ELU New data?
Train accuracy	0.9910	0.9911
Train loss	0.0420	0.0389
Test accuracy	0.9330	0.9333
Test loss	0.2030	0.2127
Real-time test	67% (37/55)	51% (28/55)

We looked into the data

What's wrong?

	ELU Old data	ELU New data?
Train accuracy	0.9910	0.9911
Train loss	0.0420	0.0389
Test accuracy	0.9330	0.9333
Test loss	0.2030	0.2127
Real-time test	67% (37/55)	51% (28/55)

When hand isn't detected it'll be filled with 0
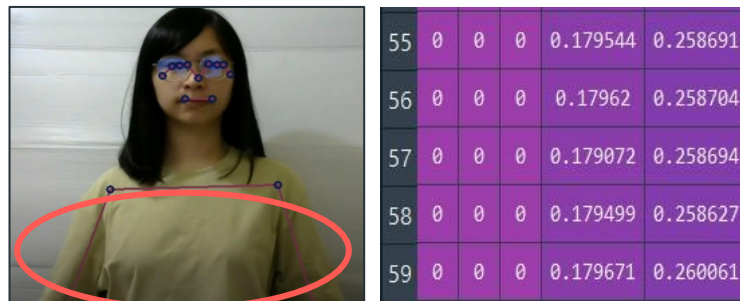
What have we done?

6. Implemented mask to preserve "0"

	ELU New data (0)
Train accuracy	0.9810
Train loss	0.0640
Test accuracy	0.9390
Test loss	0.2110
Real-time test	67% (37/55)
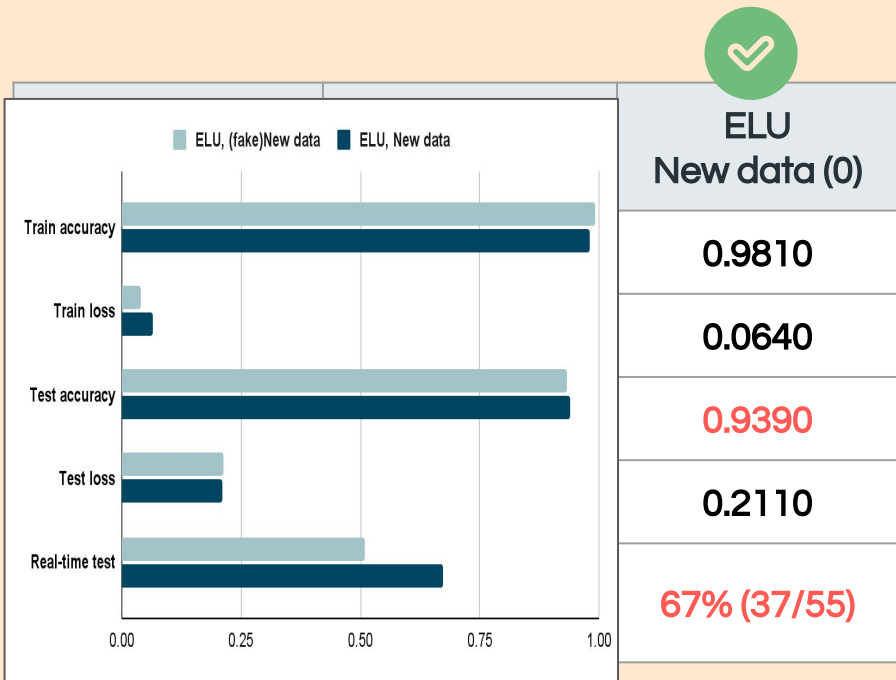
Only keep the results where the value is true
(0 - nose = obtain value)
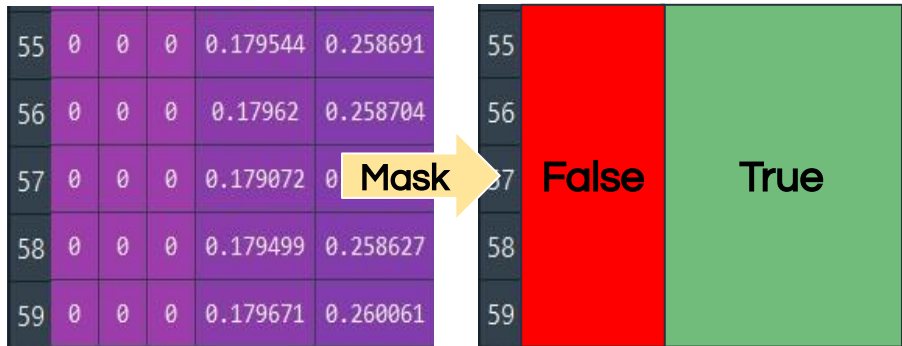
Tuning Model

