

# ASL Recognition and Production Update

**Pr. Mohamed Bekkali**

**Noureddine Ech-chouky    Med Amine Bouzmoul**

**Module :**

**Intelligence Artificielle 2**

**Filière :**

**Ingénierie Numérique En Data Science Et Intelligence Artificielle**

**Année Universitaire :**

**2021 | 2022**

## Table des matières

Chapter 1 : Introduction	3
Chapter 2 : Sentence-Level Approaches	4
About LSTM :	4
Chapter 3 : Our Aiming Goals	6

## Chapter 1 : Introduction

Things we have achieved : Word-Level Sign Language Recognition

For now we have achieved an accurate, and well performing word-level SL Recognition.

### Main Approach :

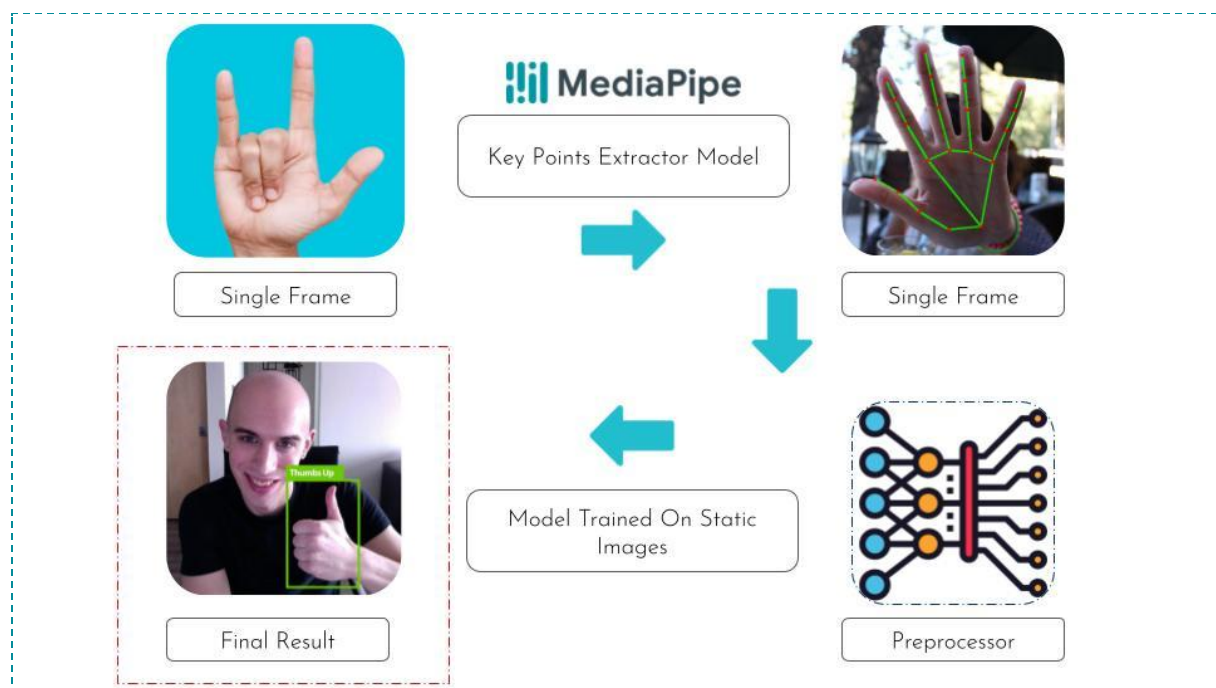


Figure 1.1 : Word-Level Sign Language Approach

## Chapter 2 : Sentence-Level Approaches

In Research for the best approach

### Problem Statement :

In the opposite of the word-level Sign Language Recognition, a sentence in sign language is done in multiple frames, so the first approach is not going to work here.

### First Approach : LSTM Model

#### About LSTM :

In simple terms, LSTM or Long Short Term Memory Model, with its Architecture shown below, figure 2.1, uses a series of 'gates' which control how the information in a sequence of data comes into, is stored in and leaves the network.

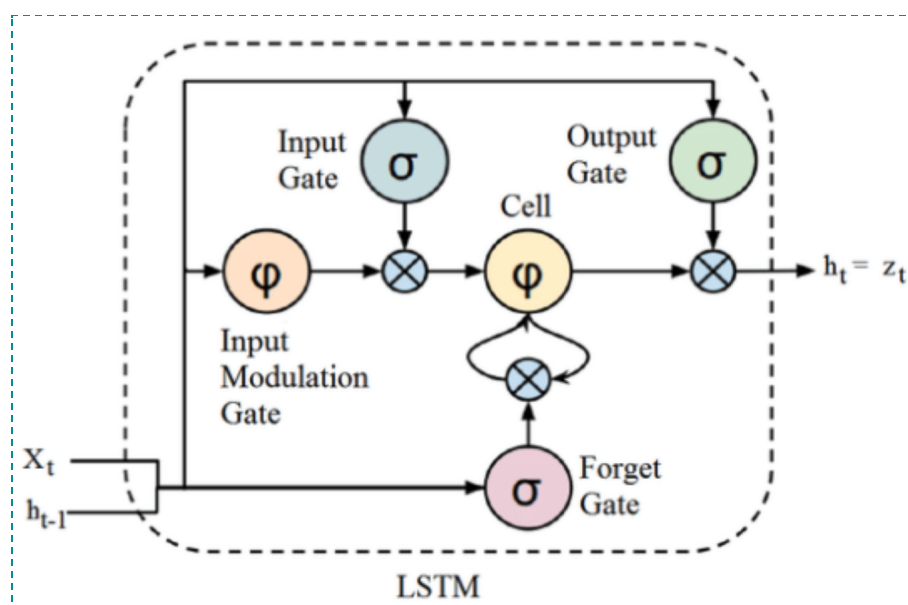


Figure 2.1 : LSTM Architecture

So, Using LSTM Model, we have established an accurate Sign Language Recognition System, and are ready to train it on sentence level videos dataset, But the problem we are facing with LSTM is **latency**. Taking for example the situation below :



Figure 2.2 : Latency Problem In The LSTM Mod

So i have shotted two videos concerning the **"Testing LSTM Model"**:

- **The first one : LSTM using the last 30 frame :**

**Video Drive link :**

[https://drive.google.com/file/d/1X0iypeDUqC7k7cBM\\_GcH4IHULqa\\_4J49/view?usp=sharing](https://drive.google.com/file/d/1X0iypeDUqC7k7cBM_GcH4IHULqa_4J49/view?usp=sharing)

- **The Second video : LSTM using the last 10 frame :**

**Video Drive link :**

[https://drive.google.com/file/d/1mQQsatmORI3H5GU-W\\_ODL9M7Ud-NQDOe/view?usp=sharing](https://drive.google.com/file/d/1mQQsatmORI3H5GU-W_ODL9M7Ud-NQDOe/view?usp=sharing)

**Note :**

The Sentence level datasets have an average of ~120 frames per label (Sentence), So the LSTM approach would perform very slowly.

## Chapter 3 : Our Aiming Goals

Find Better Approaches

- Develop a Robust ConvoLSTM Model approach
- Try Research Based Approaches (Video2Gloss, Gloss2English using Machine Translation Model) : Transformers.
- Redevelop the word-level System with a large american Dataset (Universal Language ), because we have done it previously using Indian SL dataset.