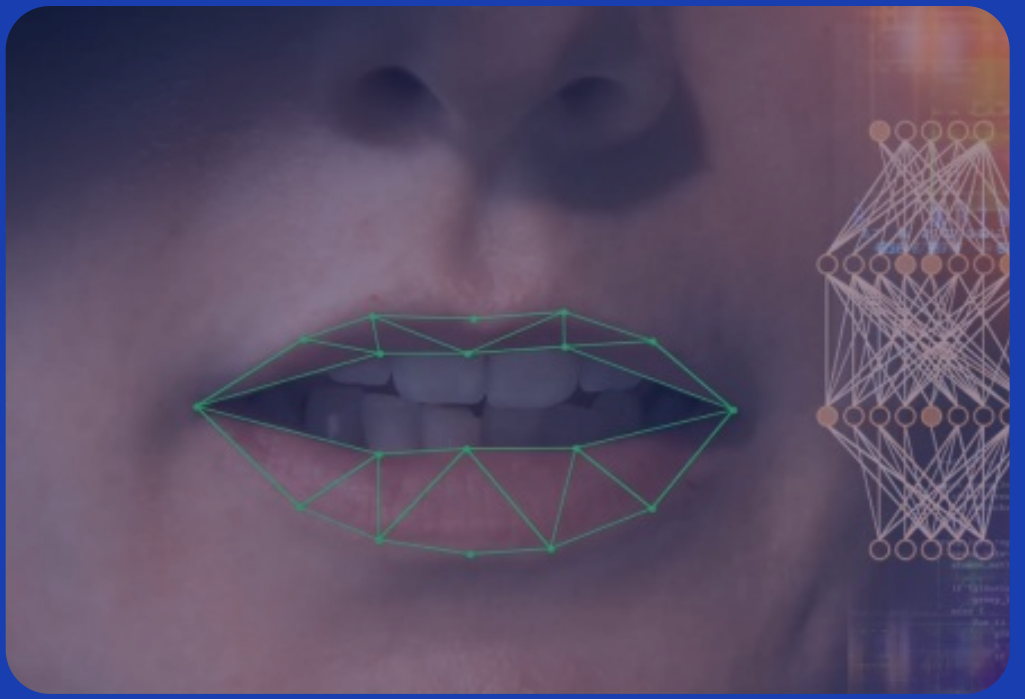
SAMSUNG

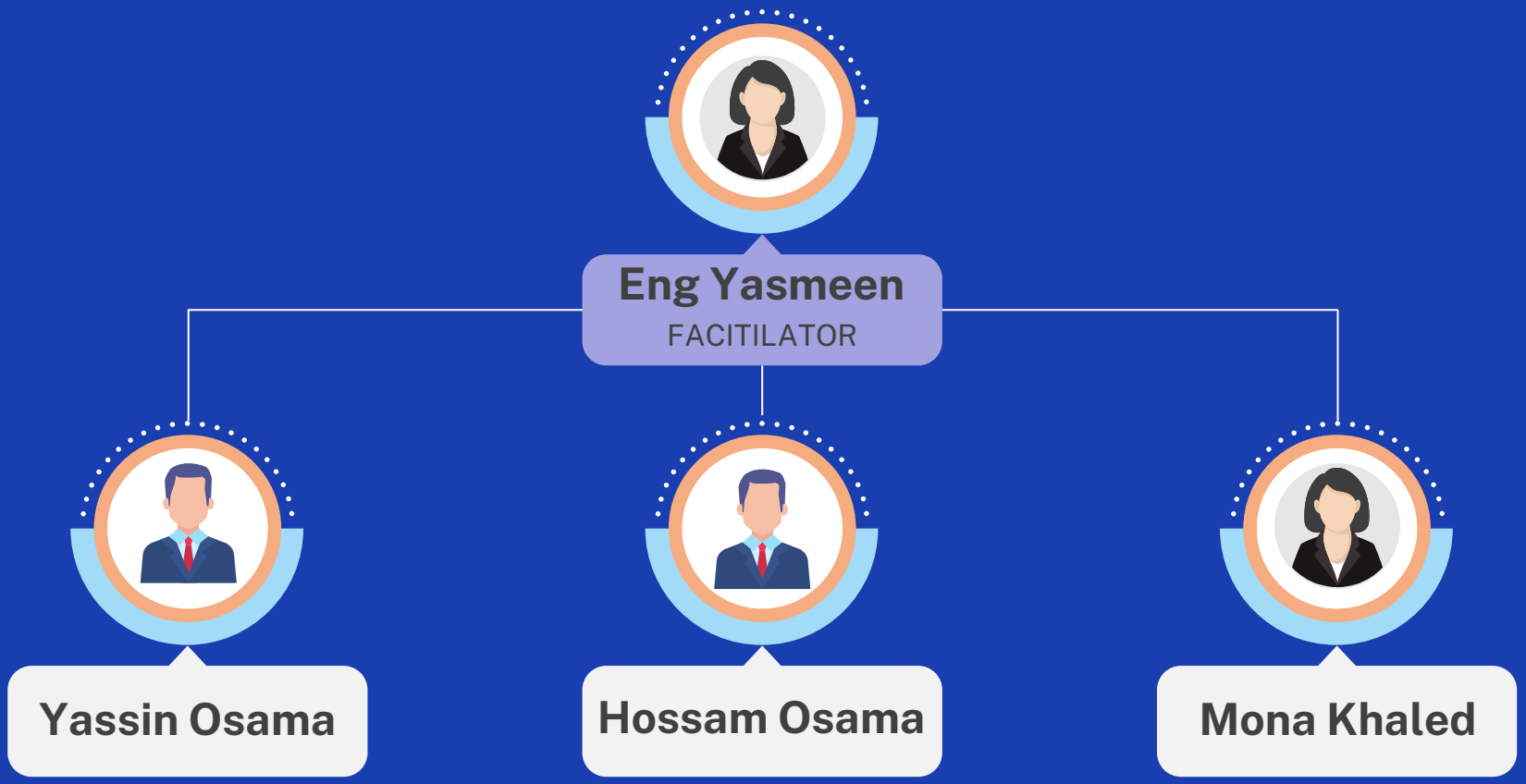
Together for Tomorrow!
Enabling People
Education for Future Generations

