Weakly-supervised Fingerspelling Recognition in British Sign Language Videos

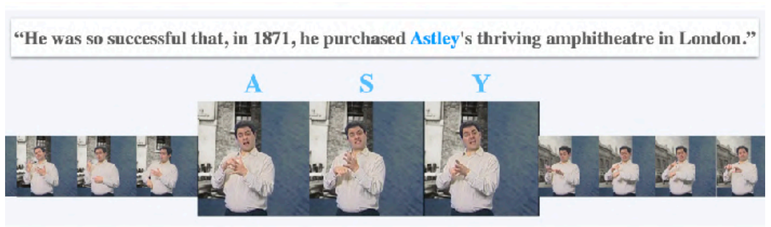
K R Prajwal[1*], Hannah Bull[2*], Liliane Momeni[1*], Samuel Albanie[3], Gül Varol[4], Andrew Zisserman[1]

[1] VGG, University of Oxford, UK [2] LISN, Univ Paris-Saclay, France [3] University of Cambridge, UK [4] IMAGINE, École des Ponts ParisTech, France

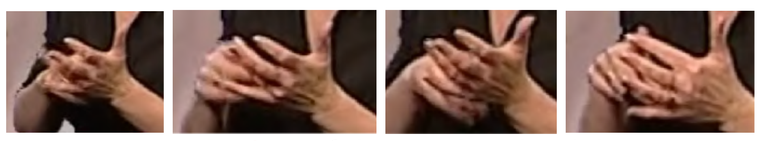
https://www.robots.ox.ac.uk/~vgg/research/transpeller/

(1) INTRODUCTION

Our goal is to **detect** and **recognise** fingerspelled sequence of letters in **continuous** British Sign Language videos.

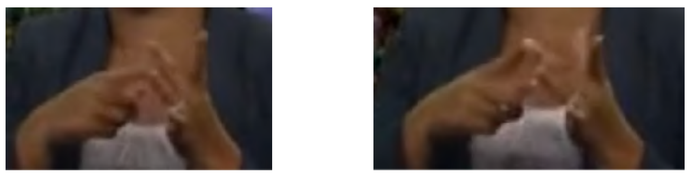


(2) CHALLENGES



Occlusion of the two hands while fingerspelling "**love**"

Partial fingerspelling: Signers fingerspell short forms or skip letters. E.g. NN is finger-spelt here for Nottingham.

