

# WLASL Recognition

01418364 Practical Deep Learning

# MEMBER



นาย ศรัณย์ วงศ์คำ 6410451415



นาย ถิรวัฒน์ พงศ์ปฏิสนธิ 6410450958



นายธนภักดิ์ เชื้อโตหลวง 6410401060

# Outline

- 
- ◆ 01 Introduction
  - ◆ 02 Dataset
  - ◆ 03 Model
  - ◆ 04 Evaluation  
and  
Results
  - ◆ 05 Duplicating  
Project

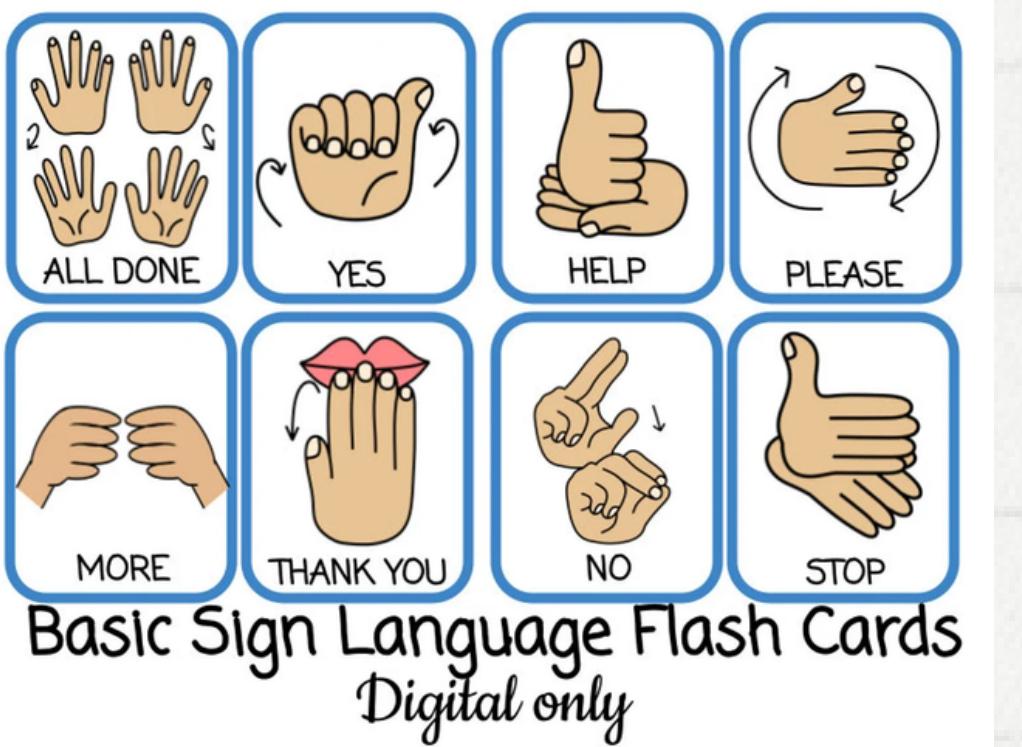
# Introduction

- American Sign Language (ASL) เป็นภาษาที่ใช้ร่างกายในการสื่อสารเป็นก้าทางแทนคำพูด
- มีกลุ่มคนจำนวนมากที่ไม่เข้าใจภาษามือ
- การสื่อสารระหว่าง signer และ non-signer เป็นเรื่องยาก



# Problem

- จำแนกสิ่งที่ผู้ใช้ภาษาเมื่อต้องการชื่อสารได้อย่างถูกต้อง



# Benefits



**Signers**

more easily communicate with non-signers



**Non-Signers**

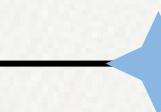
understand what signers are trying to convey

