

Group 4 - Project Proposal

1. What problem did you select and why did you select it?

Visual speech recognition (VSR), commonly referred to as lipreading, is the task of decoding spoken language solely from visual input of a speaker's mouth region. This problem is challenging due to coarticulation, subtle lip movements, and variations across speakers. Recent advances in deep learning have enabled end-to-end architectures capable of learning spatiotemporal speech patterns directly from video data.

LipNet (Assael et al., 2016) represents one of the earliest end-to-end deep learning models to achieve sentence-level lipreading

2. What database/dataset will you use? Is it large enough to train a deep network?

ECG image digitalization using Deep Learning:

<https://spandh.dcs.shef.ac.uk/gridcorpus/>

3. What deep network will you use? Will it be a standard form of the network, or will you have to customize it? Deep network choice

We are thinking about using a multi-task convolutional encoder with lightweight task-specific heads, rather than one monolithic end-to-end model. This balances accuracy, speed, and debuggability

4. What framework will you use to implement the network? Why?

Pytorch or Tensorflow, due to its flexibility for custom losses and multi-task heads, large vision ecosystem, and smooth deployment via ONNX/TensorRT; classical CV (OpenCV) and external OCR complete the pipeline.

5. What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?

Academic and Technical Papers

- 3D Convolutional Layers
- Extract spatial and temporal features using stacked Conv3D → ReLU → MaxPool3D blocks.

Bidirectional LSTM Layers

Capture temporal dependencies across the full sentence.

Dense Softmax Output Layer

Produces per-timestep character probabilities.

- CTC Loss Function
- Enables alignment-free mapping between input video frames and output text sequences.

6. How will you judge the performance of the network? What metrics will you use?
Segmentation and Detection Metrics, Precision, Recall, and F1-score, CER and WER
7. Provide a rough schedule for completing the project.
Week 1,2 - Preprocessing the data
Week 3 - Model Building
Week 4 - App building and Report making