



Multitwine: Multi-Object Compositing with Text and Layout Control

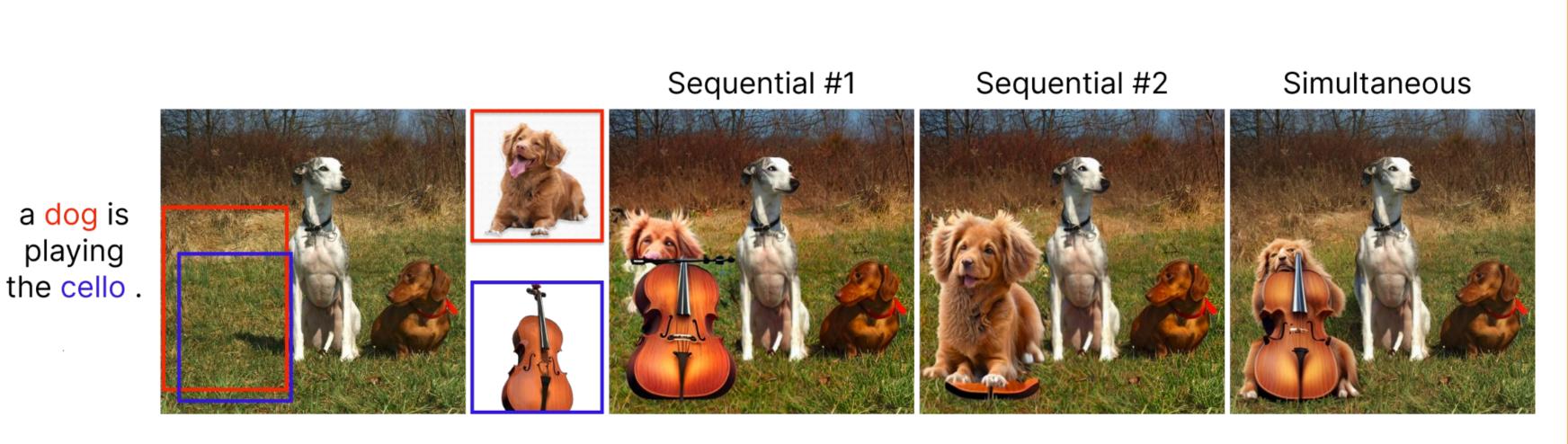
Gemma Canet Tarrés¹, Zhe Lin², Zhifei Zhang², He Zhang², Andrew Gilbert¹, John Collomosse^{1,2}, Soo Ye Kim²

¹ University of Surrey, ² Adobe Research





Motivation



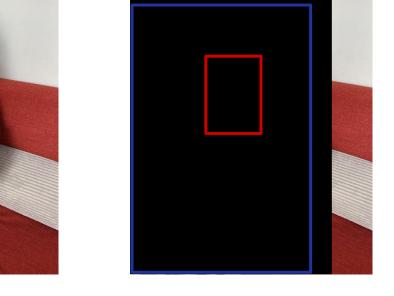
Simultaneous multi-object compositing:

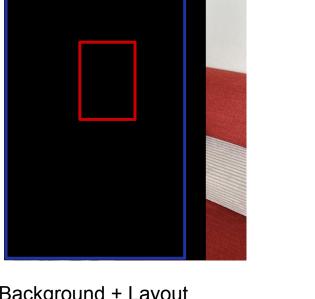
- Allows simultaneous reposing and interaction
- Ensures visual coherence
- Automatically generates props

Training Data

Paired training criteria:

- 1. Diverse Interactions
- Varied Prompts Styles
- 3. Multiview Object Images
- 4. High-Quality Images
- Large Dataset Size

