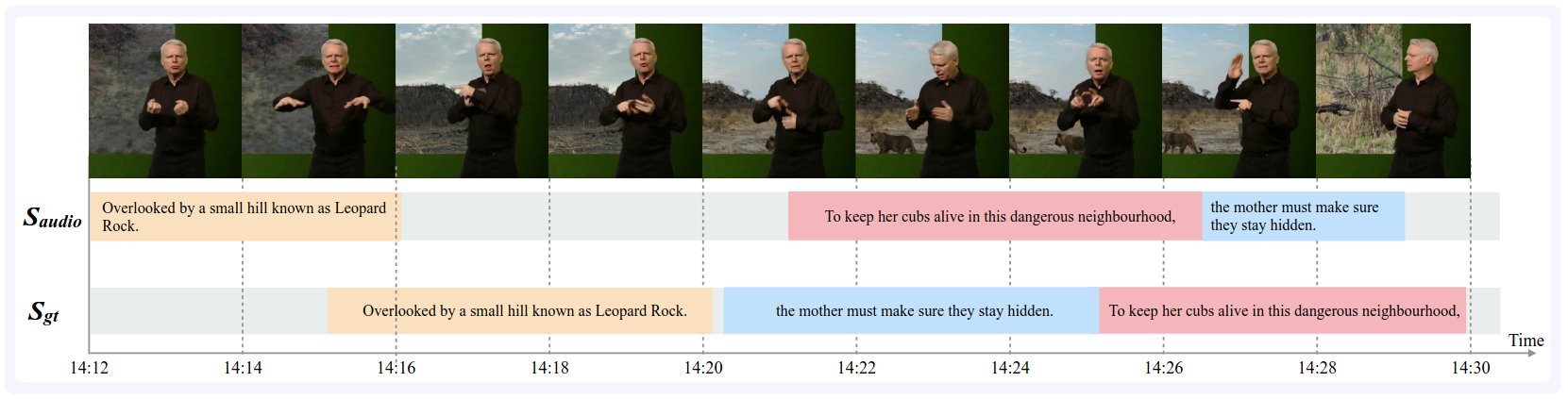
## **Topic K – Aligning Text to Sign Language Video**



**Requirements:** A GPU and familiarity with PyTorch

**Motivation:** Sign languages are languages just like French, English or Japanese, but use hands, body orientation and facial expressions as means of communication. British Sign Language (BSL) is the sign language used by the deaf community in the UK. It is a natural language like any other: it was not “invented” but rather evolved over time, it has its own grammar and is different from English, and a native BSL signer can just as easily express concepts in BSL as everyone else in their native language. Although translation systems work quite well for spoken and written languages, translation systems essentially don’t work at all for signed languages, beyond translation of very simple phrases. This project aims to work on a task even simpler than translation, yet still difficult: aligning text sentences to the correct location in sign language video.

**Description:** The task is to align text sentences (subtitles) to sign language video using the BOBSL dataset. This dataset contains around 13 hours of strongly annotated training data (text sentences which are manually aligned to sign language video), as well as over 1000 hours of weakly annotated training data (text sentences which are approximately aligned to sign language video). You can look at [1] for an idea upon which to improve.

**Dataset:** <https://www.robots.ox.ac.uk/~vgg/data/bobsl/>   
You will need to request access.

**Example project outline:**

1. Reproduce the baseline results on BOBSL from [2] using the available code [here](https://github.com/hannahbull/subtitle_align).
2. Try to improve upon these results:
   1. You could try to use additional information from automatic sign spottings, available from the project page of [3]. You could add information on the available sign spottings to the transformer-based architecture, for example by concatenating word embeddings of spottings with the associated video features in the decoder, OR
   2. You could try to take into account more context by inputting 3 consecutive subtitles instead of 1 subtitle, and then try to align all three subtitles at once.
3. Evaluate results to see whether or not there is an improvement
4. Visualise error cases

[1] Bull, H., Afouras, T., Varol, G., Albanie, S., Momeni, L., & Zisserman, A. (2021). Aligning subtitles in sign language videos. In ICCV 2021.

[Paper](https://openaccess.thecvf.com/content/ICCV2021/papers/Bull_Aligning_Subtitles_in_Sign_Language_Videos_ICCV_2021_paper.pdf) | [Project page](https://www.robots.ox.ac.uk/~vgg/research/bslalign/) | [Code](https://github.com/hannahbull/subtitle_align)

[2] Albanie, S., Varol, G., Momeni, L., Bull, H., Afouras, T., Chowdhury, H., Fox, N., Woll, B., Cooper, R., McParland, A. and Zisserman, A., 2021. BBC-Oxford British Sign Language Dataset. *arXiv preprint arXiv:2111.03635*.

[Paper](https://arxiv.org/abs/2111.03635) | [Project page](https://www.robots.ox.ac.uk/~vgg/data/bobsl/)

[3] Momeni, L., Bull, H., Prajwal, K. R., Albanie, S., Varol, G., & Zisserman, A. (2022). Automatic dense annotation of large-vocabulary sign language videos. In ECCV 2022.

[Paper](https://arxiv.org/abs/2208.02802) | [Project page](https://www.robots.ox.ac.uk/~vgg/research/bsldensify/)