# Outline



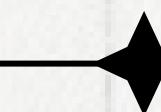
**01**

Introduction



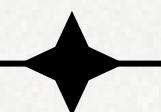
**02**

Dataset



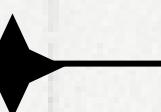
**03**

Model



**04**

Results



**05**

Duplicating  
Project

# Data Collections

## • paperswithcode

The screenshot shows the paperswithcode website interface. At the top, there is a navigation bar with links for "Search", "Browse State-of-the-Art", "Datasets", "Methods", and "More". Below the navigation bar, a section titled "WLASL (Word-Level American Sign Language)" is displayed. It includes a brief introduction by Li et al. in "Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison". It states that WLASL is a large video dataset for Word-Level American Sign Language (ASL) recognition, featuring 2,000 common different words in ASL. A "Homepage" button is visible. In the bottom left, there is a "Benchmarks" section with a table comparing datasets based on Trend, Task, Dataset Variant, Best Model, Paper, and Code. The table includes rows for "Sign Language Recognition" with variants WLASL-2000, WLASL100, and WLASL, and best models NLA-SLR, SignBERT, and SignBERT+. To the right of the benchmarks, there is a "Usage" chart showing the number of papers citing the dataset from 2020 to 2024, with WLASL being the most cited. There is also a "License" section indicating "Unknown".

## • kaggle

The screenshot shows the kaggle website interface. The left sidebar has a "Datasets" tab selected. The main content area displays the "WLASL (World Level American Sign Language) Video" dataset. It features a thumbnail image of two people signing. Below the thumbnail, it says "12k processed videos of Word-Level American Sign Language glossary performance". There are links for "Data Card", "Code (15)", "Discussion (1)", and "Suggestions (0)". On the right side, there are sections for "About Dataset", "Context", "Content", and "Acknowledgements". The "Context" section notes that WLASL is the largest video dataset for Word-Level ASL recognition, containing 2,000 words. The "Content" section describes the structure of the dataset, mentioning a "WLASL\_v0.3.json" file and approximately 12k videos per video\_id. The "Tags" section includes "Classification" and "Computer Vision".

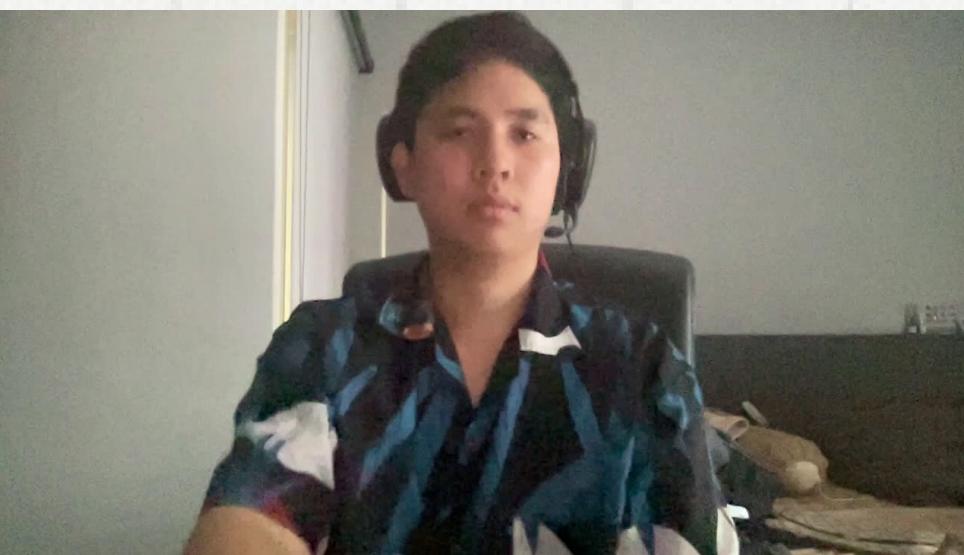
# Data Collections

(cont.)

