

Table 1: Quantitative Results: CLIP Text-Image Similarity Score on full scenes.

Methods	CLIP Text-Image Similarity Score		
	ViT-B/16 $\uparrow$	ViT-B/32 $\uparrow$	ViT-L/14 $\uparrow$
DreamFusion (Poole et al. 2023)	0.324 $\pm 0.035$	0.310 $\pm 0.029$	0.272 $\pm 0.032$
ProlificDreamer (Wang et al. 2024)	0.319 $\pm 0.042$	0.306 $\pm 0.038$	0.274 $\pm 0.046$
GaussianDreamer (Yi et al. 2024)	0.315 $\pm 0.029$	0.310 $\pm 0.030$	0.271 $\pm 0.039$
LucidDreamer (Liang et al. 2024)	0.309 $\pm 0.035$	0.300 $\pm 0.028$	0.252 $\pm 0.027$
GraphDreamer (Gao et al. 2024)	0.326 $\pm 0.040$	0.319 $\pm 0.037$	0.274 $\pm 0.042$
GALA3D (Zhou et al. 2024)	0.310 $\pm 0.035$	0.282 $\pm 0.033$	0.279 $\pm 0.042$
Trellis (Xiang et al. 2024)	0.314 $\pm 0.030$	0.293 $\pm 0.033$	0.267 $\pm 0.030$
DreamFragment (Ours)	<b>0.335 <math>\pm 0.033</math></b>	<b>0.336 <math>\pm 0.038</math></b>	<b>0.290 <math>\pm 0.032</math></b>

Table 2: Quantitative Results: CLIP Text-Image Similarity Score on separate instances.

Methods	CLIP Text-Image Similarity Score		
	ViT-B/16 $\uparrow$	ViT-B/32 $\uparrow$	ViT-L/14 $\uparrow$
GraphDreamer (Gao et al. 2024)	0.292 $\pm 0.042$	0.285 $\pm 0.042$	0.239 $\pm 0.043$
GALA3D (Zhou et al. 2024)	0.295 $\pm 0.023$	0.283 $\pm 0.028$	0.249 $\pm 0.025$
DreamFragment (Ours)	<b>0.303 <math>\pm 0.018</math></b>	<b>0.299 <math>\pm 0.020</math></b>	<b>0.260 <math>\pm 0.023</math></b>

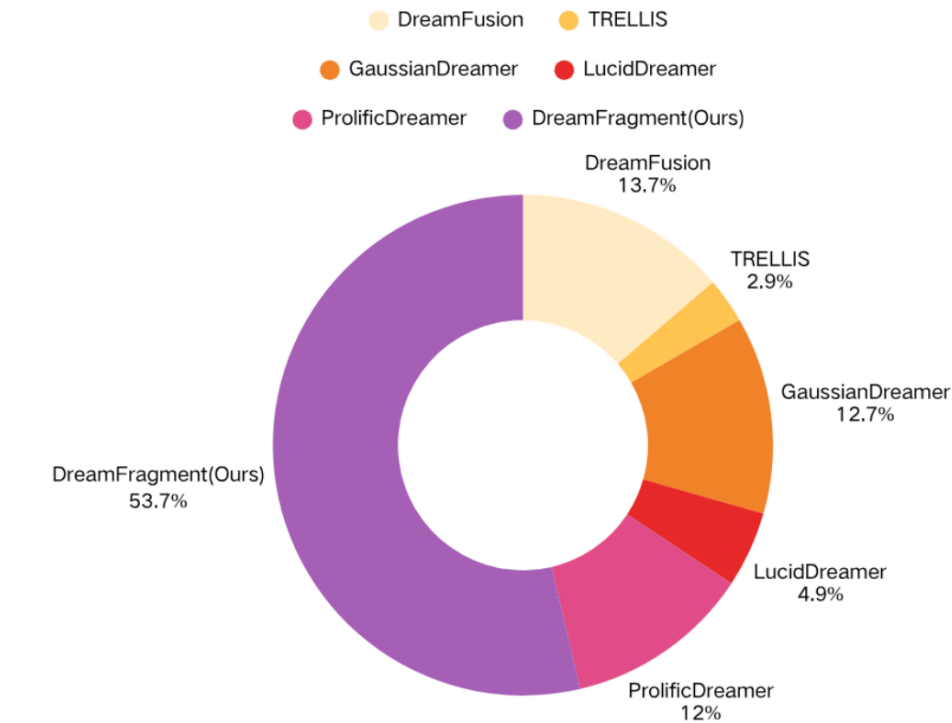


Figure 1: Evaluation on Overall Quality (Two objects)

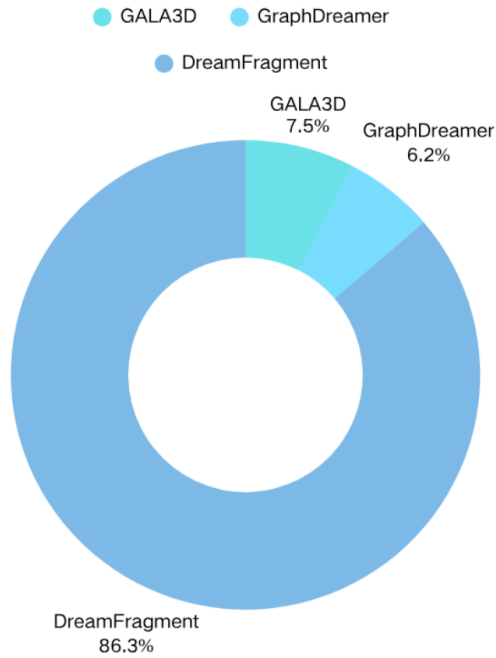


Figure 3: Evaluation on Overall Quality (Multi-objects)

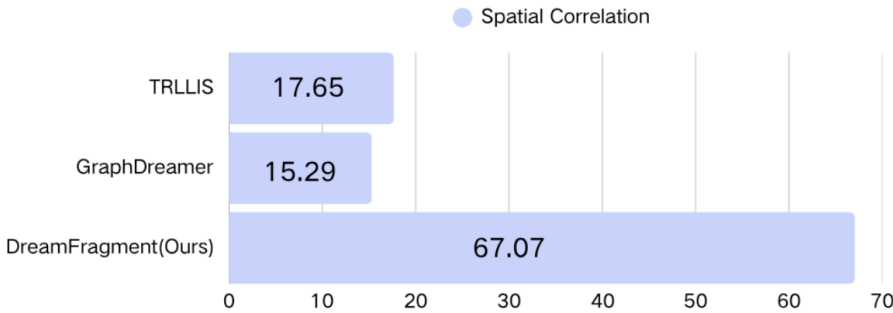


Figure 4: Evaluation on Spatial Correlation.

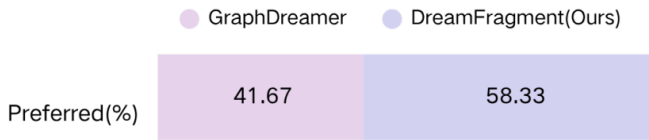


Figure 2: Long&complex Text Prompt Overall Quality