Handshape-Aware Sign Language Recognition

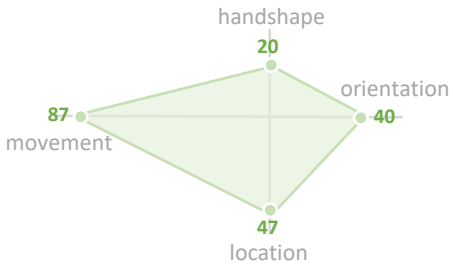
Xuan Zhang and Kevin Duh

Johns Hopkins Whiting School of Engineering

Motivation

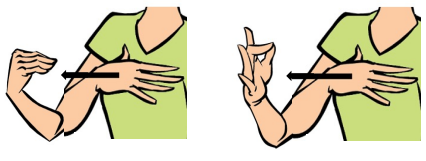
Signs are defined by 5 parameters:

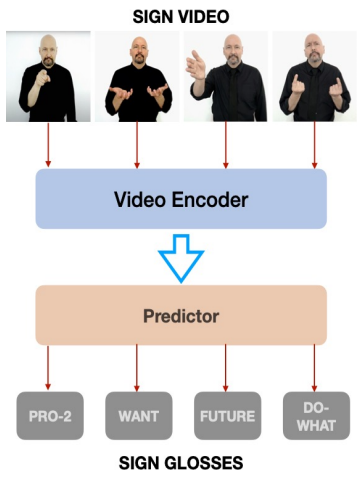
handshape, orientation, location, movement, facial expression



Human's ability to distinguish between sign minimal pairs with only handshape differences is limited. [1]

A handshape minimal pair in ASL. Difficult to distinguish when signers sign rapidly.





Sign Language Recognition (SLR) system

- video encoder:
  (1) visual encoder – spatial
  (2) sequence encoder – temporal
- predictor:
  alignment module

Challenges:

Signs are interpreted as a cohesive whole.

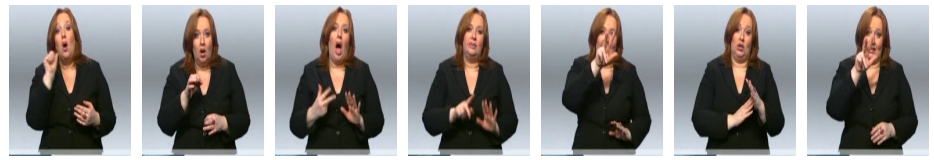
Solutions:

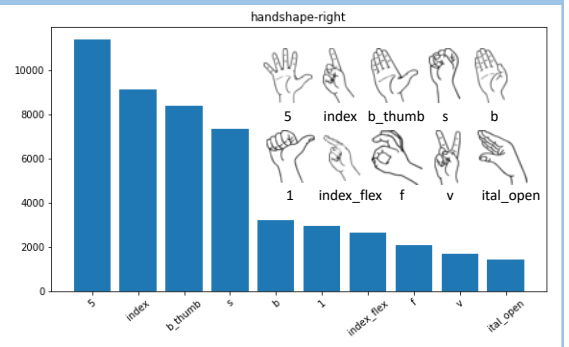
Incorporation of phonological features into SLR systems.

Contributions:

1. Extend an existing dataset with handshapes.
2. propose two handshape-inclusive SLR systems.

PHOENIX14T-HS: The Handshape-Extended Dataset





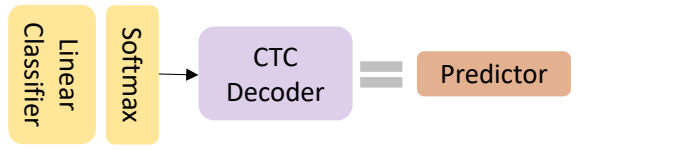
Dataset info:

German sign language for Weather forecast [2]; train/dev/test: 7,096/519/642; vocab: 1,066; average gloss length: 7.8