Lip Reading Project

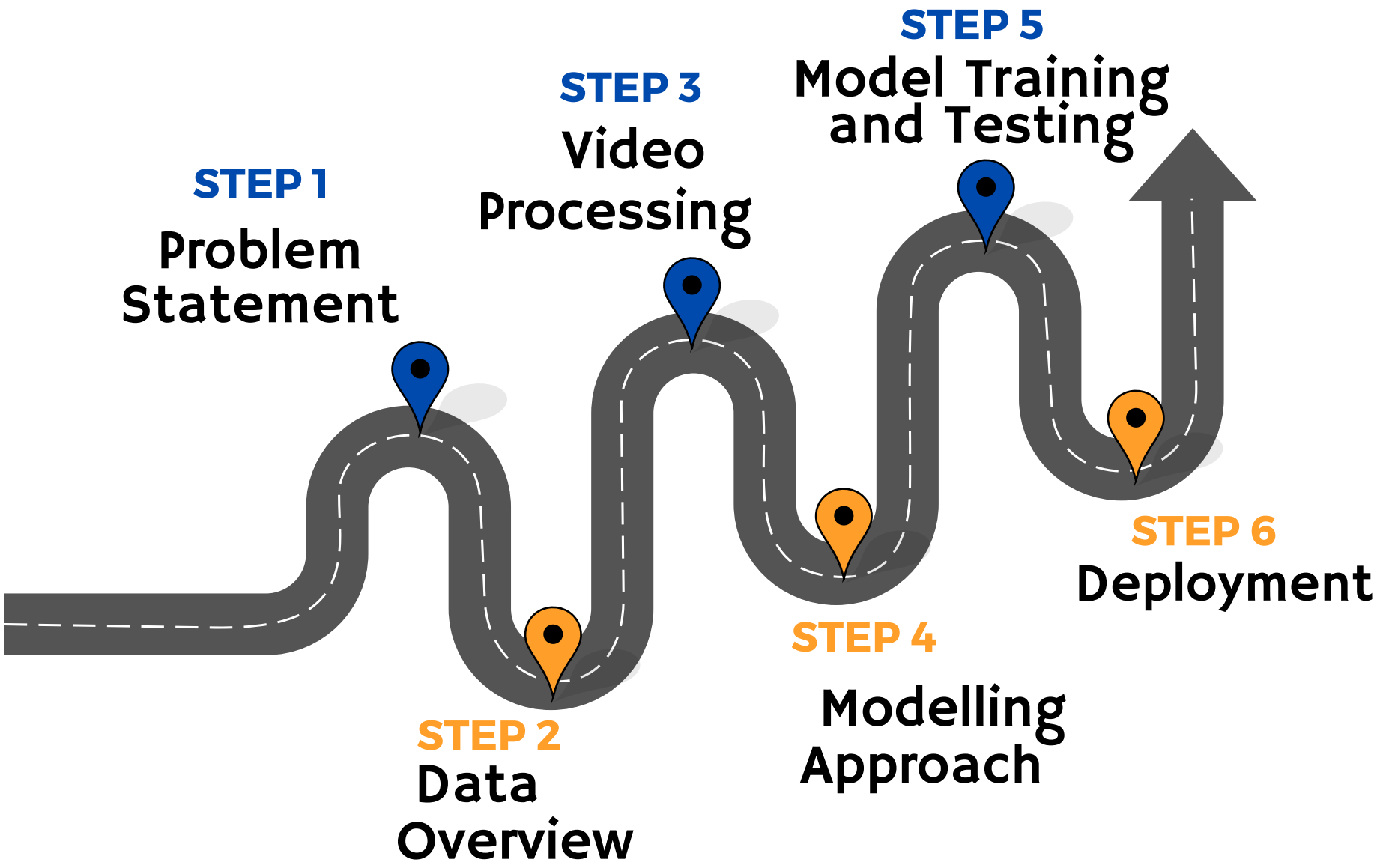
SIC 6



Meet Our Team



OUR AGENDA


STEP 1
Problem Statement
STEP 2
Data Overview
STEP 3
Video Processing
STEP 4
Modelling Approach
STEP 5
Model Training and Testing
STEP 6
Deployment

Problem Statement

Problem Statement

Lip-to-Text Conversion

The core problem we aim to address is the conversion of visual information, specifically lip movements, into accurate text. Lip reading, or speechreading, involves understanding speech by observing the movements of the lips.

Problem Statement

Lip-to-Text Conversio

Objective: Translate visual lip movements into accurate text

- **Enhance communication accessibility for individuals with hearing impairments.**
- **Improve interactions in sound-sensitive environments by enabling silent speech recognition**

Data Overview

Data Overview

GRID Corpus dataset

- Speakers: 34 unique speakers
- Sentences: Each speaker delivers 1,000 short sentences
- Total Sentences: 34,000 audio and video recordings
- Data Types:
- Align and video recordings
- Each sentence has both audio and video data
    -

