Supplementary Material for Temporal Lift Pooling for Continuous Sign Language Recognition

Lianyu Hu[1], Liqing Gao[1], Zekang Liu[1], and Wei Feng[✉1]

Tianjin University, Tianjin 300350, China
{hly2021,lqgao,lzk100953}@tju.edu.cn;wfeng@ieee.org

1 Ablations of Loss C_u and C_p

Ablations of Loss C_u and C_p are given in tab. 1. One can notice that employing either loss only give little accuracy boost. Employing both losses benefits most.

C_u	C_p	PHOENIX14 Dev(%)	PHOENIX14 Test(%)	PHOENIX14-T Dev(%)	PHOENIX14-T Test(%)
×	×	20.8	21.7	20.4	22.0
×	✓	20.4	21.4	19.9	21.7
✓	×	20.3	21.3	20.0	21.7
✓	✓	**19.7**	**20.8**	**19.4**	**21.2**

Table 1: Effectiveness of Loss C_u and C_p on two datasets.

2 Flexibility of TLP upon other CSLR methods

Tab. 2 shows the results of equipping FCN [1] with TLP, which gains +1.8% WER boost on average. As no official code is released for FCN, we manually reproduce it.

Methods	PHOENIX14 Dev(%)	PHOENIX14 Test(%)	PHOENIX14-T Dev(%)	PHOENIX14-T Test(%)
FCN [3]	24.2	24.5	23.6	25.2
FCN w/ TLP	22.4(+**1.8**)	22.7(+**1.8**)	21.9(+**1.7**)	23.1(+**2.1**)

Table 2: Flexibility of TLP upon FCN [3] on two datasets.

3 Results on the CSL dataset

Tab. 3 validates the effectiveness of our TLP on the chinese sign language dataset (CSL) [3]. The CSL dataset is collected within laboratory environments with a vocabulary size of 178 with 100 sign language sentences. Each sentence is performed by fifty signers with five times, resulting in total 25000 videos with 100+ hours. Our method achieves significant accuracy boost (+5.5%) on the CSL dataset.

Methods	CSL [3](%)
SubUNet [2]	11.0
FCN [1]	3.0
STMC [5]*	2.1
VAC [4]	1.6
Baseline	7.3
Baseline w/ TLP	1.8 (+**5.5**)

Table 3: Effectiveness of TLP on the CSL Dataset.

References

1. Cheng, K.L., Yang, Z., Chen, Q., Tai, Y.W.: Fully convolutional networks for continuous sign language recognition. In: ECCV (2020)
2. Cihan Camgoz, N., Hadfield, S., Koller, O., Bowden, R.: Subunets: End-to-end hand shape and continuous sign language recognition. In: ICCV (2017)
3. Huang, J., Zhou, W., Zhang, Q., Li, H., Li, W.: Video-based sign language recognition without temporal segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
4. Min, Y., Hao, A., Chai, X., Chen, X.: Visual alignment constraint for continuous sign language recognition. In: ICCV (2021)
5. Zhou, H., Zhou, W., Zhou, Y., Li, H.: Spatial-temporal multi-cue network for continuous sign language recognition. In: AAAI (2020)