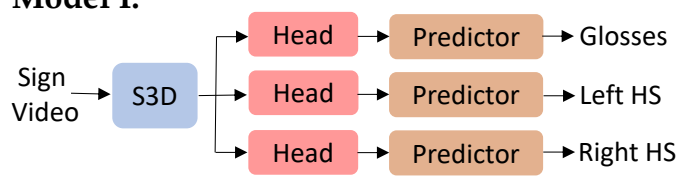
Annotation Process:

1. Look up the SignWriting dictionary (a publicly accessible, user-edited sign language dataset, with handshape annotations).
2. Manually label handshapes for signs not present in SignWriting. (1/3 vocab; the author is the annotator)

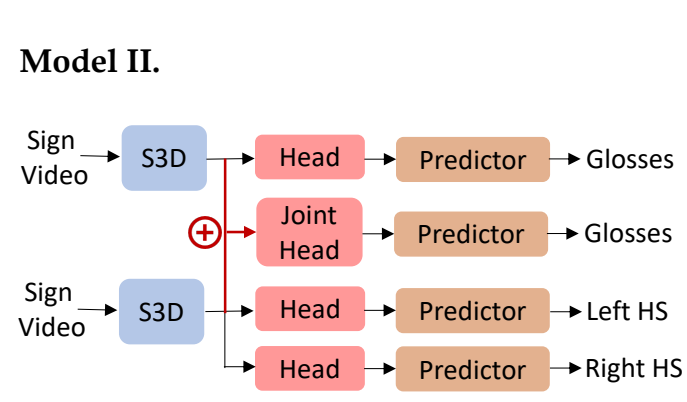
Handshape-inclusive SLR systems

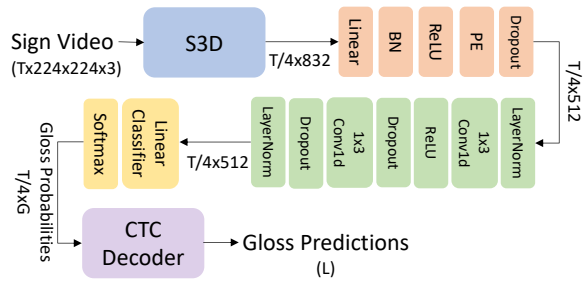


Model I.



Model II.





Baseline System [3]
(w/o handshape)

Comparison with SOTA (WER)

Method	Dev	Test
CNN-LSTM (Koller et al., 2019)*	22.1	24.1
SFL (Niu and Mak, 2020)	25.1	26.1
FCN (Cheng et al., 2020)	23.3	25.1
Joint-SLRT (Camgoz et al., 2020)	24.6	24.5
CMA (Papastratis et al., 2020)*	23.9	24.0
SignBT (Zhou et al., 2021a)	22.7	23.9
MMTLB (Chen et al., 2022a)	21.9	22.5
SMKD (Hao et al., 2021)	20.8	22.4
HS-SLR(ours)	**20.3**	**21.8**
STMC-R (Zhou et al., 2021b)*	19.6	21.0
C^2SLR (Zuo and Mak, 2022)*	20.5	20.4
TwoStream (Chen et al., 2022b)*	17.7	19.3
Single-modality	Multi-modality	

Conclusions from Ablation Studies

1. baseline 23.69 > *HS-SLR* 21.8 WER
2. *Model II* is better than Model I
3. incorporating *both hands* is not better than the only right hand
4. *pretraining* is helpful
5. *freezing* parts of S3D is helpful
6. Adding *CE-loss* is helpful

Conclusions

1. Handshape-inclusive SLR delivers SOTA performance.
2. It is beneficial to have handshape annotations.

Future directions: (1) multi-modal (2) contrastive learning

References

[1] Fahey, T., & Hilger, A. (2022). Impact of manual American Sign Language parameters on intelligibility. [2] Forster, J., Schmidt, C., Koller, O., Bellgardt, M., & Ney, H. (2014). Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather. In LREC. [3] Chen, Y., Wei, F., Sun, X., Wu, Z., & Lin, S. (2022). A simple multi-modality transfer learning baseline for sign language translation. In CVPR.