Appendix

Table 1. Settings of τ and γ in QVRS computation at the highest QA accuracy.

MLLM	Frames	VideoMME		LongVideoBench	
		τ	γ	τ	γ
Qwen2.5-VL-7B	32	0.4	0.7	0.2	0.6
Qwen2.5-VL-7B	64	0.6	0.6	0.2	0.7
Qwen2.5-VL-32B	64	0.5	0.6	0.3	0.7
InternVL3-8B	64	0.4	0.7	0.6	0.7

A. Hyperparameter Settings and Analysis

For **AKS**, the hyperparameters were set to $s_{\text{thr}} = 0.2, L = 2$ on LongVideoBench and $s_{\text{thr}} = 0.8, L = 5$ on VideoMME. For **ITS**, the hyperparameter α was set to 7.0 on LongVideoBench and 2.0 on VideoMME. In our method **ASCS**, α was set identically to ITS, and the settings of τ and γ in QVRS computation at the highest QA accuracy are shown in Tab. 1. The results indicate that although the optimal hyperparameters vary across different configurations, they generally fall within certain ranges depending on the dataset. Due to the inherent uncertainty and errors of MLLMs, however, it is difficult for the optimal hyperparameters to remain consistent across models or frame budgets.

In Tab. 2, we analyze the effect of different choices of the hyperparameters τ and γ on QA accuracy and sampling quality (measured by UKSS) when using Qwen2.5-VL-32B. As shown in the table, the parameter setting that achieves the highest QA accuracy ($\tau = 0.3$, $\gamma = 0.7$) also yields the highest UKSS, and the top three settings in accuracy are included among the top six in UKSS. These results demonstrate a certain degree of correlation between QA accuracy and UKSS, highlighting the value of the proposed metric. In addition, all hyperparameter settings in the table achieve higher accuracy than ITS and AKS, further confirming the stability and effectiveness of our method.

B. Visualization of Accuracy–UKSS Correlation

To provide an intuitive illustration of the correlation between QA accuracy and UKSS, we visualize the results as scatter plots. Fig. 1 shows the results under the four settings in Table 3 of the main paper, evaluated on both datasets. The x-axis represents UKSS values, and the y-axis represents QA accuracy. The plots reveal a clear positive correlation between accuracy and UKSS, consistent with the correlation coefficients reported in Table 3 of the main paper.

Table 2. Analysis of different τ and γ settings on QA accuracy (left) and UKSS (right) when using Qwen2.5-VL-32B.

		γ	
		0.6	0.7
τ	0.1	64.2 / 0.266	64.0 / 0.265
	0.2	65.4 / 0.266	65.2 / **0.269**
	0.3	65.4 / 0.267	**65.5** / **0.269**
	0.4	64.6 / 0.268	65.1 / 0.262
	0.5	65.1 / 0.264	64.2 / 0.262

We further observe that when UKSS lies within 0.01 of its maximum value (referred to as the UKSS-optimal range), accuracy still exhibits noticeable variance. This indicates that UKSS alone cannot perfectly identify the sampling method or hyperparameters that yield the highest accuracy. However, when accuracy lies within 1% of its maximum value (referred to as the accuracy-optimal range), UKSS is also found to be within or very close to its optimal range. This suggests that, when tuning sampling strategies with the goal of maximizing QA accuracy, UKSS can substantially narrow the search space and reduce the number of MLLM evaluations required, thereby lowering experimental cost.

C. Additional Experiments on Factors Affecting QA Accuracy

In addition to the experiments in Section 3.3 conducted on VideoMME with InternVL3-8B, we further performed experiments on LongVideoBench and with Qwen2.5-VL-7B. Tab. 3 and Tab. 4 report the effects of Key Frame Rate (KFR) and Scene Hit Rate (SHR) on QA accuracy when using InternVL3-8B and Qwen2.5-VL-7B, respectively. Tab. 5 and Tab. 6 show the effects of key frame distribution on QA accuracy under the same settings.

The observations from these results are almost entirely

consistent with those in Section 3.3, with one minor exception: in Tab. 4, when controlling SHR on VideoMME, accuracy peaks at KFR = 60% before slightly decreasing. However, since the decrease is marginal and such high KFR values rarely occur in practice, this phenomenon does not affect the design of the UKSS metric.

