UNIVERSITY OF SURREY
CVSSP Centre for Vision, Speech and Signal Processing

BMVC 2025

Beyond Gloss: A Hand-Centric Framework for Gloss-Free Sign Language Translation

Sobhan Asasi - Mohamed Ilyas Lakhal - Ozge Mercanoglu Sincan - Richard Bowden

CVSSP, University of Surrey, Guildford, United Kingdom

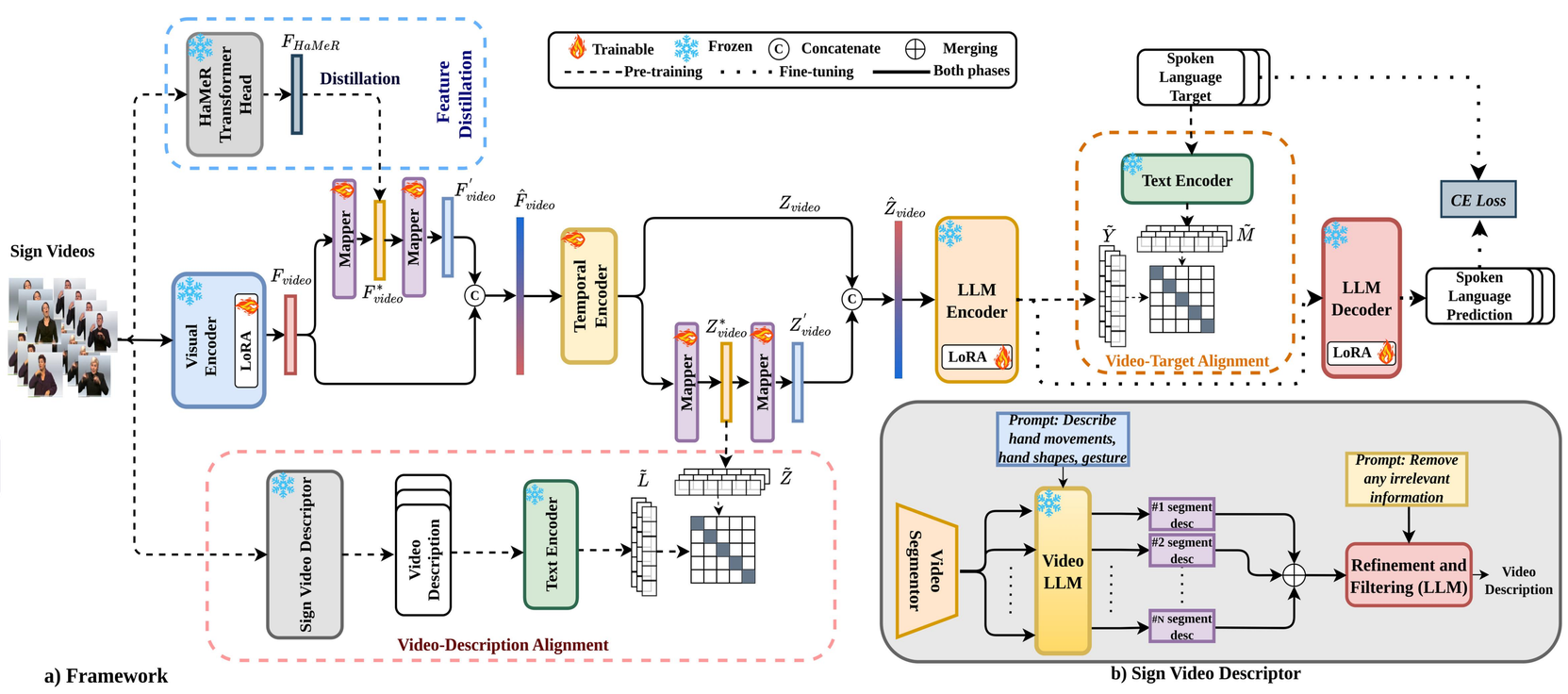
About

- **Task:** Sign Language Translation (SLT) is the process of converting visual sign language into written text to bridge communication gaps.
- **Challenge**: SLT is difficult because it requires bridging the gap between visual signs and written text and capturing very subtle hand/body and face movements.

