

Supplementary Materials: Highly Efficient No-reference 4K Video Quality Assessment with Full-Pixel Covering Sampling and Training Strategy

Anonymous Authors

1 SUPPLEMENTAL EXPERIMENTAL RESULTS

To more comprehensively evaluate the performance of our method, we conduct two additional experiments: the first one is a comparison experiment conducted on the MCML 4K UHD dataset [4]; the second one is an ablation study for the frame sampling intervals t of our method.

1.1 Additional Comparison Results

Table 1: Comparison on MCML 4K UHD. Bold: best.

Method	Source	SRCC	PLCC
BTURA [23]	TB2022	0.935	0.960
FAST-VQA [41]	ECCV2022	0.902	0.923
DOVER [43]	ICCV2023	0.888	0.910
SAMA [20]	AAAI2024	0.916	0.938
Proposed	Proposed	0.961	0.979

The comparison experiment is conducted on the MCML 4K UHD dataset [4]. Although this dataset is a 4K Video Quality Assessment (VQA) dataset, it only includes 10 source reference sequences (SRC) and is built by conducting compression and up-scaling on these SRC videos, resulting in the simplicity of video scenes. So the performances of the comparison methods on this dataset shown in

Table1 are significantly higher than those on our 4K dataset and other open-source datasets listed in our original paper. However, our proposed method still achieves the highest performance on this dataset compared to other methods, including BTURA [23], FAST-VQA [41], DOVER [43], and SAMA [20]. It should be noted that although BTURA utilizes extra information about whether a 4K content is true or pseudo, our method still outperforms it.

1.2 Additional Ablation Study

Table 2: Ablation Study on Different Frame Sampling Intervals.

t	2	4	6	8	10
SRCC	0.918	0.917	0.915	0.914	0.914
PLCC	0.935	0.935	0.932	0.932	0.932
Time(s)	0.506	0.254	0.169	0.127	0.101

We assess the performances of our method on our proposed 4K VQA dataset for different settings of t , which means the sample interval of all video frames. As shown in Table2, as t increases, the inference time for processing a one-second video grows rapidly, while the performance does not show a significant decrease. Therefore, we set t to be 10 for the final implementation.