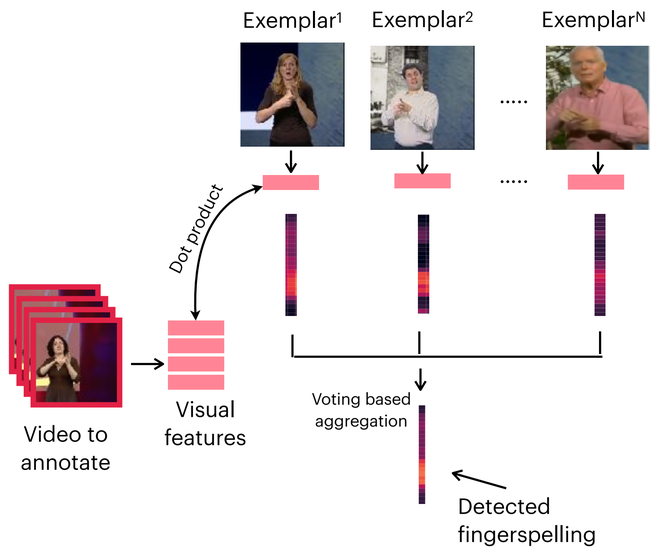


Fine-grained perceptual differences
E.g. letters **a** and **i** are shown here

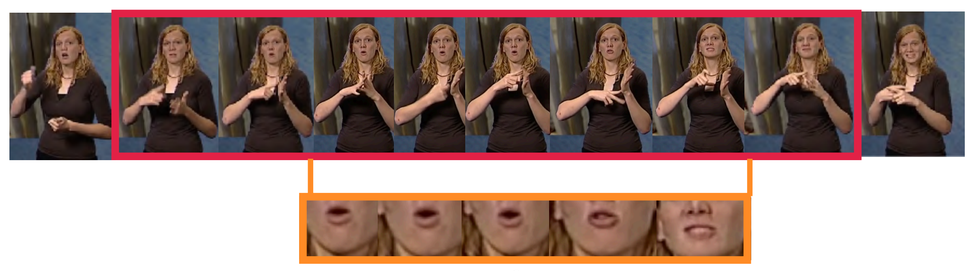
No fingerspelling datasets in British Sign Language!

- **Challenging** – requires expert annotators
- **Time-consuming** to label each letter
- **Not scalable**.

(3) AUTOMATICALLY CURATING TRAINING DATA: (i) DETECTION, (ii) RECOGNITION



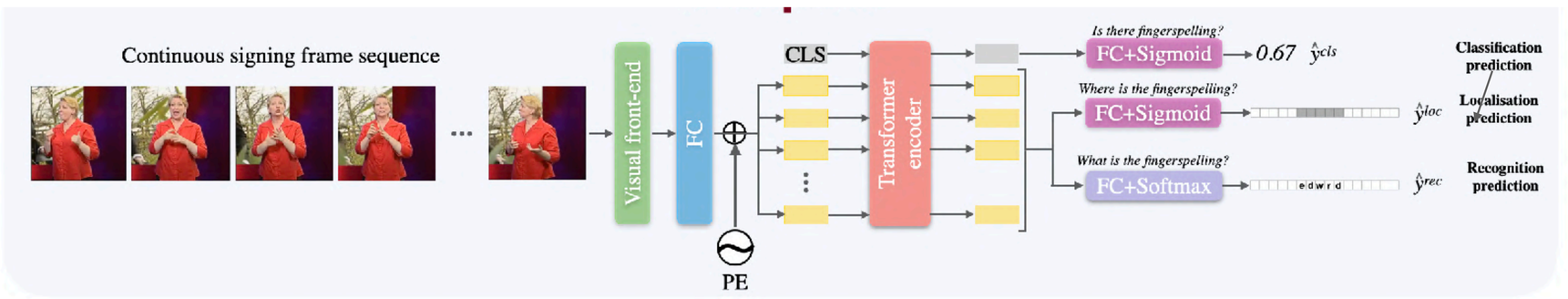
Detected Fingerspelling instance



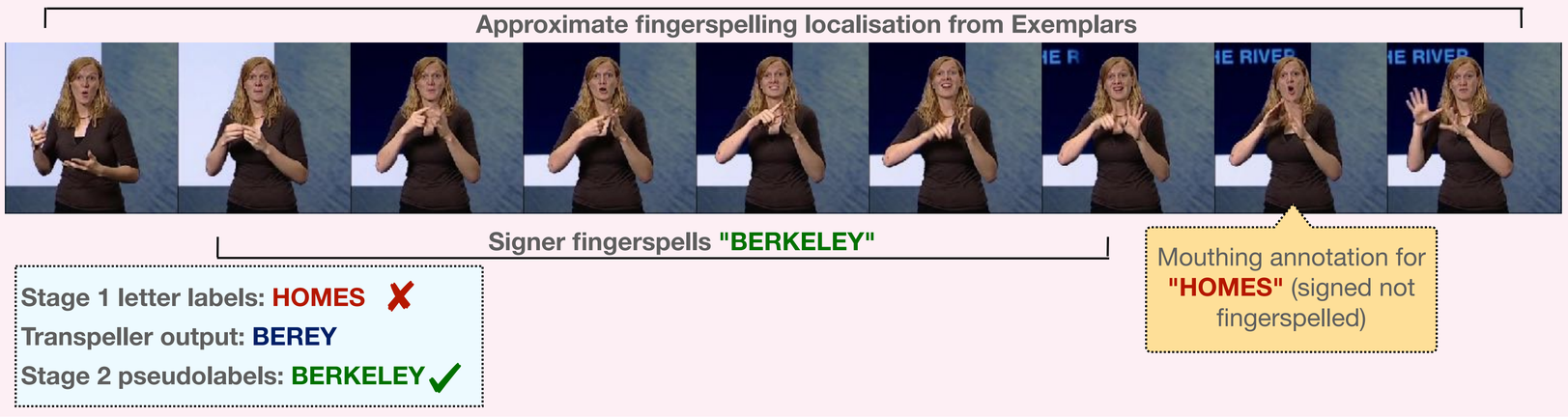
Mouthing of "**Lawrence**"

