Supplementary material

Table 1. Computational comparison with existing open-source long-term AQA methods on the RG dataset. SRCC gains and RL2 drops are computed w.r.t. ACTION-NET and GDLT, respectively.

Method	FLOPs (G)	Parameter (M)		SRCC Gain / RL2 Drop
		Online	Offline	
ACTION-NET [2]	34.7500	28.08	–	0% / –
GDLT [1]	0.1164	3.20	–	+5.1% / 0%
HGCN [3]	1.1201	0.50	–	+1.0% / −80.7%
CoFInAl [4]	0.1178	3.70	–	+10.9% / −67.8%
PHI [5]	0.2637	3.00	4.60	+11.3% / +4.2%
CaFlow (Ours)	1.0418	3.00	9.79	**+15.5% /** −2.0**%**

References

[1] Angchi Xu, Ling-An Zeng, and Wei-Shi Zheng. Likert scoring with grade decoupling for long-term action assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3232–3241, 2022. 1

[2] Ling-An Zeng, Fa-Ting Hong, Wei-Shi Zheng, Qi-Zhi Yu, Wei Zeng, Yao-Wei Wang, and Jian-Huang Lai. Hybrid dynamic-static context-aware attention network for action assessment in long videos. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2526–2534, 2020. 1

[3] Kanglei Zhou, Yue Ma, Hubert PH Shum, and Xiaohui Liang. Hierarchical graph convolutional networks for action quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(12):7749–7763, 2023. 1

[4] Kanglei Zhou, Junlin Li, Ruizhi Cai, Liyuan Wang, Xingxing Zhang, and Xiaohui Liang. Cofinal: Enhancing action quality assessment with coarse-to-fine instruction alignment. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 2024. 1

[5] Kanglei Zhou, Hubert PH Shum, Frederick WB Li, Xinxing Zhang, and Xiaohui Liang. Phi: Bridging domain shift in long-term action quality assessment via progressive hierarchical instruction. *IEEE Transactions on Image Processing*, 2025. 1