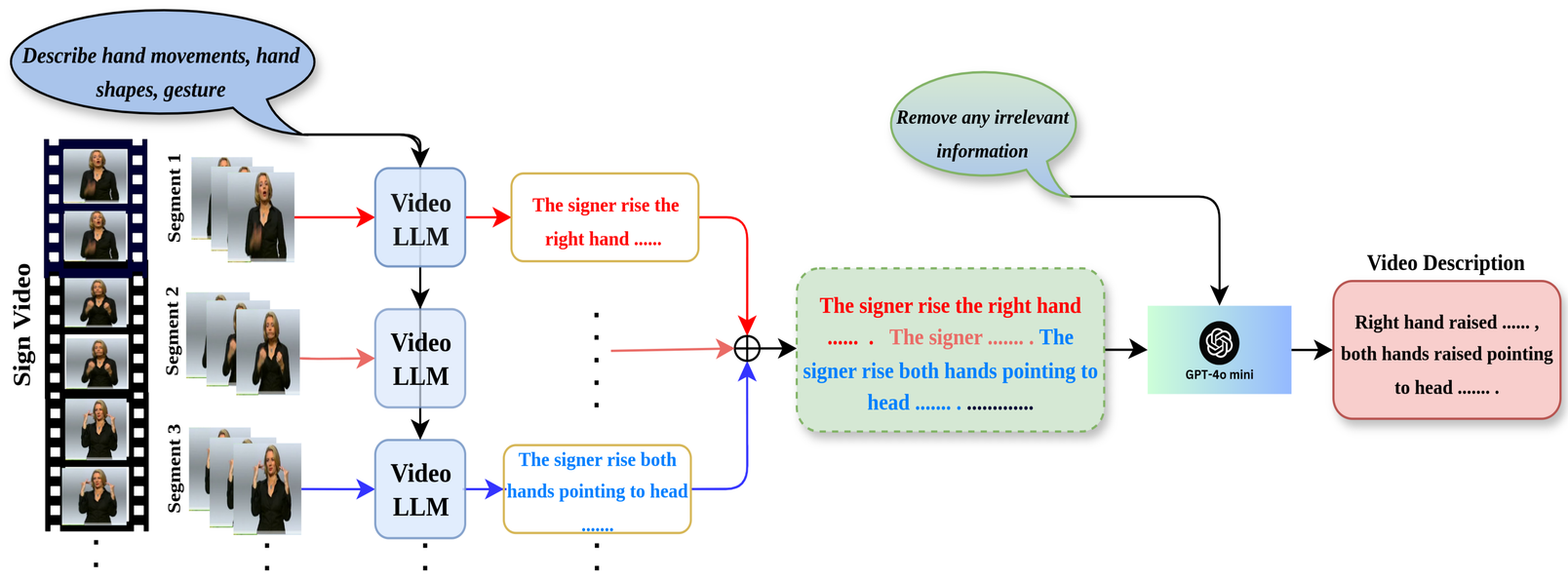
Contributions

- We introduce **BeyondGloss**, a novel, gloss-free SLT framework that leverages spatio-temporal reasoning to directly translate sign video to text.
- We propose a new pre-training strategy that learns **hand-centric representations** by aligning video with detailed motion descriptions.
- We enrich the model's understanding of hands by distilling features from **Hand Mesh Recovery (HaMeR)**.
- We use **contrastive learning** to align video representations with text, effectively closing the modality gap.

Method



Sign Video Descriptor



Quantitative Results

- **The Result: BeyondGloss** achieves **state-of-the-art performance** on major SLT benchmarks (Phoenix14T and CSL-Daily), proving the effectiveness of our approach.