Video Processing Pipeline

Video Processing Pipeline

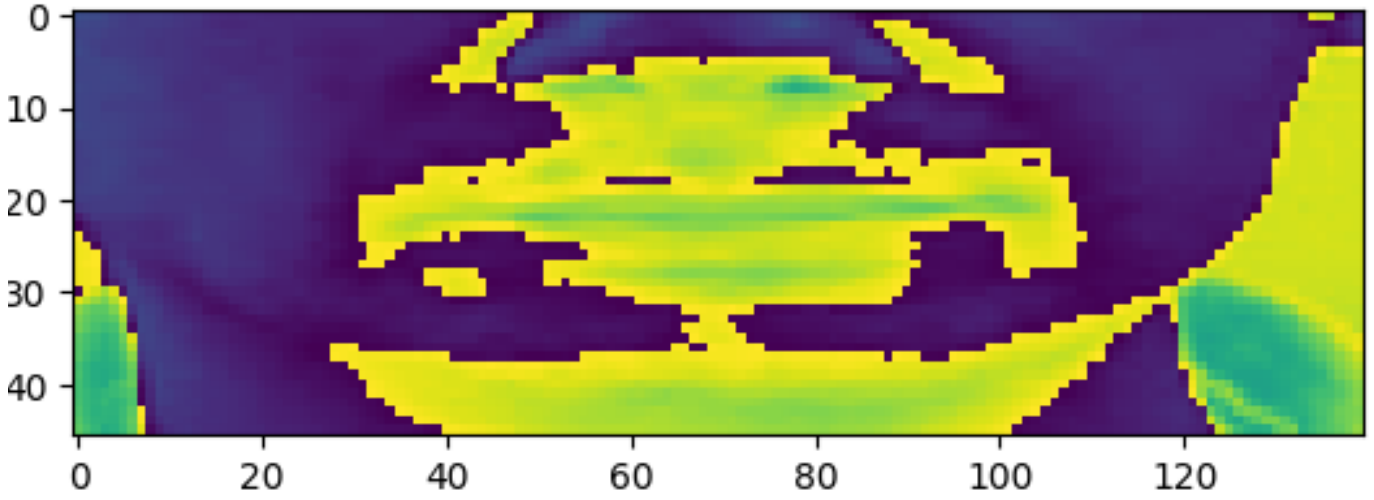
Face Detection and Landmark Prediction

- Tools: Using dlib for facial landmark detection.
- Landmark Detection:
  - shape_predictor_68_face_landmarks.dat pretrained model to detect 68 facial landmarks.
  - Specifically, the mouth region is defined by landmarks 48 to 61.
- A bounding box is created around the mouth, with added padding to ensure full coverage of the lips.

Video Processing Pipeline

Video Frame Processing

- **Frame Extraction: captures frames from the video, clips the mouth region from each frame, and resizes the mouth region to a fixed size (140x46).**
- **Grayscale Conversion: Each clipped frame is converted to grayscale and normalized.**



