Topic K – Aligning Text to Sign Language Video

Imane Elbacha & Rajae Sebai

Abstract

This paper is an introduction and plan for the final project of RecVis 2022. The chosen topic is aligning text to sign language video. The aim is to align text sentences (as subtitles) to sign language video. The dataset used is BOBSL dataset. This dataset includes about 1000 hours of poorly annotated training data and about 13 hours of fully annotated training data (text words manually aligned to sign language video) (text sentences which are approximately aligned to sign language video). There are several improvement areas using a thorough bibliography. In this report, we are going to start with a description of the topic, then an estimated plan and finally a general work distribution.

1. Topic description

For Deaf communities, sign languages are an essential means of communication. In actuality, sign languages are similar to other spoken languages except that they primarily communicate through the use of hands, body posture, and facial expressions. The purpose of this project is to convert the hand signs used in continuous signing video into subtitles. The creation of such a tool could have a wide range of uses, such as the indexing of sign language video corpora and the automatic generation of massive sign language data sets.

The process of aligning subtitles to continuous signing can be difficult. First off, the grammatical structures of sign languages differ significantly from those of spoken languages. Second, because to variations in speed and syntax, a subtitle's length varies greatly between speech and signing. Third, there is no direct one-to-one mapping between subtitle words and signs created by interpreters, and whole subtitles may not be signed.

To sum up, the goal of this project is to develop an algorithm that successfully assigns text to signing video. The tools are a comprehensive bibliography, the BOBSL dataset and the methods presented in class.

2. Plan

As a starting point, a first objective would be to fairly understand the details of the method presented in [1] and to assess its limits. A next step would be to reproduce the results obtained in the original paper using the BOBSL dataset presented in [2]. We will be using the code of the original paper for a starting pipeline. According to the authors of [1], the entire training procedure of the original paper took less than 24h on a single Tesla P42 GPU. The pretrained model is available on the official GitHub repository of the paper. We might be using it as a base model since we might not be able to reproduce the results using the same resources. Once we get comparable results to those of the original paper, we will work on one of the suggested improvements, which consists of taking more context into account by trying to align three subtitles instead of only one subtitle. The following steps would be to qualitatively and quantitatively evaluating the results and plotting the error cases in order to get an idea of what is improvable.

3. Work distribution

Overall, our project consists of two main tasks: to reproduce the original paper results and to implement an improvement to the original method. In order to collaborate effectively, we will divide each task into elementary subtasks that we will assign iteratively during the project. The idea is to work separately on technical tasks and jointly on comprehension and design tasks. As a collaboration tool, we will be working with branches and sharing the project content on a main GitHub repository.

Globally, the team will use a collaborative work strategy to ensure a coherent and rigorous detail oriented outcome while also guarantying an in depth understanding of each step of the project to everyone through exchange of resources and recurrent feedback loops.

References

[1] Hannah Bull, Triantafyllos Afouras, Gül Varol, Samuel Albanie, Liliane Momeni, and Andrew Zisserman. Aligning subtitles in sign language videos, 2021. 1

[2] Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie

Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. Bbc-oxford british sign language dataset, 2021. 1