7. Increase model size (add to GRU 256,512)

	New data (128, 256)	New data (256, 512)
Train accuracy	0.9810	0.9880
Train loss	0.0640	0.0410
Test accuracy	0.9390	0.9450
Test loss	0.2110	0.1700
Real-time test	67% (37/55)	75% (41/55)

26
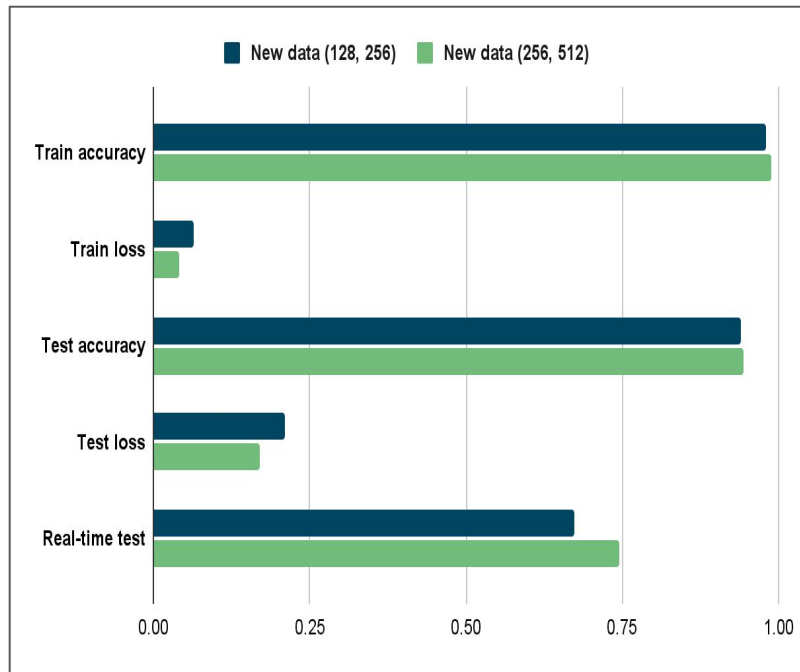
Boss 3
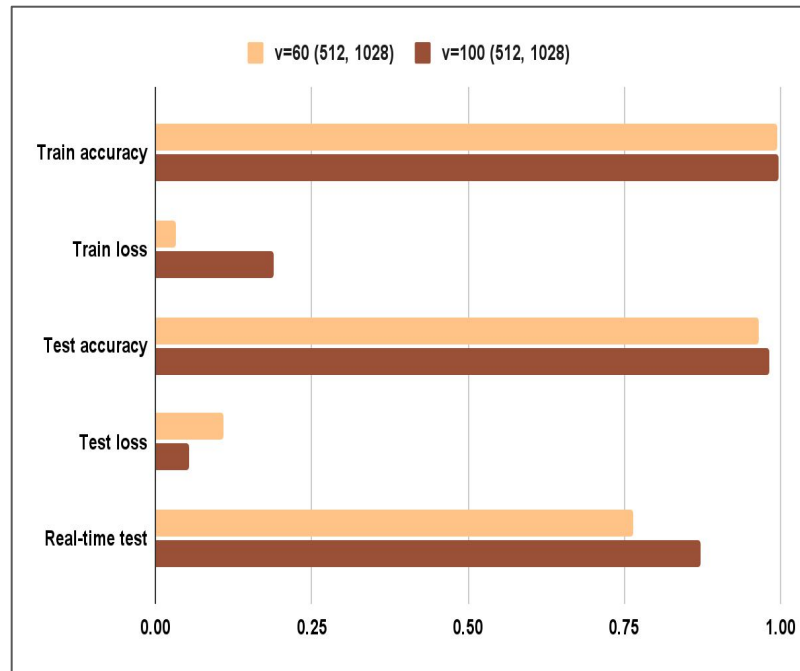
Ceiling: reach the maximum potential

What have we done?