D. Results by Single-Scene and Multi-Scene Samples

Tab. 7 reports the results separately for samples annotated with a single relevant scene and those annotated with multiple relevant scenes. We observe that our method achieves the best or second-best performance on most metrics, yielding the strongest overall QA and key frame sampling performance on both single-scene and multi-scene samples. However, our method underperforms ITS in UKSS on multi-scene samples of VideoMME, and its QA performance shows no clear advantage. Moreover, compared with the results reported in Section 5.3, the correlation between UKSS and QA accuracy is noticeably weaker on both single-scene and multi-scene samples, highlighting the limitations of UKSS and indicating that substantial room for improvement remains, particularly for multi-scene cases.

Table 3. QA accuracy on **VideoMME**$_{\text{kfs}}$ (left) and **LongVideoBench**$_{\text{kfs}}$ (right) with different Key Frame Rates and Scene Hit Rates, where the number of frames per scene is proportional to scene duration. **MLLM: InternVL3-8B.**

		Key Frame Rate					
		0%	20%	40%	60%	80%	100%
Scene Hit Rate	0%	53.8 / 52.4	-	-	-	-	-
	20%	-	64.6 / 64.6	66.7 / 66.6	67.3 / 67.8	66.8 / 69.1	67.3 / 71.2
	40%	-	65.0 / 65.0	66.6 / 66.8	67.3 / 68.3	67.5 / 70.0	68.7 / 72.0
	60%	-	65.3 / 65.3	67.6 / 67.4	67.8 / 68.9	68.2 / 70.5	68.8 / 73.3
	80%	-	65.5 / 66.0	67.8 / 68.3	68.9 / 70.2	69.9 / 71.6	71.3 / 75.7
	100%	-	66.9 / 66.9	68.7 / 69.1	70.0 / 70.9	71.2 / 72.9	73.2 / 77.8

Table 4. QA accuracy on **VideoMME**$_{\text{kfs}}$ (left) and **LongVideoBench**$_{\text{kfs}}$ (right) with different Key Frame Rates and Scene Hit Rates, where the number of frames per scene is proportional to scene duration. **MLLM: Qwen2.5-VL-7B.**

		Key Frame Rate					
		0%	20%	40%	60%	80%	100%
Scene Hit Rate	0%	49.7 / 51.2	-	-	-	-	-
	20%	-	64.6 / 65.4	66.7 / 68.6	69.3 / 70.3	67.6 / 71.6	67.3 / 74.6
	40%	-	65.3 / 65.7	67.5 / 69.3	70.5 / 71.2	68.7 / 72.1	68.8 / 75.3
	60%	-	65.8 / 66.4	68.0 / 70.0	71.6 / 71.9	70.0 / 73.3	69.7 / 76.7
	80%	-	66.1 / 66.5	68.5 / 70.5	72.1 / 73.0	71.3 / 74.8	71.1 / 79.2
	100%	-	66.2 / 67.4	69.3 / 71.6	73.0 / 74.8	73.0 / 76.5	72.5 / 81.6

Table 5. QA accuracy on **VideoMME**$_{\text{kfs}}$ (left) and **LongVideoBench**$_{\text{kfs}}$ (right) under different distributions of key frames across scenes controlled by the parameters C and β. **MLLM: InternVL3-8B.**

		β					
		0	0.2	0.5	1	2	5
C	0.05	68.2 / 73.3	68.4 / 73.3	68.6 / 74.1	68.0 / 73.9	68.4 / 73.9	68.6 / 73.4
	0.2	68.6 / 74.1	68.0 / 74.8	68.7 / 75.0	69.0 / 73.8	68.7 / 74.0	68.2 / 73.9
	1	71.1 / 75.7	70.9 / 75.9	69.4 / 75.3	69.8 / 74.7	70.3 / 74.7	69.5 / 73.9
	5	72.2 / 77.5	72.3 / 77.7	72.4 / 77.4	71.7 / 76.7	70.7 / 76.3	69.6 / 74.1
	20	73.1 / 77.9	72.7 / 77.8	72.7 / 77.4	72.4 / 77.6	71.2 / 76.1	70.7 / 74.9

