SpatialScore: Towards Comprehensive Evaluation for Spatial Intelligence

Haoning Wu[1,2*], Xiao Huang[1*], Yaohui Chen[1] Ya Zhang[1,2], Yanfeng Wang[1], Weidi Xie[1]

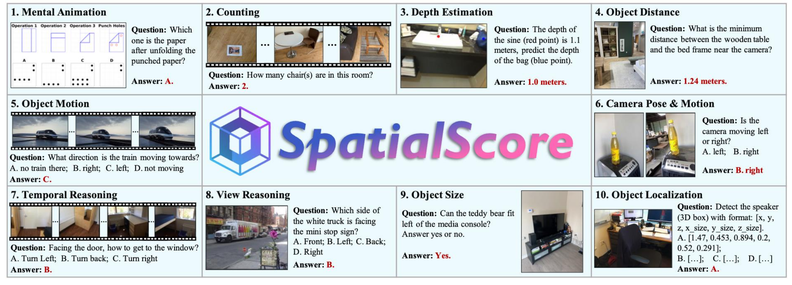
[1]*School of Artificial Intelligence, Shanghai Jiao Tong University* [2]*Shanghai Artificial Intelligence Laboratory*

CVPR JUNE 3-7, 2026 DENVER COLORADO

Contributions

- **SpatialScore**: the most comprehensive and diverse ***spatial intelligence benchmark*** to date, comprising ***~5K challenging samples*** spanning ***30 distinct tasks***;
- **SpatialCorpus**: a ***large-scale training resource*** with ***331K multimodal QA samples*** for supervised fine-tuning (SFT) on spatial understanding tasks;
- **SpatialAgent**: a ***multi-agent system*** with ***12 perception tools*** that supports ***Plan-Execute*** and ***ReAct*** reasoning, enabling substantial gains in a ***tuning-free manner***;
- **Evaluation**: extensive assessments over ***49 representative MLLMs*** reveal a substantial gap between current models and human-level spatial intelligence.

