



Project Page



Dataset

Scalable 3D Captioning with Pretrained Models

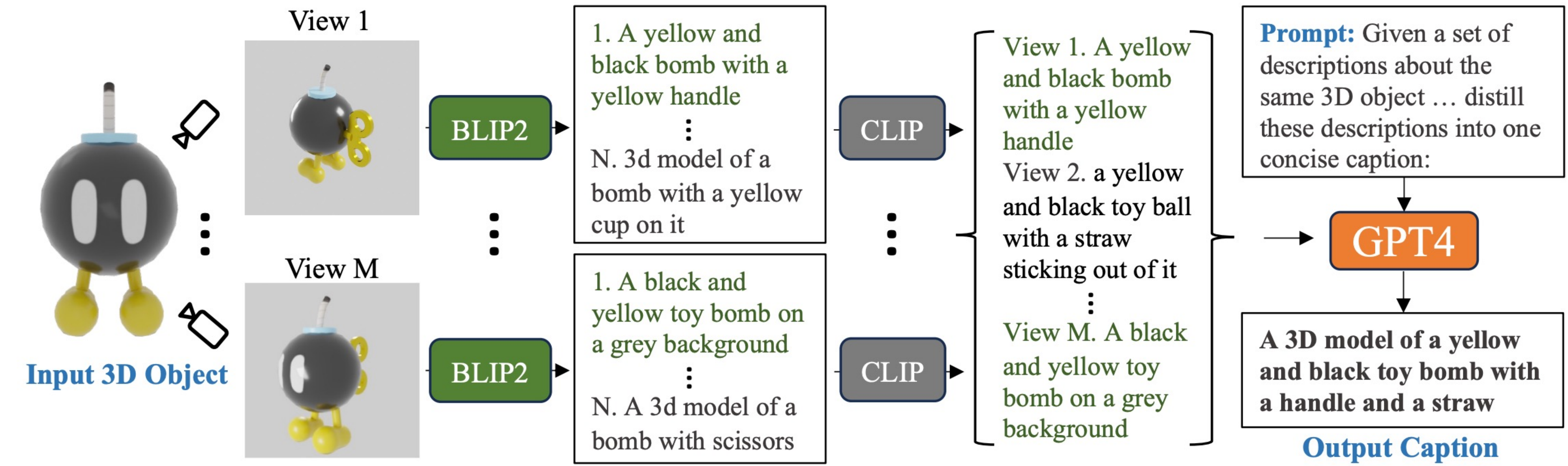
Tiange Luo*, Chris Rockwell*, Honglak Lee†, Justin Johnson†
University of Michigan, LG AI Research



Task: 3D Captioning

3D model of a sakura soft drink can with purple and yellow gradient, Japanese writing, and purple flowers.	A 3D model of a blue grand piano with spikes and sharp teeth resembling a shark mouth.	A 3D model of a metal cube featuring a skull, pizza, and various stickers.	3D model of a yellow Pikachu-themed Pokémon ball with a black and gold stripe and lightning bolt.
3D model of Notre Dame Cathedral, a Gothic cathedral with spires in Paris.	Loki bust 3D model featuring a green and yellow horned helmet.	A 3D model featuring a basketball hoop, ball, racquet, bowling ball, stand, and pin.	3D model of a purple and green Halloween spider bowl on a metal stand, containing purple liquid.
3D model of an armored character with purple horns and spikes on the back.	3D model of a robotic scorpion with multiple arms and guns.	A cluster of five glass sphere light bulbs suspended from a single thin wire.	L-shaped sectional sofa with a chaise, U-shaped backrest, curved armrests, and a footstool on one side.

Method: Cap3D



Cap3D is better, cheaper, and faster than crowdsourced annotation

Method	A/B Human Testing Win % (Tie %)	Cost per 1k Objects	Annotation Speed
Human	37.8% ± 0.5% (9.5%)	\$87.18	1.4k / day
Cap3D	52.3% ± 0.5% (9.5%)	\$8.35	65k / day

Apply Cap3D on Objaverse and finetune Text-to-3D models

	Pretrained					Finetuned on Cap3D				
	FID↓	CLIP Score	CLIP R@1	CLIP R@5	CLIP R@10	FID↓	CLIP Score	CLIP R@1	CLIP R@5	CLIP R@10
Ground Truth Images	-	81.6	32.7	55.1	64.3	-	81.6	32.7	55.1	64.3
Point-E (Text-to-3D) [88]	36.1	72.4	6.0	16.2	22.4	32.8	75.6	12.4	28.1	36.9
S. Diff. [22] (CNet) [63]+ [88](Im-to-3D)	54.7	73.6	11.0	23.4	30.0	53.3	74.6	12.4	26.2	33.8
S. Diff. [22] (LoRA) [103]+ [88](Im-to-3D)	54.7	73.6	11.0	23.4	30.0	53.7	74.4	11.6	24.6	31.4

All Cap3D captions and annotated objects have commercial-friendly licenses.

1k Objects Cost Breakdown	
BLIP2	\$3.79
CLIP	\$0.38
GPT4	\$4.18
Cap3D Total Cost	\$8.35