8. Increase model size (add to GRU 512,1028)
9. Increase number of videos (add to 100)

	v=60 (512, 1028)	v=100 (512, 1028)
Train accuracy	0.9940	0.9970
Train loss	0.0320	0.1890
Test accuracy	0.9650	0.9820
Test loss	0.1090	0.0530
Real-time test	76% (42/55)	87% (48/55)

Hyperparameters Fine Tuning

10. Used AutoKeras instead of grid search

RNNBlock

autokeras.RNNBlock(return_sequences=False, bidirectional=None, num_layers=None, layer_type=None)

DenseBlock

autokeras.DenseBlock(num_layers=None, num_units=None, use_batchnorm=None, dropout=None)

```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 60, 165)]         0

cast_to_float32 (CastToFloat (None, 60, 165)           0

bidirectional (Bidirectional (None, 60, 330)           328680

bidirectional_1 (Bidirection (None, 330)               492030

dense (Dense)                (None, 64)                21184

batch_normalization (BatchNo (None, 64)                256

re_lu (ReLU)                 (None, 64)                0

dense_1 (Dense)              (None, 55)                3575

classification_head_3 (Softm (None, 55)                0
=================================================================
Total params: 845,725
Trainable params: 845,597
Non-trainable params: 128
```

29

Hyperparameters Fine Tuning

10. Used AutoKeras instead of grid search

	v=100 (512, 1028)	AutoKeras
Train accuracy	0.9970	0.9991
Train loss	0.189	0.0270
Test accuracy	0.982	0.9927
Test loss	0.053	0.0504
Real-time test	87% (48/55)	95% (52/55)

```
Layer (type)                  Output Shape              Param #
=================================================================
input_1 (InputLayer)          [(None, 60, 165)]         0

cast_to_float32 (CastToFloat  (None, 60, 165)           0

bidirectional (Bidirectional  (None, 60, 330)           328680

bidirectional_1 (Bidirection  (None, 330)               492030

dense (Dense)                 (None, 64)                21184

batch_normalization (BatchNo  (None, 64)                256

re_lu (ReLU)                  (None, 64)                0

dense_1 (Dense)               (None, 55)                3575

classification_head_3 (Softm  (None, 55)                0
=================================================================
Total params: 845,725
Trainable params: 845,597
Non-trainable params: 128
```