Table 6. QA accuracy on **VideoMME**$_{\text{kfs}}$ (left) and **LongVideoBench**$_{\text{kfs}}$ (right) under different distributions of key frames across scenes controlled by the parameters C and β. **MLLM: Qwen2.5-VL-7B.**

		β					
		0	0.2	0.5	1	2	5
C	0.05	68.1 / 76.1	68.0 / 75.9	68.1 / 77.0	67.9 / 77.4	68.2 / 77.8	68.0 / 77.6
	0.2	68.6 / 77.4	68.0 / 77.5	68.1 / 77.6	68.7 / 77.0	68.0 / 77.6	67.5 / 77.7
	1	69.8 / 79.7	69.8 / 79.7	68.4 / 79.2	68.8 / 79.2	68.7 / 79.1	68.6 / 78.4
	5	72.1 / 81.0	72.5 / 80.8	70.9 / 80.9	72.2 / 79.8	70.8 / 79.8	69.3 / 78.4
	20	72.2 / 81.2	72.5 / 81.6	72.0 / 81.2	72.4 / 81.0	71.8 / 79.8	70.3 / 78.7

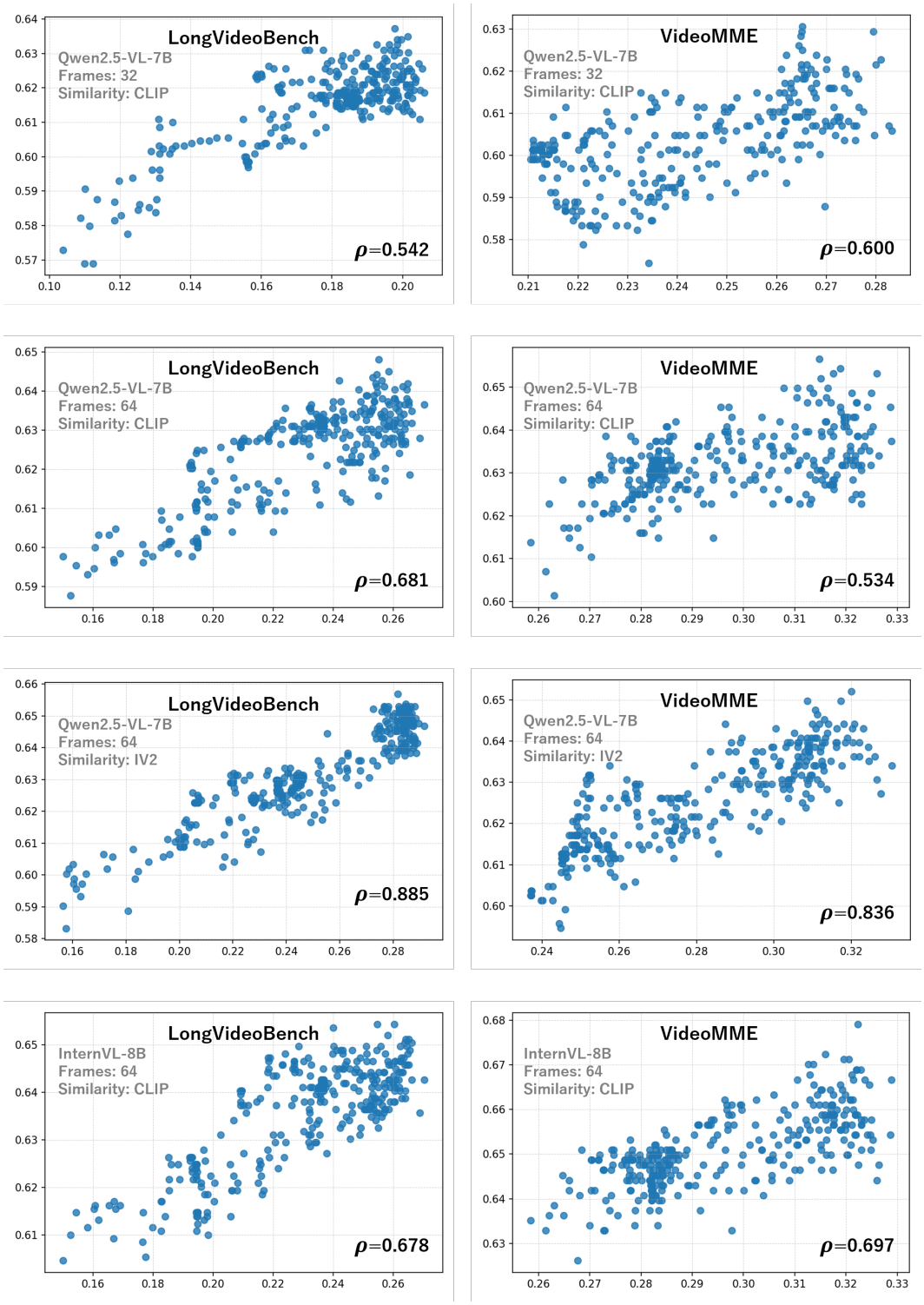

Figure 1. Scatter plots of QA accuracy versus UKS. The x-axis represents UKSS values, and the y-axis represents QA accuracy. A clear positive correlation is observed, consistent with the correlation coefficients. While accuracy still varies within the *UKSS-optimal range* (UKSS within 0.01 of its maximum), the *accuracy-optimal range* (within 1% of maximum accuracy) consistently corresponds to UKSS values close to their optimum, showing that UKSS can effectively narrow the search space and reduce evaluation cost.

Table 7. Experimental results on VideoMME and LongVideoBench (LVB). QA accuracy (acc.) and UKSS are reported separately for the single-scene and multi-scene samples of the KFS-Bench subsets.