Method	Phoenix14T					CSL-Daily				
	B-1	B-2	B-3	B-4	R	B-1	B-2	B-3	B-4	R
Gloss-based										
MMTLB	53.97	41.75	33.84	28.39	52.65	53.31	40.41	30.87	23.92	53.25
SLTUNET	52.92	41.76	33.99	28.47	52.11	54.98	41.44	31.84	25.01	54.08
TS-SLT	54.90	42.43	34.46	28.95	53.48	55.44	42.59	32.87	25.79	55.72
Gloss-free										
NSLT+Luong	29.86	17.52	11.96	9.00	30.70	34.16	19.57	11.84	7.56	34.54
CSGCR	36.71	25.40	18.86	15.18	38.85	–	–	–	–	–
GFSLT-VLP	43.71	33.18	26.11	21.44	42.49	39.37	24.93	16.26	11.00	36.44
Sign2GPT	49.54	35.96	28.83	22.52	48.90	41.75	28.73	20.60	15.40	42.36
SignCL	49.76	36.85	29.97	22.74	49.07	47.47	32.53	22.62	16.16	48.92
FLa-LLM	46.29	35.33	28.03	23.09	45.27	37.13	25.12	18.38	14.20	37.25
SignLLM	45.21	34.78	28.05	23.40	44.49	39.55	28.13	20.07	15.75	39.91
LLaVA-SLT	51.20	37.51	29.39	23.43	50.44	52.15	36.24	26.47	20.42	51.26
BeyondGloss	**52.38**	**38.57**	**30.74**	**25.49**	**52.89**	**53.12**	**38.63**	**27.82**	**21.53**	**53.46**

Qualitative Results

Reference: am tag wechseln sonne und wolken einander ab teilweise ist es auch langere zeit sonnig .
(During the day sun and clouds alternate partly it is sunny for a long time)

GFSLT-VLP: am tag wechseln sonne und wolken einander ab es bilden sich langere zeit viel sonnenschein.
(During the day, sun and clouds alternate There is a lot of sunshine for a long time)

BeyondGloss: am tag wechseln sonne und wolken einander ab zeigt sich die auch langere zeit sonnig .
(During the day, sun and clouds alternate, and the weather remains sunny for longer periods.)

Reference: am tag nur hier und da einige sonnige momente vor allem an den alpen .
(During the day only here and there some sunny moments, especially in the Alps)

GFSLT-VLP: morgen zeigt sich mal die sonne wenn dann vor allem an den alpen.
(Tomorrow the sun will show up , especially in the Alps)

BeyondGloss: und morgen scheint und da einige die sonnige vor allem an den alpen .
(Tomorrow will shine and there will be some sunny weather, especially in the Alps .)

Reference: am freitag scheint abseits der nebelgebiete haufig die sonne .
(On Friday the sun often shines outside of the foggy areas)

GFSLT-VLP: am freitag sobald der nebel weg ist sonnenschein.
(On Friday as soon as the fog is gone there will be sunshine)

BeyondGloss: am freitag scheint abseits der nebelgebiete haufig die sonne .
(On Friday the sun often shines outside of the foggy areas)

Ablation

	Feature Distillation	Video-Description Alignment	Video-Target Alignment	B-4	R
(1)	–	–	–	16.31	36.33
(2)	–	–	✓	21.87	44.30
(3)	–	✓	✓	23.47	47.17
(4)	✓	✓	–	22.47	44.87
(5)	✓	✓	✓	**25.68**	**53.85**

Acknowledgements

This work was supported by the SNSF project 'SMILE II' (CR SII5 193686), the Innosuisse IICT Flagship (PFFS-21-47), EPSRC grant APP24554 (SignGPT-EP/Z535370/1) and through funding from Google.org via the AI for Global Goals scheme. This work reflects only the authors' views, and the funders are not responsible for any use that may be made of the information it contains.

Paper Link