- yes



- time



- what



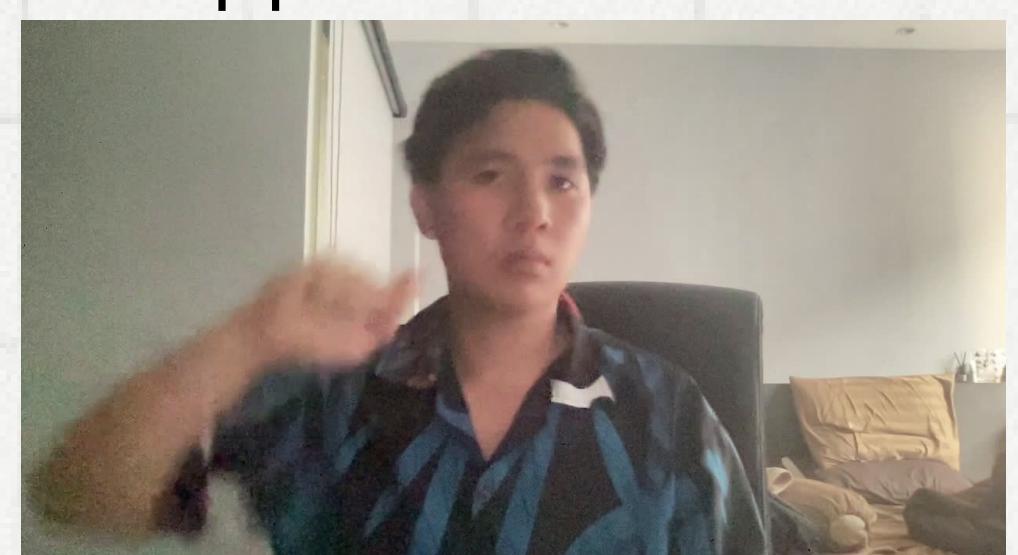
- done



- eat



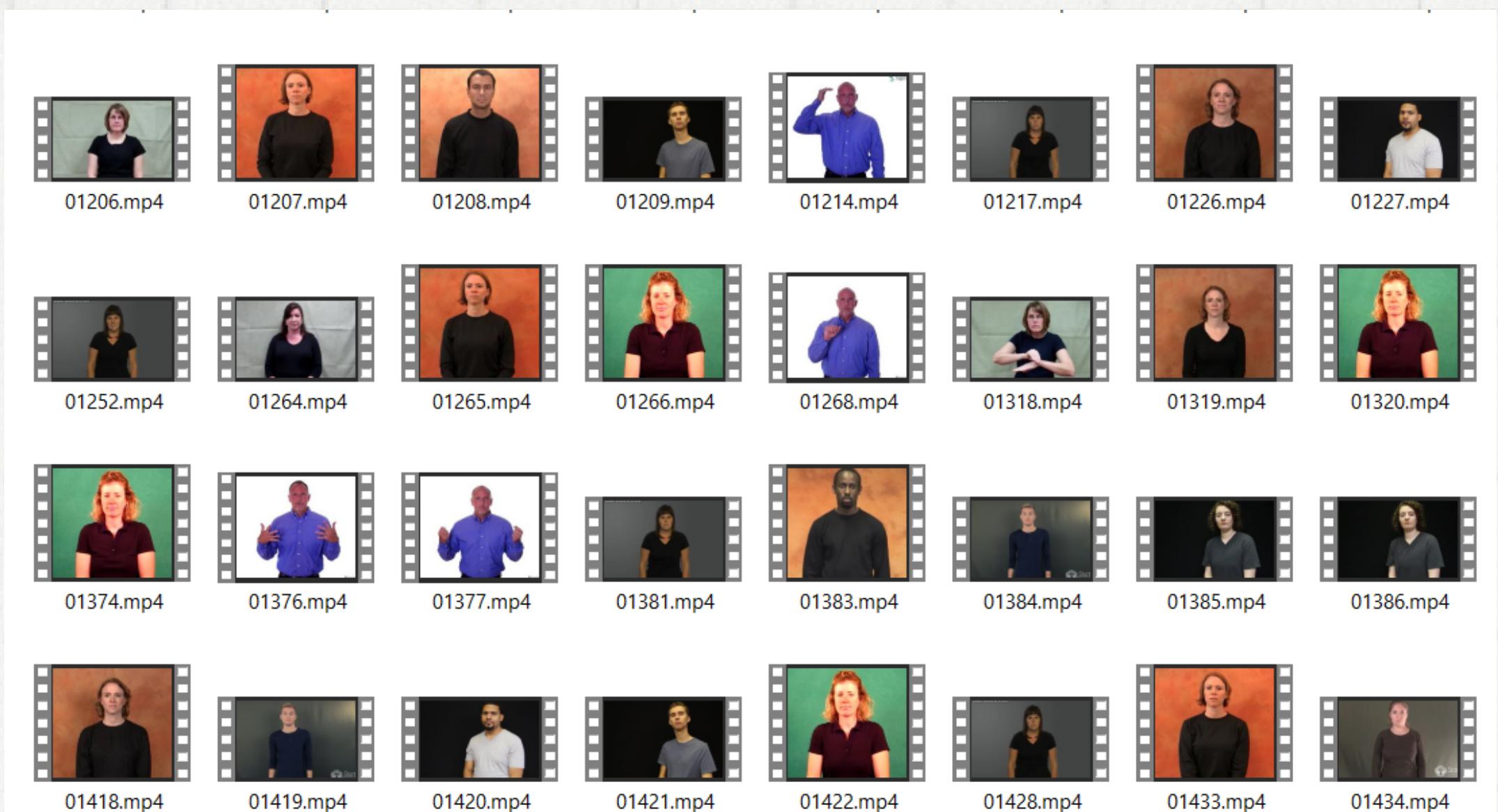
- apple



# Dataset Descriptions



- ประกอบด้วยวิดีโอ กึ่งหนด 200 วิดีโอ
- แต่ละวิดีโอเป็นการสื่อสารด้วยภาษา มือ สำหรับ 1 คำ ความยาวประมาณ 3-5 วินาที
- โดยแต่ละคำจะมีคลิปวิดีโออย่างน้อย 8 วิดีโอที่สื่อถึงคำนั้นๆ



# Dataset Descriptions (cont.)

- ไฟล์ json ที่มีรายละเอียดจาก WLASL เพื่อจำแนกรายการของคำศัพท์
  - gloss** : คำศัพท์ภาษาอังกฤษ
  - video\_path** : ที่อยู่ของวิดีโอภาษาอังกฤษซึ่งเป็นเส้นทางไฟล์ที่ซึ่งมีวิดีโอภาษาอังกฤษนั้น
  - frame\_start** : เฟรมแรกที่มีการเก็บข้อมูลคำศัพท์ในวิดีโอ
  - frame\_end** : เฟรมสุดท้ายที่มีการเก็บข้อมูลในวิดีโอ โดยค่า -1 หมายถึงไม่มีการระบุเฟรมสุดท้าย