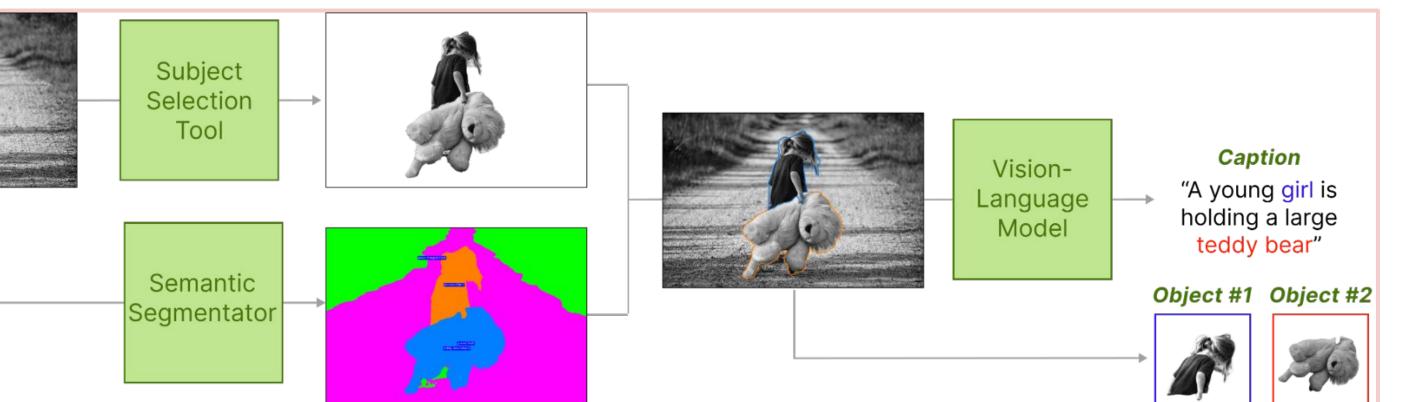


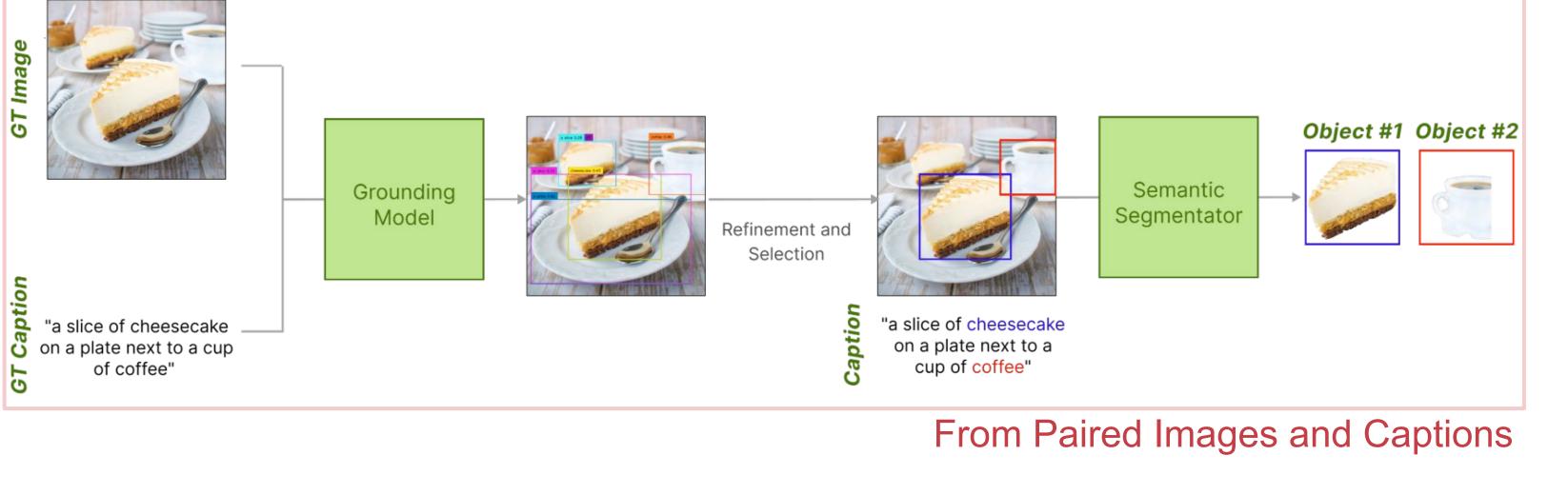


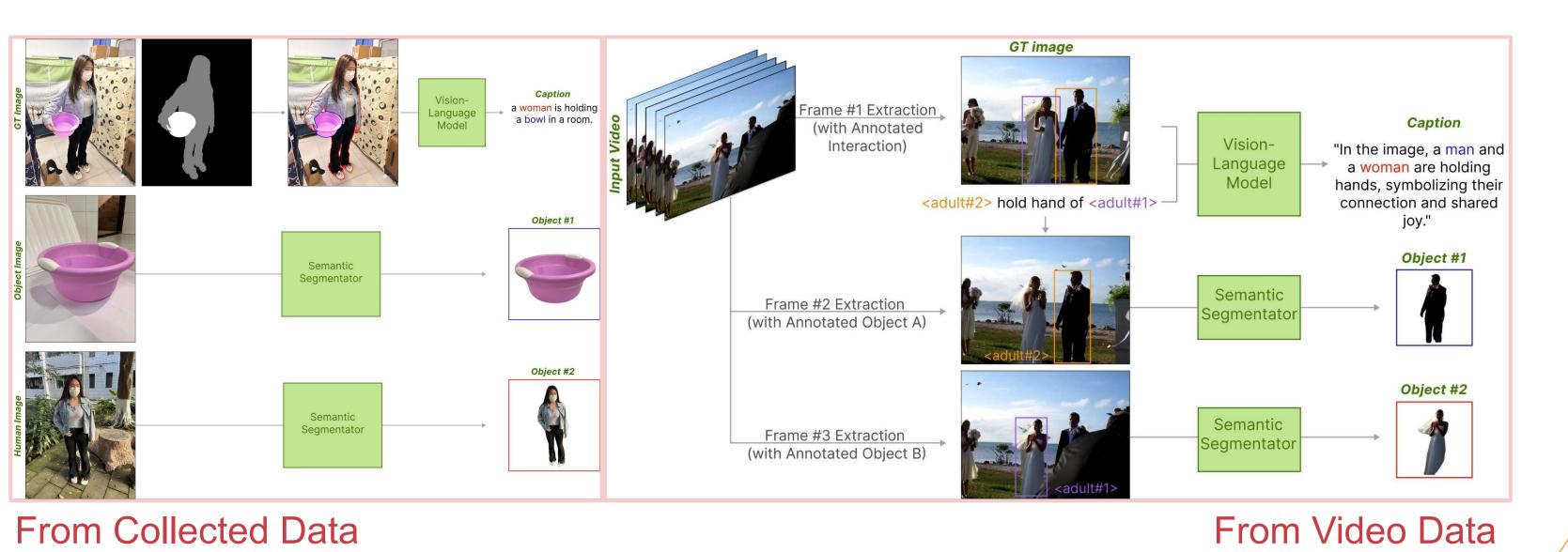