We automatically create an initial training set of about 59K training clips by:
(i) using exemplar-based techniques to detect instances of fingerspelling,
(ii) associating word labels with the help of a visual keyword spotting[2] model.

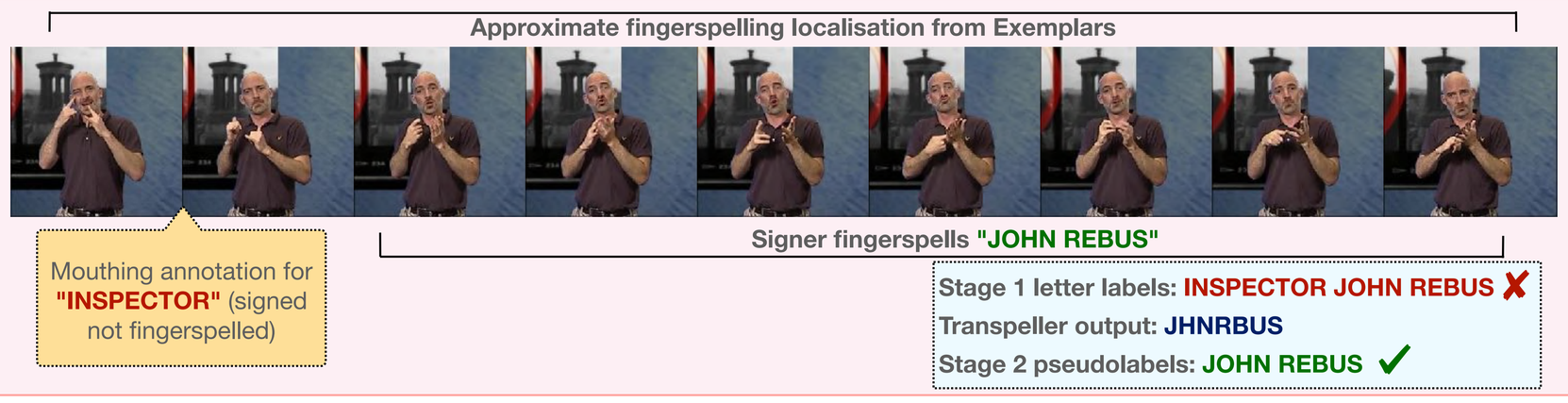
(4) TRANSPELLER



(5) PSEUDOLABELING to expand and enrich the training data

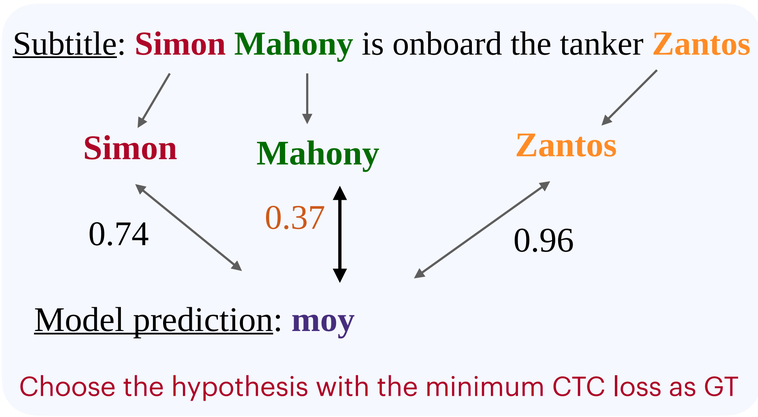


Subtitle: "This area will be developed quickly now that **Berkeley** Homes have got their foot on this corner."