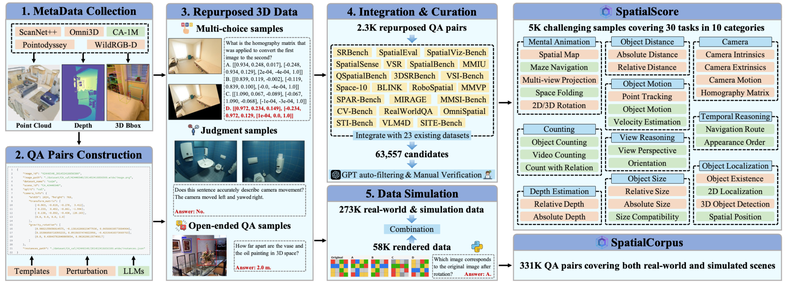
SpatialScore Benchmark Overview



Qualitative Comparison

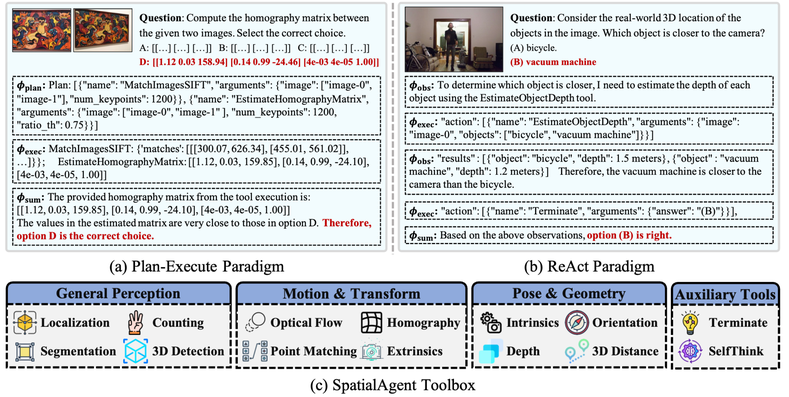
Data Construction

Dataset	Publication	Data Types			Input Modalities			QA Formats			#Tasks	#Samples
		Real	Simulated	AIGC	Image	Sequence	Video	MCQ	Yes/No	Open		
QSpatialBench [44]	EMNLP 2024	✓	✗	✗	✓	✗	✗	✗	✗	✓	2	271
SpatialEval [81]	NeurIPS 2024	✓	✓	✗	✓	✗	✗	✓	✓	✗	4	4,635
VSI-Bench [98]	CVPR 2025	✓	✗	✗	✓	✗	✓	✓	✗	✓	8	5,156
RoboSpatial-Home [72]	CVPR 2025	✓	✗	✗	✓	✗	✗	✗	✓	✓	3	350
3DSRBench [56]	ICCV 2025	✓	✓	✗	✓	✓	✓	✓	✓	✗	4	2,772
STI-Bench [42]	ICCV 2025	✓	✗	✗	✗	✗	✓	✓	✗	✗	8	2,064
VLM4D [113]	ICCV 2025	✓	✗	✓	✗	✗	✓	✓	✓	✗	4	1,816
SITE-Bench [88]	ICCV 2025	✓	✓	✗	✓	✓	✓	✓	✗	✗	6	8,068
MIRAGE [47]	NeurIPS 2025	✓	✗	✗	✓	✗	✗	✗	✓	✓	3	1,710
SPAR-Bench [110]	NeurIPS 2025	✓	✓	✗	✓	✓	✓	✓	✓	✗	20	7,211
SpatialViz-Bench [86]	ICLR 2026	✗	✓	✗	✓	✗	✗	✓	✗	✗	12	1,180
MMSI-Bench [103]	ICLR 2026	✓	✓	✗	✗	✓	✓	✓	✗	✗	11	1,000
OmniSpatial [33]	ICLR 2026	✓	✓	✗	✓	✓	✓	✓	✓	✗	50	1,533
SpatialScore (Ours)	CVPR 2026	✓	✓	✓	✓	✓	✓	✓	✓	✓	**30**	**5,025**



