Model	Macro Avg. (%)	Overall Acc. (%)	Puzzle Category Accuracy (%)					
			BWB	PSAC	SRO	SR	SP	TLN
GPT$_{5}$	**39.6**	**37.7**	29.4	**38.3**	48.0	29.9	**33.8**	**58.5**
MiMo$_{\text{VL-7B-RL-2508}}$	33.7	29.1	20.6	27.7	48.0	34.5	27.6	43.9
Ovis$_{\text{2.5-9B}}$	32.3	28.8	26.5	27.1	48.0	**41.4**	29.0	22.0
VL-Rethinker$_{\text{7B}}$	32.2	26.1	26.5	24.4	**52.0**	29.9	24.1	36.6
Qwen$_{\text{2.5-VL}}$	31.7	28.6	20.6	28.0	48.0	28.7	28.3	36.6
SkyWork$_{\text{R1V3-38B}}$	31.3	26.9	32.4	25.6	40.0	24.1	29.0	36.6
Kimi$_{\text{VL-A3B-Thinkin}}$	30.5	27.9	26.5	27.3	48.0	28.7	28.3	24.4
Eagle$_{\text{2.5-8B}}$	29.0	27.0	26.5	27.0	44.0	31.0	23.5	22.0
GLM$_{\text{4.1V-9B-Thinkin}}$	28.8	27.6	32.4	27.3	28.0	28.7	26.9	29.3
MiniCPM$_{\text{V-4.5}}$	28.8	26.2	17.7	25.4	44.0	27.6	26.2	31.7
Intern$_{\text{VL-2.5-78B}}$	28.6	27.1	26.5	26.8	40.0	26.4	27.6	24.4
GPT$_{\text{o3}}$	27.8	25.0	32.4	23.9	28.0	27.6	25.5	29.3
Yi$_{\text{VL-34B}}$	27.5	26.4	29.4	26.3	32.0	27.6	25.5	24.4
NVILA$_{\text{15B}}$	27.4	25.5	29.4	24.2	36.0	27.6	30.3	17.1
Idefics$_{\text{2-8B}}$	26.6	24.7	29.4	24.2	32.0	27.6	24.1	22.0
MM-Eureka$_{\text{Qwen-32B}}$	26.4	26.3	23.5	26.5	28.0	21.8	26.9	31.7
Phi$_{\text{3.5-vision-instruct}}$	26.0	24.0	26.5	22.6	32.0	21.8	31.0	22.0
Pixtral$_{\text{12B-2409}}$	25.5	26.5	17.7	26.3	24.0	21.8	31.7	31.7
CogVLM$_{\text{2-Llama3-19}}$	23.0	22.2	**38.2**	21.8	16.0	25.3	22.1	14.6
DeepSeek$_{\text{VL2}}$	22.3	21.3	23.5	20.2	28.0	25.3	24.8	12.2
MMaDA$_{\text{8B-MixCoT}}$	20.0	25.3	14.7	26.7	12.0	26.4	25.5	14.6