04

CONCLUSION & PROSPECT

Review: Comparison of All Model

Legend:
- V1_LSTM(10 words)
- V2_GRU(10 words)
- V3_GRU(55 words)
- V4_GRU(ReLU)
- V5_GRU(ELU)(old)
- V6_GRU(ELU)(fake New data)
- V7_GRU(ELU)(New data)(128,256)
- V8_GRU(ELU)(New data)(256,512)
- V9_GRU(512,1028)(v=60)
- V10_GRU(512,1028)(v=100)
- V11_AutoKeras
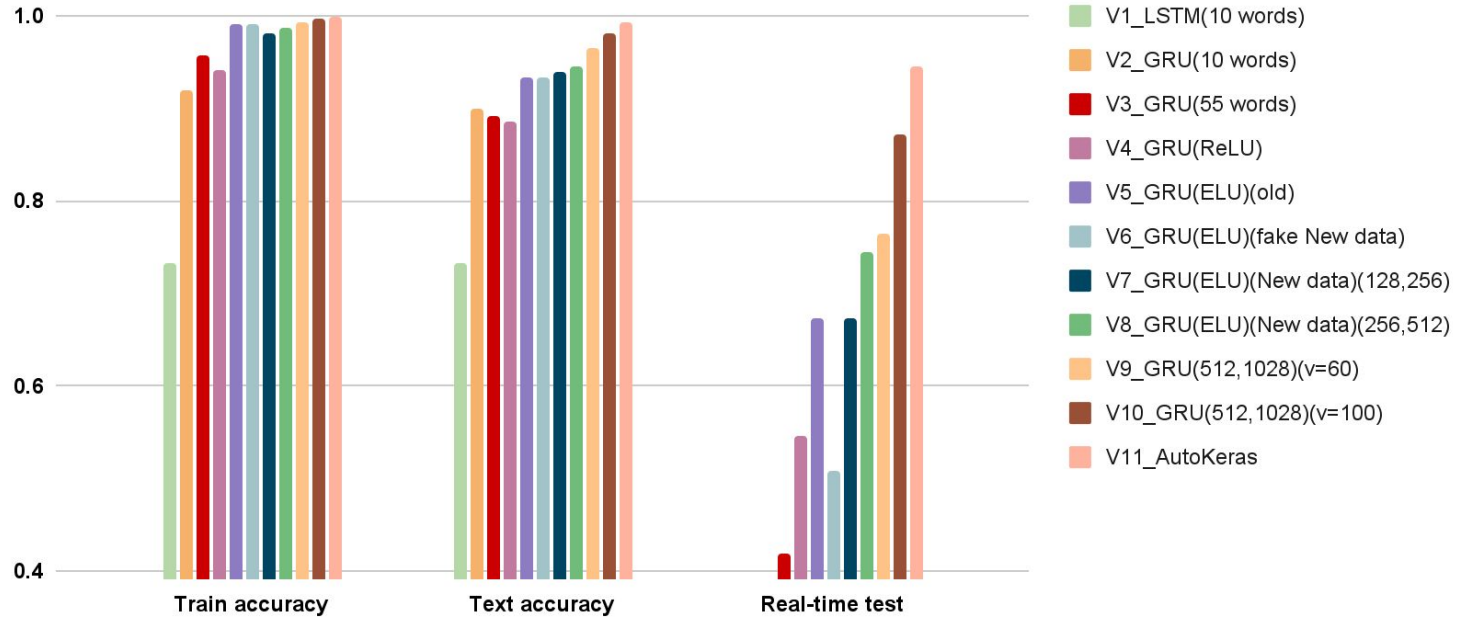
Review: Comparison of Datasets
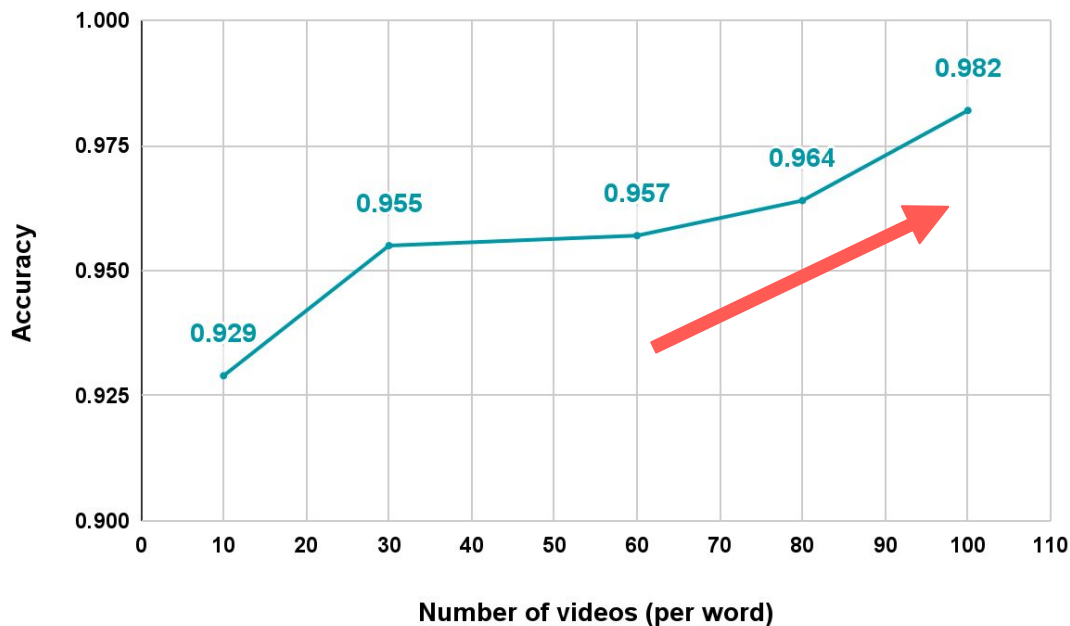
More words

More videos

32

Conclusion

- AutoKeras outperforms V10 (GRU 512) on all numerical aspects
- Currently V10 has the upper hand in terms of real-time due to technical difficulties (time)

Conclusion

- ## GRU performs better on smaller datasets
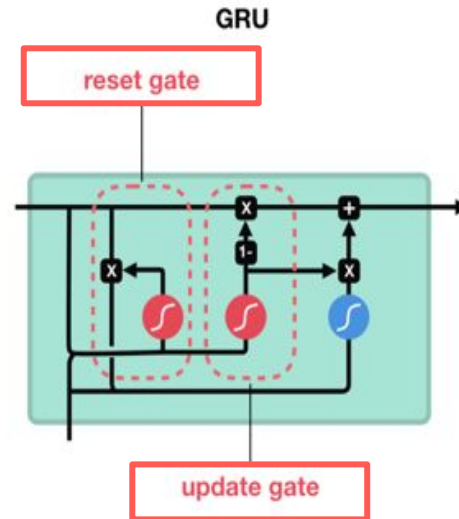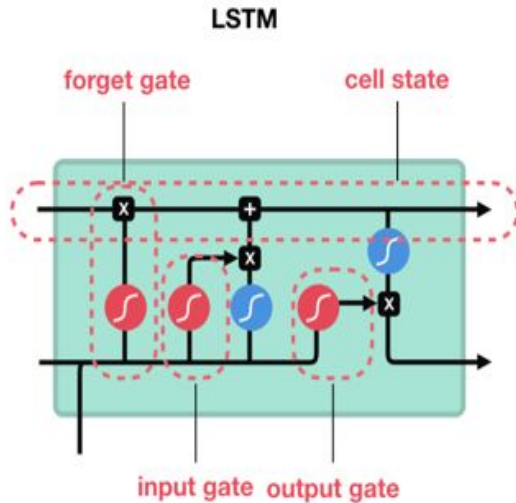
Comparable with previous TSL project