(a) Data Construction Pipeline of SpatialScore and SpatialCorpus

SpatialAgent Framework



(a) Plan-Execute Paradigm (b) ReAct Paradigm

(c) SpatialAgent Toolbox

Quantitative Comparison

Methods	Overall	Rank	Mental.	Count.	Depth.	View-Rea.	Obj-Size.	Obj-Loc.	Obj-Dist.	Obj-Mo.	Camera.	Temp-Rea.
Baselines												
Blind GPT-5 (Text-only) [12]	30.62	-	18.79	20.34	29.36	40.81	31.05	45.34	24.20	25.54	32.01	26.47
Chance-level (Random)	28.29	-	23.71	22.80	22.88	31.84	30.29	38.59	24.06	25.06	28.79	28.68
Human-level	**86.60**	-	**96.87**	**89.72**	**82.33**	**92.15**	**85.18**	**81.64**	**78.96**	**94.46**	**86.89**	**84.19**
Small-scale models (1B~4B)												
InternVL2.5-1B [15]	31.66	13	30.20	38.42	22.26	32.29	32.29	51.08	25.12	26.02	27.76	25.74
InternVL3-1B [14]	33.03	12	26.85	47.69	24.74	32.06	34.06	57.53	24.02	32.29	25.71	19.85
InternVL3.5-1B [87]	33.55	11	31.10	40.27	33.47	30.72	30.95	53.08	26.72	24.82	29.43	29.41
Qwen3-VL-2B [5]	41.41	5	35.35	**52.74**	34.64	35.65	49.69	61.69	35.42	38.80	30.59	39.34
Qwen2.5-VL-3B [6]	36.92	8	32.21	46.07	33.07	34.53	38.48	56.53	29.73	37.59	30.59	24.26
SpaceQwen2.5VL-3B [13]	31.25	14	20.81	47.74	22.51	32.96	38.82	39.45	25.13	30.60	29.95	24.26
SpaceThinker-3B [13]	34.47	10	27.96	48.30	34.64	34.08	36.89	44.76	28.07	35.42	28.66	26.84
SpatialThinker-3B [7]	37.16	7	29.75	50.79	37.45	37.22	31.69	58.97	31.87	35.18	30.08	22.79
VST-3B-SFT [101]	47.37	2	25.73	48.45	**53.11**	**45.74**	53.02	61.41	42.35	60.48	35.35	50.74
VST-3B-RL [101]	**48.04**	**1**	24.61	52.38	51.75	43.72	54.16	**61.84**	**43.03**	**62.89**	**38.95**	47.43
Spatial-MLLM-4B [91]	42.59	3	33.56	47.61	34.80	43.27	**56.16**	57.68	40.49	36.39	28.41	**53.31**
Qwen3-VL-4B [5]	42.52	4	**37.81**	48.22	37.68	34.75	48.20	59.40	33.40	53.01	34.06	38.24
InternVL2.5-4B [15]	36.18	9	27.52	45.15	41.25	35.65	35.43	57.25	29.23	29.64	29.18	23.53
InternVL3.5-4B [87]	37.81	6	34.23	40.40	40.97	34.75	29.59	60.11	31.15	38.07	30.46	34.19
Middle-scale models (7B~14B)												
LLaVA-1.5-7B [49]	25.32	19	20.58	21.64	29.34	22.42	22.25	47.20	14.43	24.58	27.63	2.21
LLaVA-OneVision-7B [38]	36.57	15	25.50	45.97	38.92	36.32	42.31	56.67	29.01	29.40	32.13	16.18
Qwen2.5-VL-7B [6]	40.53	9	42.06	48.67	43.44	37.00	40.88	57.82	34.02	43.61	28.66	26.84
SpaceR-7B [63]	43.36	6	44.74	51.13	45.31	37.44	48.48	58.39	35.98	45.06	33.03	31.62
VST-7B-SFT [101]	50.06	2	39.37	54.00	53.67	43.05	52.91	63.27	46.44	62.65	**39.72**	**45.96**
VST-7B-RL [101]	**52.47**	**1**	41.39	**54.92**	**56.80**	**50.00**	**55.48**	**64.56**	**48.65**	**63.37**	**42.80**	45.59
SpatialThinker-7B [7]	40.83	8	39.82	49.46	44.41	34.75	41.88	59.40	33.13	44.34	31.36	23.90
SpatialReasoner-8B [57]	36.31	16	36.02	41.21	36.12	36.32	34.63	53.23	28.00	33.73	31.75	26.10
InternVL2.5-8B [15]	39.63	11	40.27	45.85	45.39	35.43	45.25	58.82	34.07	27.71	28.53	28.31
InternVL3-8B [114]	41.57	7	37.81	51.95	39.46	38.79	43.05	64.13	36.02	41.69	29.95	28.31
InternVL3.5-8B [87]	37.75	12	38.48	42.64	41.25	31.39	33.29	60.83	32.27	36.39	32.39	13.60
Qwen3-VL-8B [5]	45.48	3	38.26	50.35	47.12	37.67	52.12	61.12	37.22	55.90	34.19	41.54
LLaMA-3.2V-11B [29]	35.13	17	40.72	47.81	38.96	36.55	25.98	51.94	25.33	33.98	29.05	17.28
LLaMA-3.2V-11B-CoT [95]	37.51	13	28.64	39.71	40.51	35.87	47.98	54.52	31.84	31.77	29.18	25.74
LLaVA-1.5-13B [49]	26.67	18	25.06	25.91	31.52	20.63	23.59	47.63	16.10	26.51	28.15	1.84
SpaceLLaVA-13B [13]	23.97	20	34.23	32.14	26.93	22.65	19.35	36.01	15.13	19.52	15.30	23.16
InternVL3-14B [114]	44.89	4	44.52	53.05	42.92	37.89	48.86	62.84	40.86	43.37	35.35	35.29
InternVL3.5-14B [87]	43.88	5	**45.64**	46.63	41.66	40.36	46.95	**64.56**	36.82	42.65	35.60	29.04
Kimi-VL-16B-A3B [76]	40.10	10	29.08	53.32	38.25	41.03	45.17	59.11	35.85	43.61	27.38	25.74
Kimi-VL-A3B-Thinking [76]	37.06	14	30.20	40.70	32.03	35.43	45.74	52.37	28.22	32.77	33.16	33.82
Large-scale models (30B~78B)												
Qwen3-VL-30B-A3B [5]	50.71	3	46.31	58.86	48.49	47.98	56.52	66.43	40.73	59.52	39.20	45.59
Qwen2.5-VL-32B [6]	47.23	8	45.41	50.38	52.28	39.46	49.25	61.84	40.35	56.14	40.23	29.04
Qwen3-VL-32B [5]	54.11	2	43.40	61.16	51.80	50.22	59.53	69.58	**50.94**	66.75	41.26	47.79
InternVL2.5-38B [15]	45.60	10	43.62	52.98	52.71	37.67	50.47	62.70	39.16	50.84	35.09	21.69
InternVL3-38B [114]	48.95	5	45.64	58.06	49.10	42.15	50.61	67.58	41.04	57.83	39.46	33.82
InternVL3.5-38B [87]	45.74	9	40.27	43.98	48.58	40.13	52.95	62.84	39.31	48.19	39.72	29.04
LLaVA-OneVision-72B [38]	43.29	11	37.58	51.57	51.45	38.57	50.24	62.12	31.84	37.59	38.43	19.49
Qwen2.5-VL-72B [6]	48.42	7	53.69	49.49	**56.23**	36.10	50.91	62.55	38.42	59.52	**42.93**	22.43
InternVL2.5-78B [15]	48.71	6	51.45	54.15	54.61	43.50	53.68	62.84	41.67	52.05	39.20	25.74
InternVL3-78B [114]	50.67	4	50.34	59.19	48.74	45.74	50.32	65.71	42.50	60.48	41.52	43.75
Qwen3-VL-235B-A22B [5]	**56.63**	**1**	**57.27**	**65.19**	54.04	**52.47**	**59.90**	**70.01**	50.40	**69.40**	42.80	**49.63**
Proprietary Models (Commercial APIs)												
Claude-4.5-Sonnet [4]	45.68	4	51.01	49.67	47.33	39.91	54.08	53.66	38.68	46.99	36.12	41.18
Gemini-2.5-Pro [18]	56.37	3	73.29	**64.04**	51.15	46.41	58.24	66.52	46.93	57.84	48.78	56.25
Gemini-3-Pro [28]	**60.12**	**1**	70.40	62.99	**58.15**	**62.44**	**60.02**	64.20	**49.50**	**71.12**	**51.37**	59.93
GPT-5 [62]	58.13	2	**78.08**	57.59	55.13	54.04	59.22	**67.39**	45.01	57.11	50.90	**62.13**

