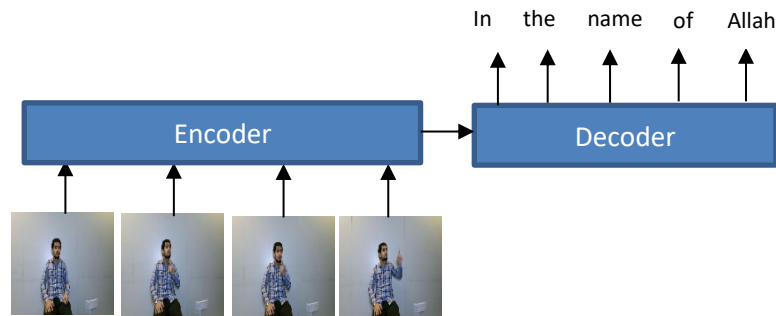


Homework 4

Date: 27/11/2021

Due date: 11/12/2021

Sign language recognition is the task of recognizing the sign performed by deaf people. Sign gestures can be isolated (one sign per video) or continuous (several signs per video). In this assignment, you will develop a deep learning model for sign language recognition at the sentence level. This task is similar to the video captioning task where the input will be a sign video and the output will be the performed sign(s) by the signer in the text format. An example of the input and output of the system is shown below.



In this assignment you need to do the following:

- [10 points] Read and prepare the dataset
 - o The dataset is provided as images ([access link](#))
 - o There are 10 different sentences performed by three signers
 - o Each video sample is already normalized to 80 frames.
- [20 points] Use word embedding to represent the ground truth text when you feed it to the model
- [30 points] Develop an **encoder-decoder model** to recognize the sign video (video captioning)
 - o Select the best architecture that gives good results
- [30 points] Develop an **encoder-decoder model with attention** to recognize the sign video (video captioning)
- [10 points] Report the results using word error rate (WER) metric
- **Bonus [40 points]**
 - o [10 points] The first three best results (lowest WER) will get +10 points bonus
 - o [30 points] Develop a transformer model for the same problem.