Author Response for Submission3942: MuDI

Figure R1: **Qualitative comparison** of DreamBooth and MuDI with cartoon style characters.

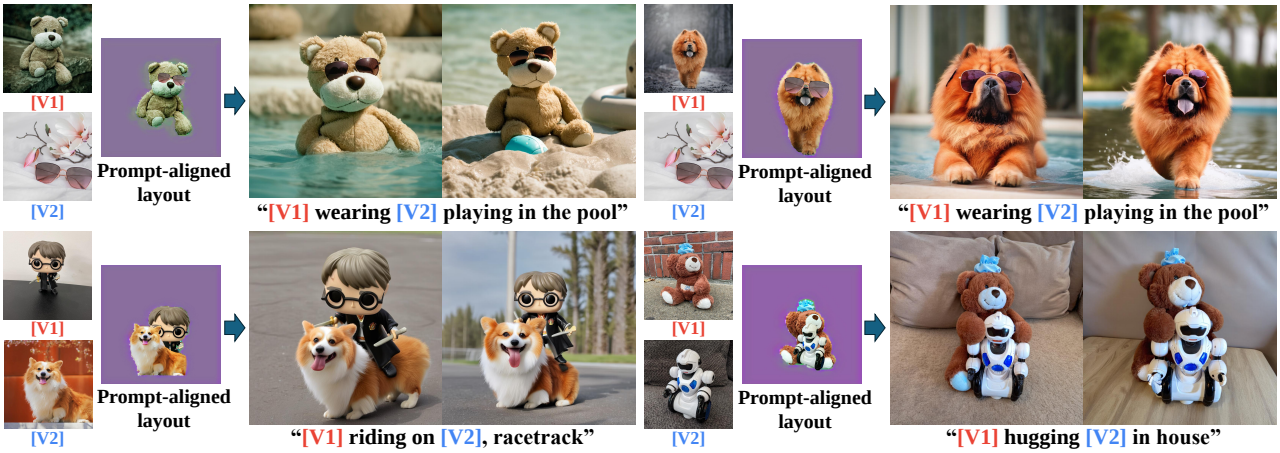
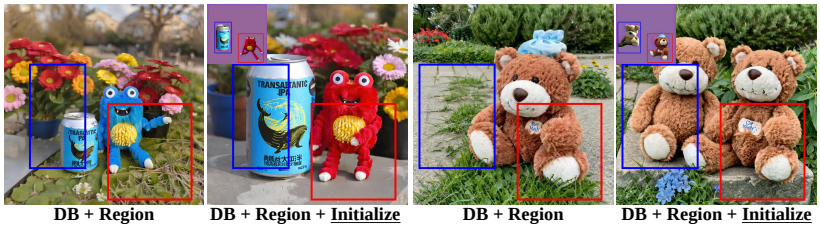


Figure R2: **Examples of rich semantic interaction between subjects** using prompt-aligned layout.

Figure R3: **Seg-Mix does not affect the background of the generated images or text alignment**. (a) Inference with various places. (b) Inference with various styles.

Figure R4: FastComposer fails to personalize Corgi and Chow Chow.



Method	Multi-Subject Fidelity		Text Fidelity	
	D&C-DS↑	D&C-DINO↑	ImageReward†↑	CLIPs†↑
DB	0.371	0.388	0.579	0.255
DB+R	0.340	0.379	0.349	0.245
DB+R+I	0.516	0.513	0.307	0.224
MuDI (Ours)	**0.637**	**0.610**	**0.770**	**0.263**

Figure R5: Results of applying inference initialization for DreamBooth combined with region control. **(Left) Generated images with and without initialization** where the red and blue box denotes the layout of the subjects, and the purple image at the left-top corner visualizes the initialization latent. **(Right) Quantitative comparison. R** denotes Region Control and **I** denotes our initialization.