Methods	Overall	Mental.	Count.	Depth.	View-Rea.	Obj-Size.	Obj-Loc.	Obj-Dist.	Obj-Mo.	Camera.	Temp-Rea.
Qwen3-VL-4B											
Blind (text only)	28.10	27.29	11.20	27.48	25.34	42.99	32.71	25.81	21.69	30.46	20.22
Zero-shot	42.52	37.81	48.22	37.68	34.75	48.20	59.40	33.40	53.01	34.06	38.24
One-shot	44.03	38.26	48.76	41.87	34.75	52.31	61.41	36.09	52.53	35.86	33.09
Two-shot	45.66	40.27	51.32	43.25	34.75	51.88	62.84	37.14	57.59	37.79	36.03
Four-shot	46.59	43.85	51.78	45.16	35.65	53.70	**64.56**	38.19	56.39	35.99	41.91
w/ SpatialCorpus (Ours)	**52.99**	**65.55**	52.24	**58.49**	33.86	43.92	58.97	**47.17**	**64.10**	55.78	**44.85**
w/ SpatialAgent-PE (Ours)	48.93	56.15	**54.98**	45.83	36.55	**54.70**	62.41	36.68	60.48	41.90	38.24
w/ SpatialAgent-ReAct (Ours)	50.30	46.53	53.46	51.75	**39.01**	50.51	58.11	39.20	53.25	**59.00**	42.28
Qwen3-VL-8B											
Blind (text only)	30.73	21.70	17.21	28.78	30.49	42.30	45.77	28.25	20.72	31.75	20.59
Zero-shot	45.48	38.26	50.35	47.12	37.67	52.12	61.12	37.22	55.90	34.19	41.54
One-shot	46.26	41.16	51.71	45.83	39.46	52.64	61.26	38.74	56.39	37.53	34.19
Two-shot	47.61	41.39	54.01	48.97	40.81	53.81	60.11	38.54	61.93	38.17	38.60
Four-shot	49.00	47.87	54.35	49.64	39.01	56.76	**64.13**	39.56	61.45	36.76	41.18
w/ SpatialCorpus (Ours)	**54.71**	**57.27**	52.67	**64.12**	36.77	56.81	60.26	**49.50**	**64.34**	53.73	44.85
w/ SpatialAgent-PE (Ours)	52.75	50.11	**54.48**	54.44	43.50	**58.72**	**64.13**	43.57	61.45	48.71	43.38
w/ SpatialAgent-ReAct (Ours)	53.81	44.30	53.49	54.21	**45.74**	56.52	62.84	43.67	60.96	**58.10**	**51.84**