Benchmark	QA Pairs	Data	Env.	Scene			View		Evaluation		Spatio-Temporal			
				D	I	O	Ego	Allo.	Num.	Desc.	Dist.	Dir.	Vel.	Traj.
SAT	218k	**I**	**S**	✗	✓	✗	✓	✓	✓	✓	✗	✗	✗	✗
VSI-Bench	5,156	**V**	**R**	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗	✓
EmbSpatial-Bench	3,640	**I**	**R**	✗	✓	✗	✓	✗	✗	✓	✗	✗	✗	✗
EmbodiedAgentInterface	448	-	**S**	✗	✓	✗	✓	✗	-	-	✗	✗	✗	✗
EmbodiedEval	328	**I/V**	**S**	✗	✓	✓	✓	✗	-	-	✗	✗	✗	✗
EmbodiedBench	1,128	**I**	**S**	✗	✓	✓	✓	✗	-	-	✗	✗	✗	✗
WorldSense	3,172	**V**	**R**	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗
MLVU	3,102	**V**	**R**	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗
Video-MMMU	300	**V**	**S**	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
ST-Bench	2,039	**V**	**R**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 1: **Comparison of ST-Bench with existing benchmarks.**