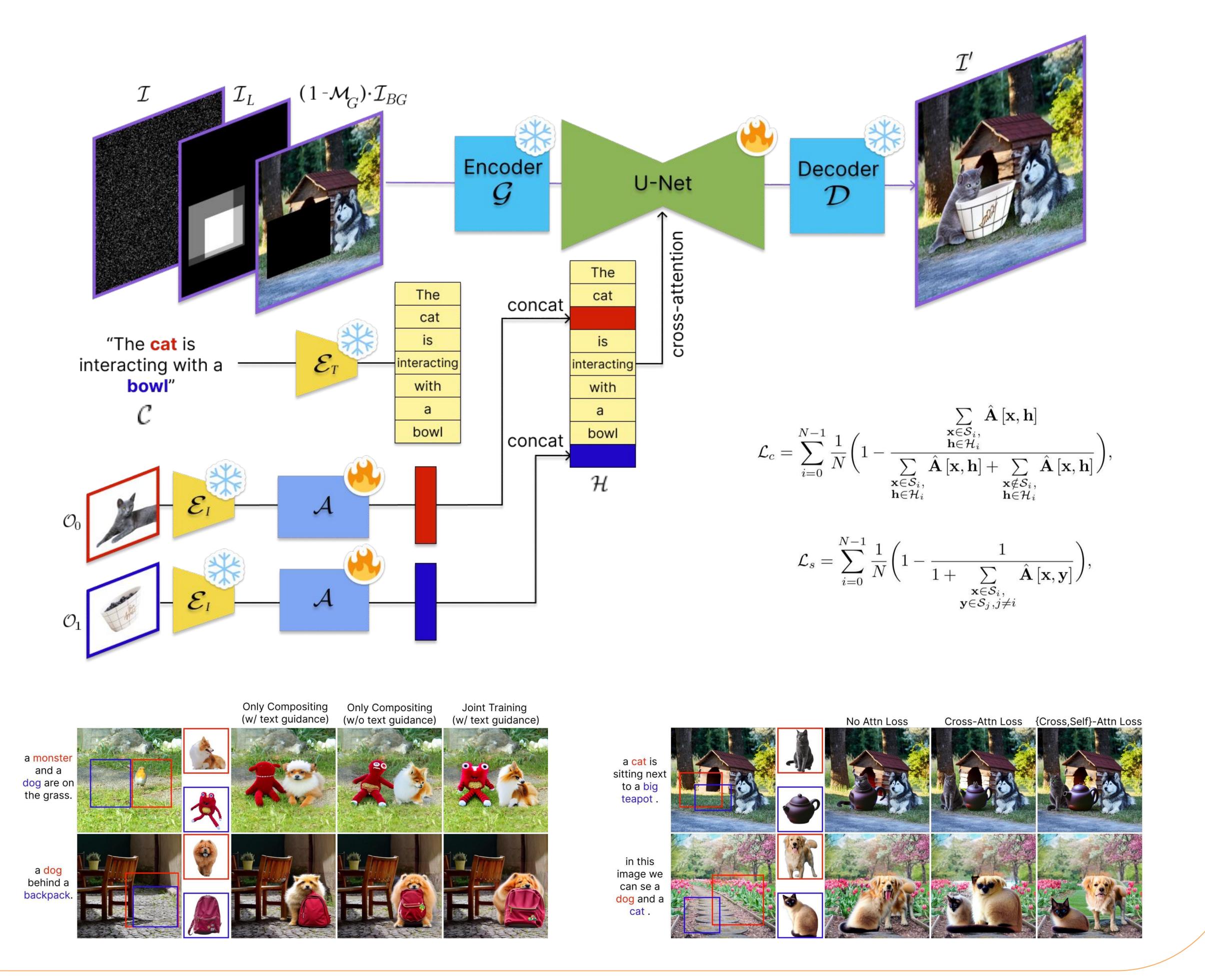








Training Methodology



Applications and Controllability