Video Processing Pipeline

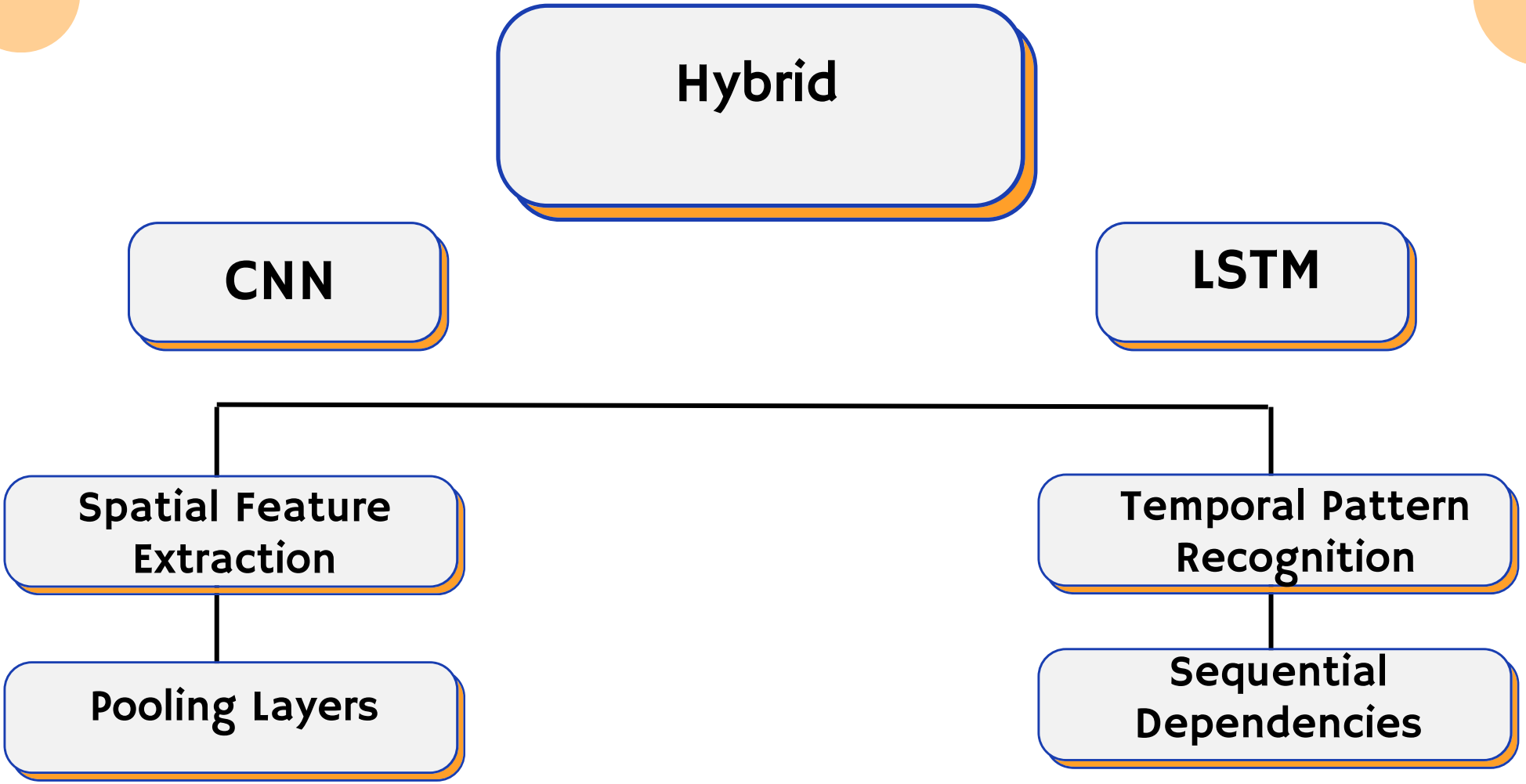
Character Tokenization and Mapping

- Input : text from align files
- Take the characters as a tokens
- Mapping the tokens into numerical tokens.