```
{  
    "gloss": "drink",  
    "video_path": "videos\\69302.mp4",  
    "frame_start": 1,  
    "frame_end": -1  
},  
,  
{  
    "gloss": "drink",  
    "video_path": "videos\\65539.mp4",  

```

# Dataset Descriptions

(cont.)

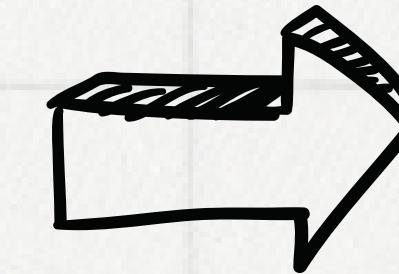
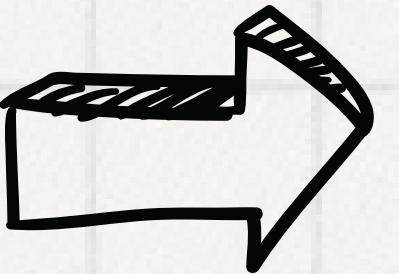
word list

<b>my</b>	<b>you</b>	<b>yes</b>	<b>no</b>
<b>help</b>	<b>fine</b>	<b>eat</b>	<b>name</b>
<b>water</b>	<b>time</b>	<b>what</b>	<b>apple</b>
<b>wrong</b>	<b>done</b>	<b>drink</b>	<b>hello</b>



# Preprocessing

- สำหรับแต่ละเฟรมในวิดีโอ สร้าง landmark กึ่งหมุด 180 จุด และเก็บตำแหน่งของแต่ละจุดไว้



(0.12359459, 126459789)  
(0.9815621, 0.4597511)

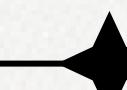
...