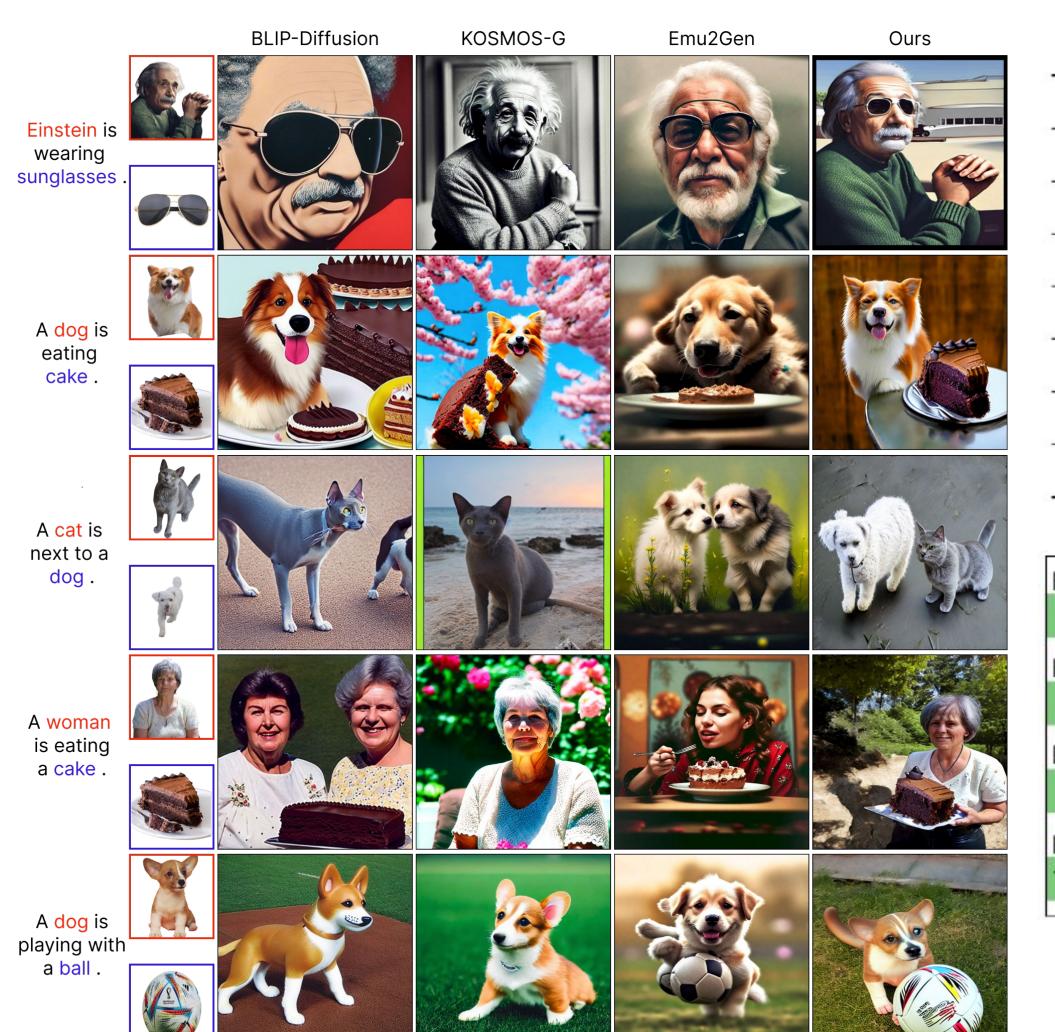
Multi-Object Compositing Comparison to State of the Art

