(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-spelled here for Nottingham.
- Fine-grained perceptual differences E.g. letters a and i are shown here
- Challenging — requires expert annotators
- Time-consuming to label each letter
- Not scalable.

(6) MULTIPLE HYPOTHESES CTC LOSS

Subtitle: Simon Mahony is onboard the tanker Zantos

\[
\begin{align*}
0.74 & \quad \text{Simon} \\
0.37 & \quad \text{Mahony} \\
0.96 & \quad \text{Zantos}
\end{align*}
\]

Model prediction: may

Choose the hypothesis with the minimum CTC loss as GT

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

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.

(5) PSEUDOLABELING to expand and enrich the training data

Approximate fingerspelling localisation from Exemplars

Signer fingerspells “BERKELEY”

Stage 1 letter labels: HOMES
Transpeller output: METER
Stage 2 pseudolabels: BERKELEY

Mouthing annotation for “HOMES” (signed not fingerspelled)

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

Approximate fingerspelling localisation from Exemplars

Signer fingerspells “JOHN REBUS”

Mouthing annotation for “INSPECTOR” (signed not fingerspelled)

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

RESULTS

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

<table>
<thead>
<tr>
<th>Character Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
</tr>
<tr>
<td>57.25</td>
</tr>
<tr>
<td>55.5</td>
</tr>
<tr>
<td>53.75</td>
</tr>
<tr>
<td>52</td>
</tr>
</tbody>
</table>

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