MLLM	Frames	Sampling Method	**VideoMME**				**LVB**			
			Single		*Multi*		*Single*		*Multi*	
			Acc.	UKSS	Acc.	UKSS	Acc.	UKSS	Acc.	UKSS
Qwen2.5-VL-7B	32	Uniform	63.7	0.198	46.0	0.211	59.5	0.083	51.2	0.161
		K-means	64.7	0.208	46.0	0.202	62.0	0.094	**54.2**	0.173
		AKS	66.4	0.241	47.8	**0.263**	65.2	0.185	51.2	0.195
		ITS	64.5	0.258	**50.4**	0.256	65.6	**0.192**	51.8	**0.237**
		ASCS	**68.1**	**0.269**	49.3	0.255	**67.7**	0.189	52.4	0.228
Qwen2.5-VL-7B	64	Uniform	67.4	0.265	53.3	0.301	62.0	0.123	54.2	0.224
		K-means	67.1	0.278	54.0	0.292	62.0	0.138	53.0	0.230
		AKS	69.2	0.278	51.1	0.322	**67.3**	0.230	52.1	0.249
		ITS	69.4	0.304	**55.1**	**0.334**	66.5	0.244	54.5	0.293
		ASCS	**69.7**	**0.317**	52.2	0.318	67.1	**0.257**	**56.1**	**0.308**
Qwen2.5-VL-32B	64	Uniform	69.5	0.265	55.1	0.301	61.4	0.123	57.0	0.224
		K-means	70.2	0.278	**58.4**	0.292	63.1	0.138	57.6	0.230
		AKS	71.3	0.278	54.4	0.322	66.6	0.230	54.5	0.249
		ITS	**71.5**	0.304	56.6	**0.334**	66.1	0.244	56.7	0.293
		ASCS	71.2	**0.306**	58.0	0.328	**68.0**	**0.252**	**58.5**	**0.325**
Intern VL3-8B	64	Uniform	68.7	0.265	54.4	0.301	62.4	0.123	55.5	0.224
		K-means	69.1	0.278	55.8	0.292	64.4	0.138	56.7	0.230
		AKS	69.9	0.278	56.2	0.322	67.4	0.230	54.2	0.249
		ITS	71.2	**0.304**	56.9	**0.334**	**67.8**	0.244	56.7	0.293
		ASCS	**71.3**	0.304	**59.5**	0.329	67.7	**0.253**	**58.2**	**0.317**