# Outline



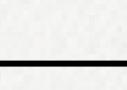
**01**

Introduction



**02**

Dataset



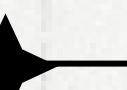
**03**

Model



**04**

Results



**05**

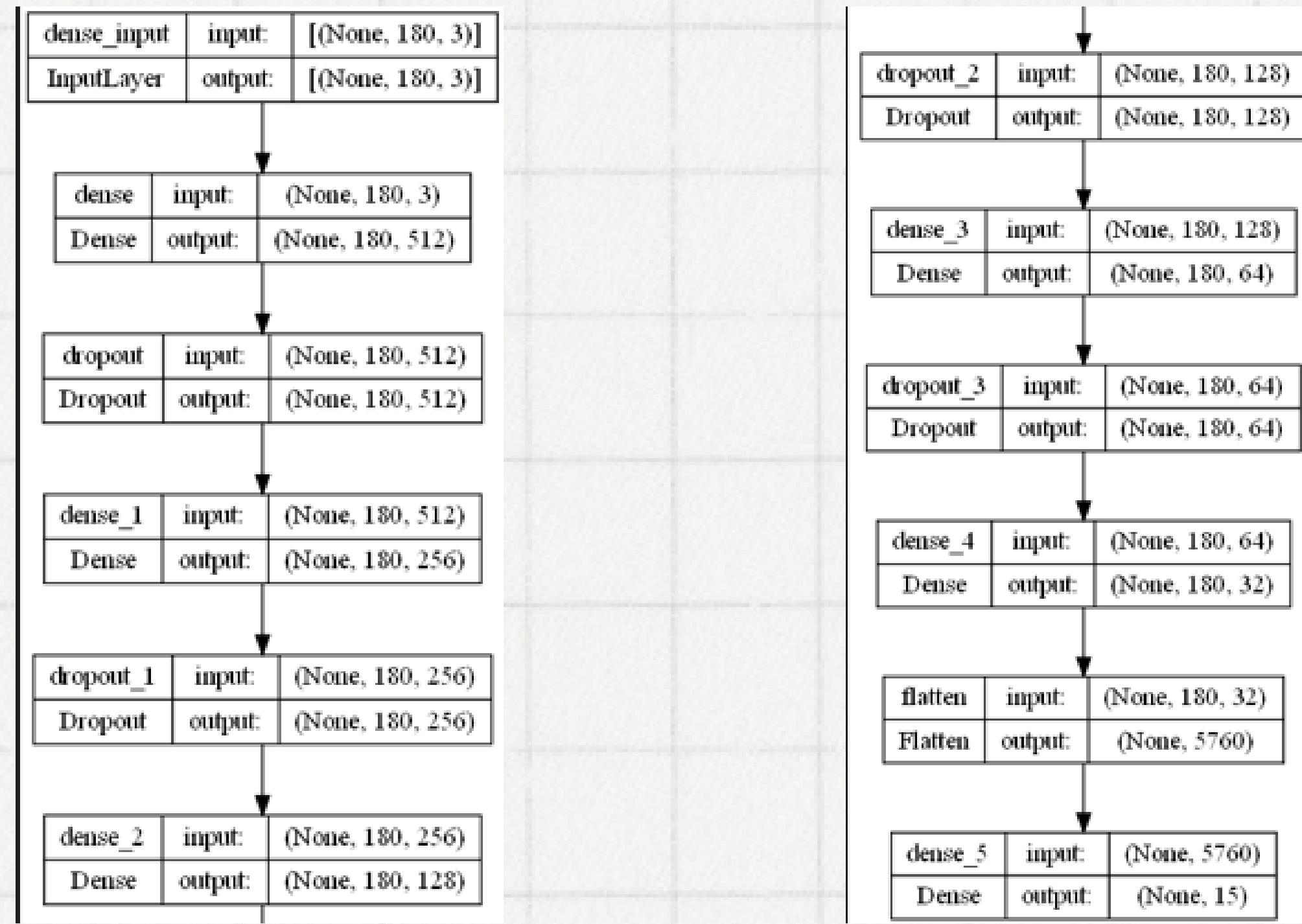
Duplicating  
Project

# Model Selection

- Assumption: ทุก frame ที่อยู่ในคลิปของคำหนึ่ง ถือว่าเป็นก้าของคำนั้น
- ไม่คำนึงถึงความต่อเนื่องของข้อมูลแต่ละเฟรม
- Selected Model : MLP  
(Multilayer Perceptron)



# Model Structure



# Outline

◆ 01

Introduction

◆ 02

Dataset

◆ 03

Model

◆ 04

Results

◆ 05

Duplicating  
Project

# Model Evaluation

1. การวัดประสิทธิภาพโดยการทดลองนำ instance บางส่วน  
ไปที่ถูกแบ่งเป็น test set มาคำนายเพื่อหา accuracy

```
In [60]: # Evaluate the model on test data
print('Test Loss: {:.6f}, Accuracy: {:.6f}'.format(*model.evaluate(X_test, y_test, verbose=0)))
model_evaluation_history = model.evaluate(X_test, y_test, verbose=0)
```

```
Test Loss: 0.615403, Accuracy: 0.797504
```



