Prompt	Prompt-Image Alignment ↑			Safe Generation Rate ↓					
	CLIPScore	VQAScore	A-Avg.	GPT-4o	DeepSeek-V3	Qwen2.5-VL	Q16	FLUX-Filter	S-Avg.
I2P	<u>0.7319</u>	<u>0.6843</u>	<u>0.7082</u>	0.8241	0.8226	0.8258	0.8483	0.7418	0.8125
Sneakyprompt	0.6662	0.6246	0.6454	0.7711	0.7773	0.7741	<u>0.8201</u>	<u>0.7445</u>	0.7774
MMA Diffusion	0.6663	0.5658	0.6161	<u>0.7299</u>	<u>0.7312</u>	<u>0.7295</u>	0.8204	0.7059	<u>0.7434</u>
Misbinding (Ours)	**0.8257**	**0.7698**	**0.7977**	**0.6987**	**0.6800**	**0.6702**	**0.7671**	**0.6730**	**0.6951**

Table 4: Evaluating different baselines on Prompt-Image Alignment and Safe Generation Rate. **Bold** and <u>underline</u> indicate the best and second-best, respectively (for attack prompts, higher Prompt-Image Alignment and lower Safe Generation Rate indicate superior attack effectiveness).