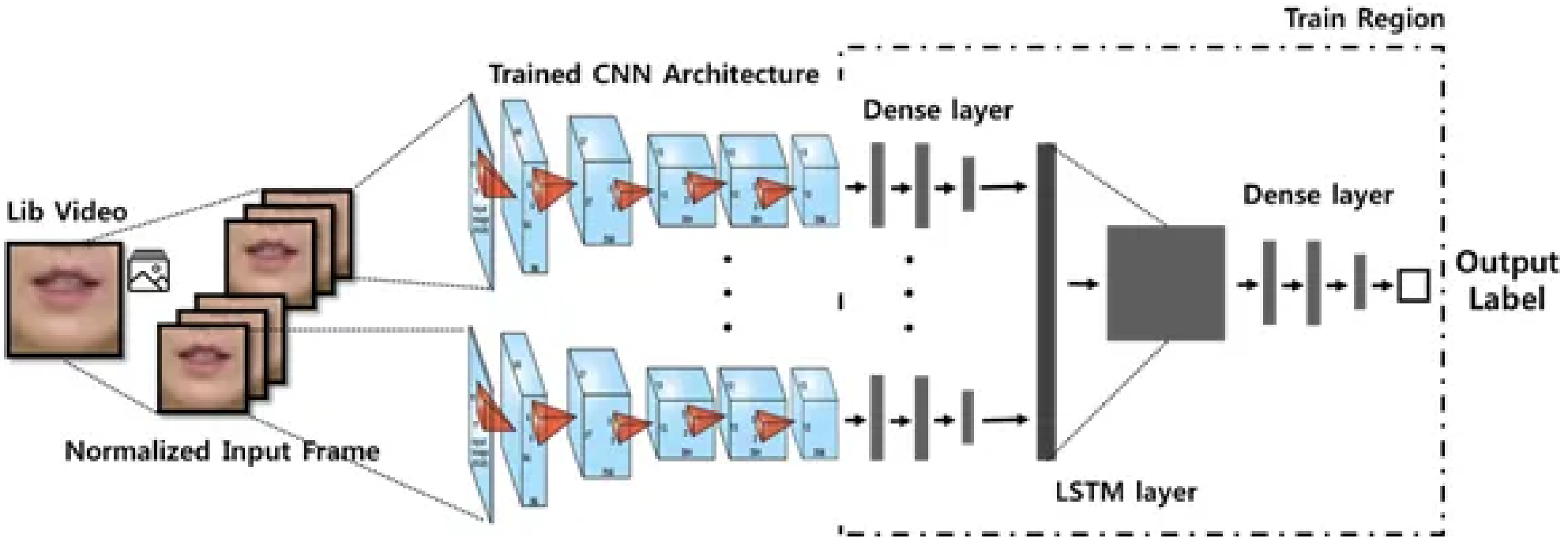
': 0, 'a': 1, 'b': 2, 'c': 3, 'd': 4, 'e': 5, 'f': 6, 'g': 7, 'h': 8, 'i': 9, 'j': 10, 'k': 11, 'l': 12, 'm': 13, 'n': 14, 'o': 15, 'p': 16

