

Model	Visual Effects			Data Fidelity		Layout Structure		Avg.
	Color	Style	Aspect	Scaling	Encoding	Position	Occlusion	
Proprietary								
Gemini-3-Pro	98.3	98.9	89.0	92.9	98.8	90.9	96.2	95.0
GPT-5.2	89.7	79.5	66.5	79.4	82.9	82.5	77.1	79.7
Seed1.6-VL	91.5	68.2	50.0	44.7	50.6	77.3	69.4	64.5
Claude-Sonnet-4.5	66.7	40.9	59.3	51.8	59.2	40.3	57.3	53.6
Open-Source (thinking)								
MiMo-VL-SFT	80.3	37.5	29.1	27.0	61.0	40.3	46.5	46.0
Qwen3-VL-30B	58.1	28.4	18.7	22.0	37.8	29.9	29.3	32.0
MiMo-VL-RL	65.8	20.4	14.3	7.8	34.8	23.4	27.4	27.7
LLaVA-Critic-R1	7.7	9.1	2.2	9.9	15.8	3.9	5.7	7.8
ThinkLite-VL-7B	12.0	2.3	0.5	3.5	14.0	1.3	3.2	5.3
Open-Source (non-thinking)								
Qwen2-VL-7B	15.4	10.2	11.5	16.3	29.3	4.5	13.4	14.4
Qwen2-VL-72B	36.8	37.5	23.6	25.5	42.7	36.4	43.3	35.1
DeepSeek-VL	0.9	0.0	0.0	0.7	0.0	0.0	0.0	0.2
Qwen2.5-VL-7B	8.6	3.4	1.7	6.4	8.5	1.9	3.8	4.9
Qwen2.5-VL-72B	58.1	58.0	54.9	39.0	51.8	77.9	52.9	56.1
GLM-4V	58.8	29.5	20.3	20.0	43.1	29.2	23.9	32.1
MiMo-VL-SFT (NoThink)	74.4	39.8	37.9	27.7	37.2	50.6	49.7	45.3
MiMo-VL-RL (NoThink)	60.7	33.0	36.3	26.2	31.1	53.9	48.4	41.4
Kimi-VL	47.0	12.5	46.7	34.0	40.2	26.0	33.1	34.2
InternVL2.5-38B	33.3	14.8	9.3	8.5	21.3	27.9	31.9	21.0
InternVL2.5-8B	26.5	15.9	9.3	7.8	18.9	14.3	20.4	16.2
Molmo	7.7	8.0	6.6	8.5	4.9	6.5	8.3	7.2
Molmo2	34.2	13.6	16.5	27.7	34.8	20.8	26.1	24.8
Qwen3-VL-8B	48.7	33.0	35.2	29.8	35.4	57.8	41.4	40.2
Qwen3-VL-32B	93.2	55.7	69.8	57.5	65.2	77.3	63.7	68.9
InternVL3.5-8B	33.3	26.1	13.2	5.7	9.8	13.6	20.4	17.4
InternVL3.5-38B	72.7	20.4	28.6	41.1	75.6	40.3	50.3	47.0
Average	49.2	30.6	28.9	27.7	38.6	35.7	36.3	35.3