Subtitle: "For much of that time, I've been writing stories about the murder investigations of my fictional detective, Inspector **John Rebus**."

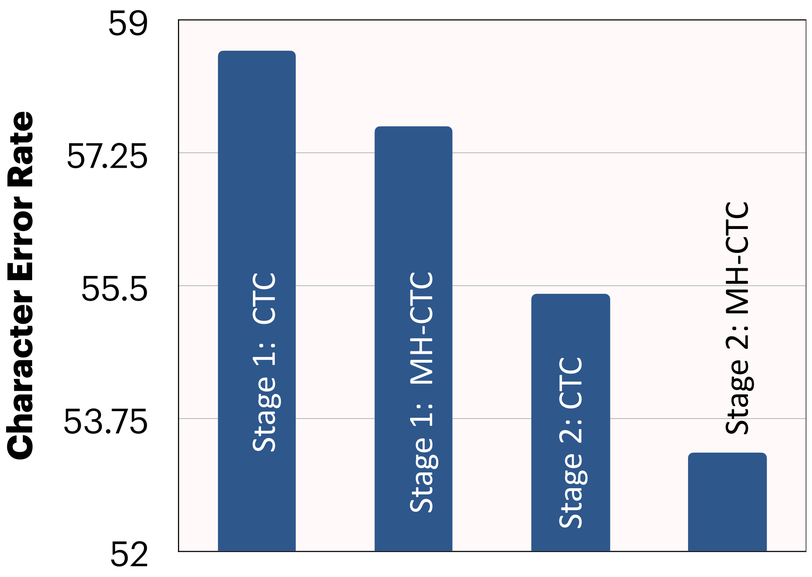
(6) MULTIPLE HYPOTHESES CTC LOSS

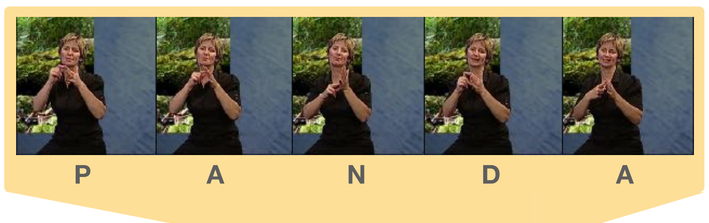


Choose the hypothesis with the minimum CTC loss as GT

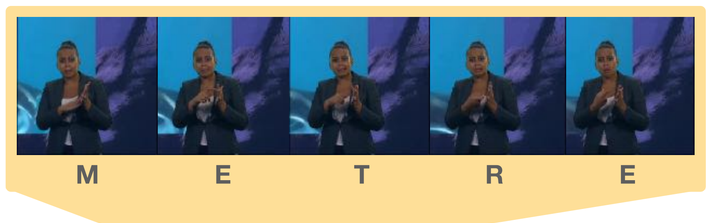
RESULTS

~ 5K verified fingerspelling clips
From the **BOBSL[1] test set**





Predicted: PENDA
Ground Truth: PANDA



Predicted: METER
Ground Truth: METRE

Limitations and future directions:
- How do we supervise partially fingerspelt letters?
- Hard negatives to suppress false fingerspelling detections
- Generalising to unseen words during inference

[1] BBC-Oxford British Sign Language Dataset, Albanie et al., arXiv preprint arXiv:2111.03635, 2021
[2] Visual Keyword Spotting with Attention, Prajwal et al., BMVC 2021