Modelling Approach

Model Architecture



Model Training

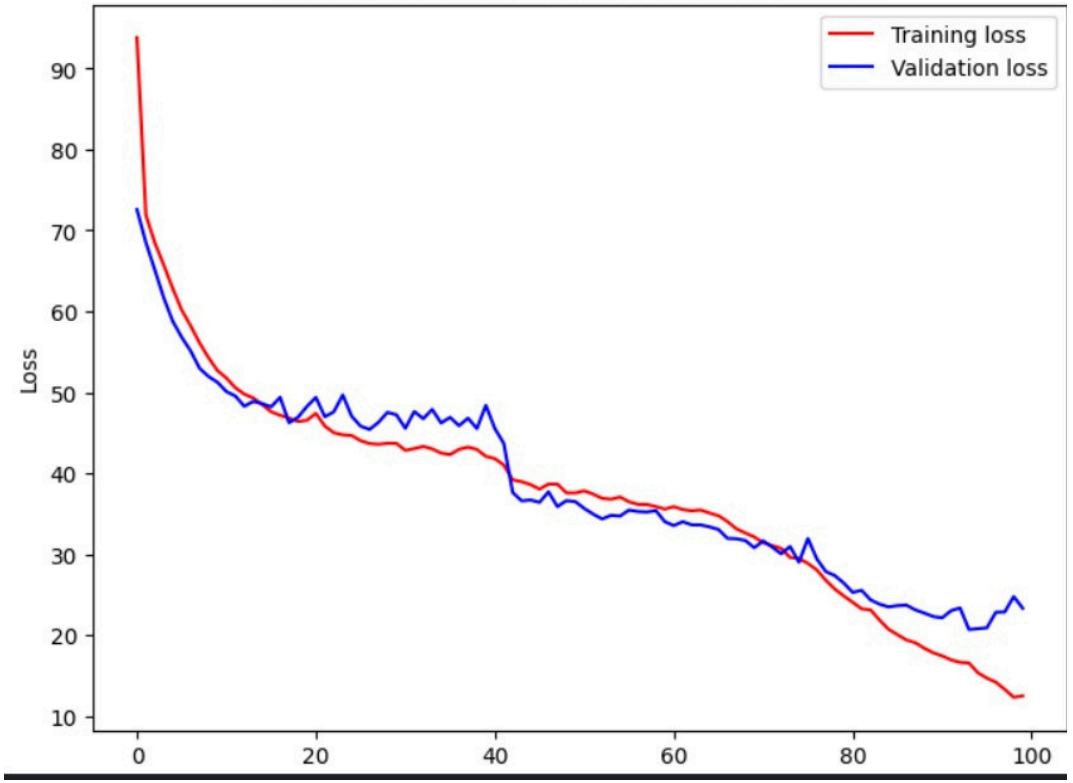


- We use the Adam optimizer, a widely used optimization algorithm in deep learning.

GUESS OUR LOSS Function?

CTC LOSS

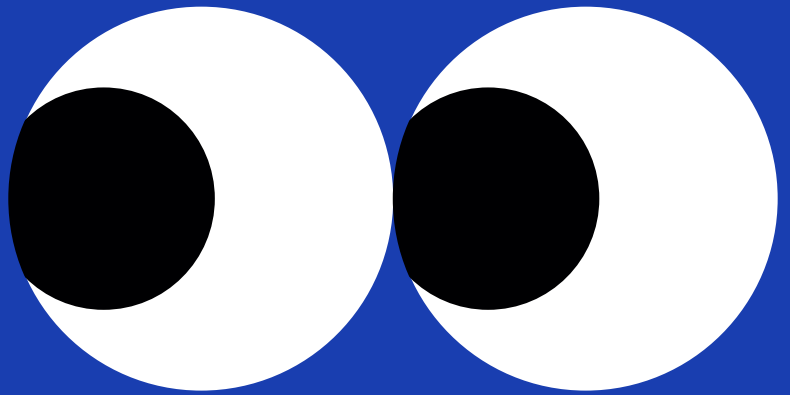
WHY IS CTC LOSS OFTEN PREFERRED OVER CROSS ENTROPY LOSS ?