Method	DreamBooth		MultiComp -overlap		MultiComp -nonoverlap	
	CLIP-I↑	DINO↑	CLIP-I↑	DINO↑	CLIP-I↑	DINO↑
PbE [48]	0.778	0.799	0.693	0.383	0.720	0.423
ControlCom [50]	0.743	0.705	0.707	0.478	0.740	0.543
AnyDoor [7]	0.806	0.836	0.727	0.520	0.763	0.593
IMPRINT [40]	0.830	0.889	0.713	0.525	0.739	0.576
TOTB [43]	0.809	0.856	0.716	0.485	0.740	0.531
Ours	0.803	0.892	0.741	0.532	0.768	0.579
w/o text	0.816	0.903	0.729	0.505	0.754	0.548
sequential w/ text		(1)	0.729	0.517	0.760	0.583
sequential w/o text	-	-	0.723	0.510	0.756	0.578

Ours (Interaction) Ours (Quality)	Baseline (Interaction) Baseline (Quality)	
AnyDoor		Ours
35.3	64.7	
26.3	73.7	
ControlCom		Ours
33.3	66.7	
37.5	62.5	
IMPRINT		Ours
47.1	52.9	
30.0	70.0	9
PaintByExample		Ours
45.0	55.0	
2.9	97.1	
Thinking Outside the BBox		Ours
41.7	58.3	3.3.3.
36.0	64.0	



Customization Comparison to State of the Art



Method	CLIP-I↑	DINO↑	CLIP-Igl↑	$\mathbf{DINO} gl \uparrow$	CLIP-T $gl \uparrow$		
	Inpu	t: Text, Ob	ject Images		-		
BLIP-Diffusion [22]	n -	n=:	0.675	0.455	0.249		
KOSMOS-G [28]	1,-	-	0.704	0.465	0.279		
UNIMO-G [23]	-	-	0.699	0.485	0.293		
	Input: Te	ext, Object	Images, Layoi	ıt			
Emu2-Gen [42]	0.595	0.414	0.616	0.434	0.287		
Ours	0.783	0.599	0.688	0.454	0.308		
Emu2Gen		Text Alig	nment		Ours		
47.	5			52.5			
Emu2Gen		Layout Alignment					
23.0		77.0					
Emu2Gen		Objects Ali	gnment		Ours		
35.0			6	5.0			
Emu2Gen	В	ackground i	Alignment		Ours		
			The second section is a second		Salt Annies Salt		

Visualizations