	Previous	Ours
Words	7	55 ↑
Number of videos	200	100 ↓
Frames	90	60 ↓
Model	LSTM	GRU
Best test accuracy	0.9963	0.9927

Real-time test:

95%

(52/55)

Confusion Matrix
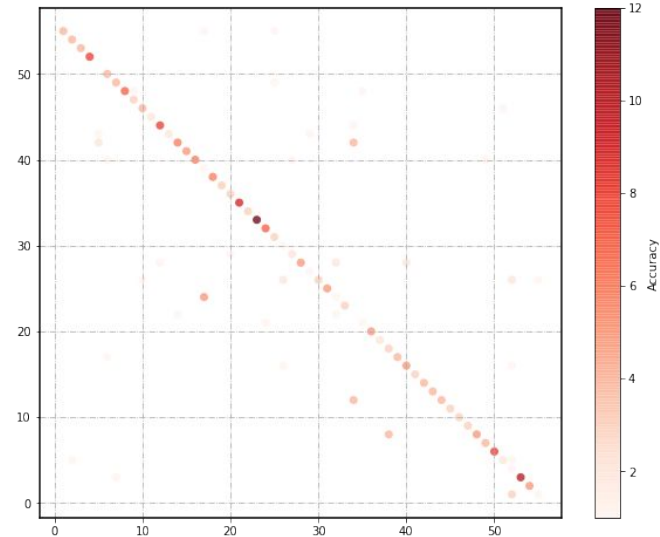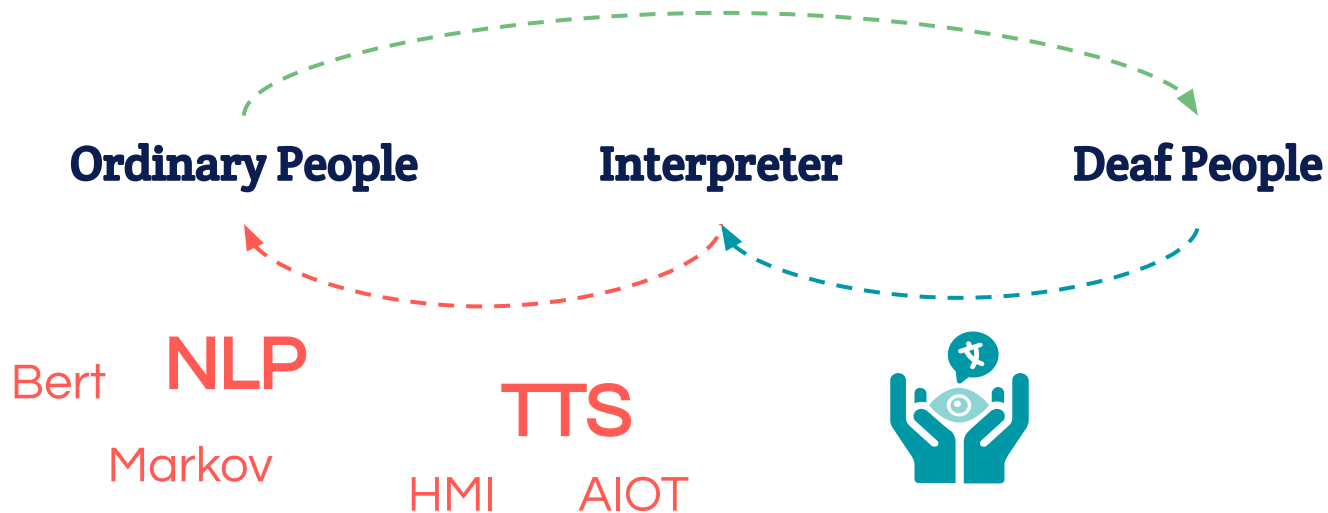
Thanks!

Taiwan Sign Language Translation

王舜婷 林于湘 林妘鑫 洪睿謙 陳郁婷 張育禔 楊畯棋 賴藍平