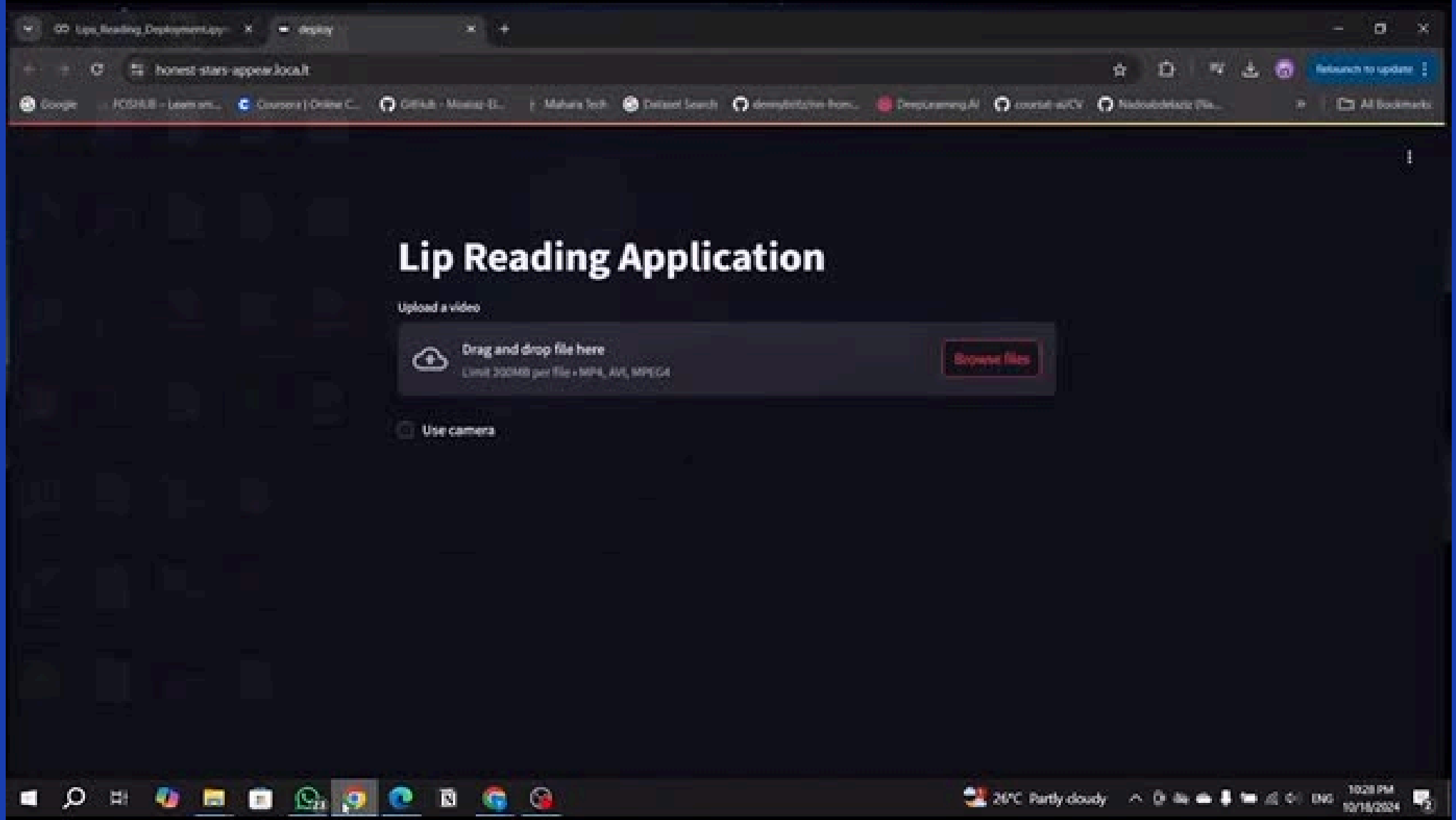


it accommodates variable-length input and output sequences without requiring explicit alignment.

Deployment

Let’s watch....




honest-stars-appear.loca.lt
Lip Reading Application
Upload a video
Drag and drop file here
Browse files
Use camera
26°C Partly cloudy
ENG
10:28 PM
10/18/2024

Thank You