

Table 1: Temporal grounding and complex reasoning under the shared 256-frame protocol. Temporal grounding uses mIoU; reasoning columns use official benchmark scores. Bold marks the best score within each backbone group.

Model	Temporal Grounding			Complex Reasoning			
	QV Highlights	Activity Net	Charades STA	Video Holmes	Video MMMU	Video TT	E.T. Bench
Qwen3-VL-4B Backbone							
Qwen3-VL-4B [4]	54.31	41.99	43.64	39.96	55.44	36.90	34.55
+ SFT-Desc	55.63	42.18	43.52	41.26	55.00	41.70	32.57
+ SFT-DVC	57.13	42.36	40.18	41.97	56.22	41.20	37.78
+ EventLens	58.50	45.76	47.99	40.28	55.56	38.80	41.47
Qwen3-VL-8B Backbone							
Qwen3-VL-8B [4]	50.03	39.96	44.79	39.96	60.11	40.00	32.77
+ SFT-Desc	51.44	39.70	44.80	42.24	58.00	41.80	33.67
+ SFT-DVC	57.61	43.70	42.56	42.57	57.33	42.00	39.45
+ EventLens	59.12	46.35	48.25	40.07	59.44	39.60	38.40