Table 2: Evaluation of reasoning accuracy between methods on two metrics.

Method	Token ↑	Semantic ↑
PAR	58.2	0.63
EmbodiedGPT	65.9	0.68
PERIA (ours)	**97.6**	**0.98**
- w/o perceive pretrain	80.2	0.83
- w/o vision planning	83.7	0.79

Table 3: Comparisons of FID (↓) between methods on three task domains.

Methodology	Blocks	Letters	VIMA
SuSIE (+oracle)	18.9	18.1	19.4
CoTDiffusion	13.1	15.8	17.6
PERIA (ours)	**10.2**	**13.5**	**11.4**
- w/o alignment	12.3	14.2	15.9