# Model Evaluation (cont.)

2). วัดประสิทธิภาพโดยแสดงก่าทางแบบ real-time

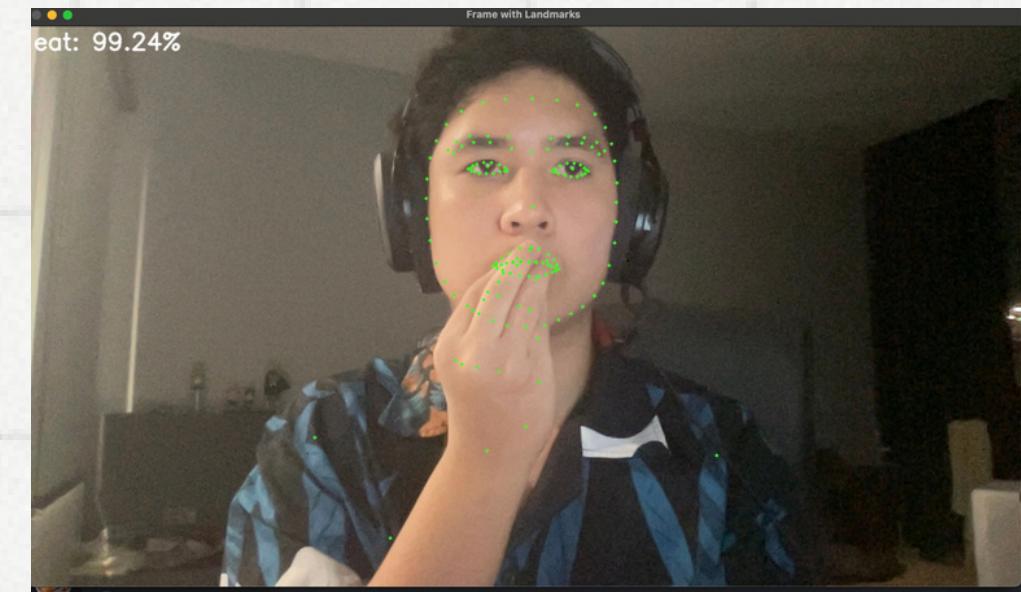


คำตอบที่ได้ และ Probability ของก่านั้นๆที่ปราฏ  
อยู่ในการจับภาพแบบ real-time



# 70%

## Accuracy on Test Set



# Limitation

- อุปกรณ์ที่ใช้ในการถ่าย/รับภาพ
- ความไม่เสถียรของแหล่งข้อมูล
- จำนวนคำที่ model สามารถแปลภาษาได้



# Outline

◆ 01

Introduction

◆ 02

Dataset

◆ 03

Model

◆ 04

Evaluation  
and  
Results

◆ 05

Duplicating  
Project

# Gathering Dataset



[github : Simple-Word-Level-American-Sign-Language-Recognition](https://github.com/zoneul/Word-Level-American-Sign-Language-Recognition)

You now have push access to the zoneul/Word-Level-American-Sign-Language-Recognition repository.

**Word-Level-American-Sign-Language-Recognition** Public

main 1 Branch 0 Tags

Go to file Add file Code

zoneul Add code and relevant data a178d5f · 4 hours ago 2 Commits

.ipynb\_checkpoints Add code and relevant data 4 hours ago

Model Add code and relevant data 4 hours ago

landmarks Add code and relevant data 4 hours ago

videos Add code and relevant data 4 hours ago

README.md first commit 4 hours ago

WLASL\_data.json Add code and relevant data 4 hours ago

wlasl\_model.ipynb Add code and relevant data 4 hours ago

wlasl\_recognition.ipynb Add code and relevant data 4 hours ago

wordlist.txt Add code and relevant data 4 hours ago

**README**

Simple-Word-Level-American-Sign-Language-...

About

No description, website, or topics provided.

Readme Activity 0 stars 1 watching 0 forks Report repository

Releases

No releases published Create a new release

Packages

No packages published Publish your first package

Contributors 2

zonntp Zon zoneul

# Continuing on our model

- Run 'wlasl\_recognition.ipynb' for real-time recognition of asl
- Continue Develop model in 'wlasl\_model.ipynb'



**Thank you !**

