



Figure 1: Visualization of the gradient from the $\mathcal{L}_{i2t}$ on the generated image.

Table 1: T2I-CompBench result. The best score is in **blue**, with the second-best score in **green**.

Model	Attribute Binding			Object Relationship		Complex $\uparrow$
	Color $\uparrow$	Shape $\uparrow$	Texture $\uparrow$	Spatial $\uparrow$	Non-Spatial $\uparrow$	
StructureDiffusion	0.4990	0.4218	0.4900	0.1386	0.3111	0.3355
Composable Diffusion	0.4063	0.3299	0.3645	0.0800	0.2980	0.2898
Attend-and-Excite	0.6400	0.4517	0.5963	0.1455	0.3109	0.3401
TokenCompose	0.5055	0.4852	0.5881	0.1815	0.3173	0.2937
PixArt- α	0.6690	0.4927	0.6477	0.2064	0.3197	0.3433
Playground-v2	0.6208	0.5087	0.6125	0.2372	0.3098	0.3613
SD1.5	0.3758	0.3713	0.4186	0.1165	0.3112	0.3047
CoMat-SD1.5 (Ours)	0.6734	0.5064	0.6243	0.2073	0.3166	0.3575
	(+0.2976)	(+0.1351)	(+0.2057)	(+0.0908)	(+0.0054)	(+0.0528)
SDXL	0.5879	0.4687	0.5299	0.2131	0.3119	0.3237
CoMat-SDXL (Ours)	0.7827	0.5329	0.6468	0.2428	0.3187	0.3680
	(+0.1948)	(+0.0642)	(+0.1169)	(+0.0297)	(+0.0068)	(+0.0443)

Table 2: Extended user study result

Model	Entities	Attributes	Relations	Global
SD1.5	0.76	0.62	0.77	0.71
CoMat-SD1.5	0.90	0.81	0.84	0.81

Table 3: Numeracy scores for different models

Model	Numeracy
SD1.5	0.4627
CoMat-SD1.5	0.5211
SDXL	0.4988
CoMat-SDXL	0.5291