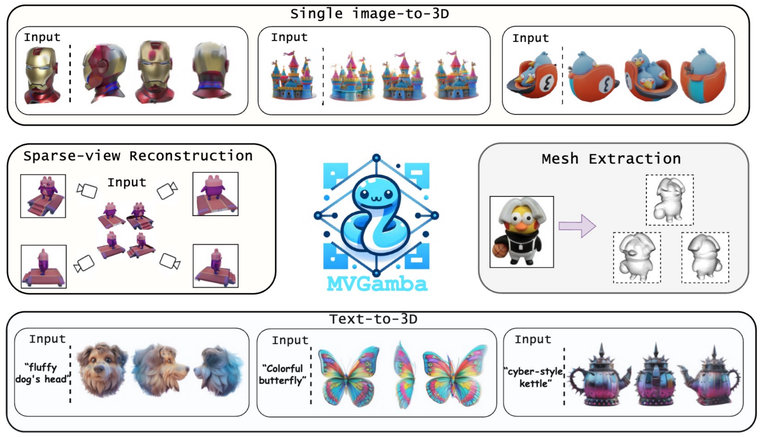
MVGamba: Unify 3D Content Generation as State Space Sequence Modeling

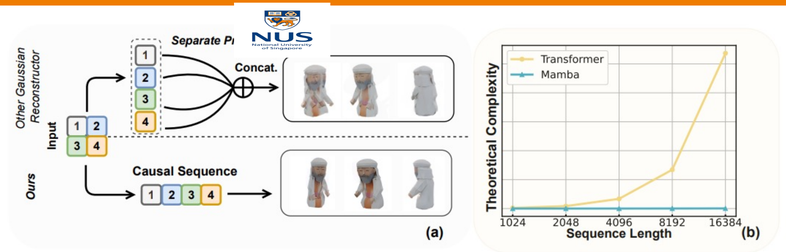
Xuanyu Yi*, Zike Wu*, Qiuhong Shen*, Qingshan Xu, Pan Zhou
Joo-Hwee Lim, Shuicheng Yan ,Xinchao Wang ,Hanwang Zhang,

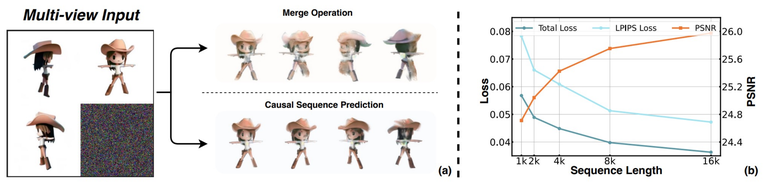
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
NUS National University of Singapore
SMU SINGAPORE MANAGEMENT UNIVERSITY
Agency for Science, Technology and Research SINGAPORE
天工 Skywork

MVGamba



Motivation





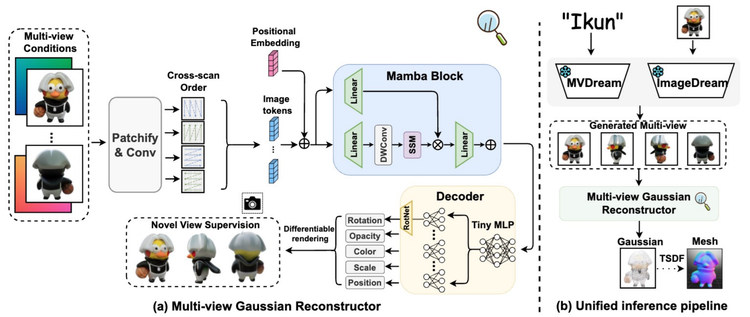
1. Previous works inevitably compromise the integrity of *multi-view information propagation* to manage *computational costs*.

2. The imperfect generation of multi-view images might be unavoidable bottom neck.

3. The computationally efficiency as well as Gaussian representation is far from perfect.

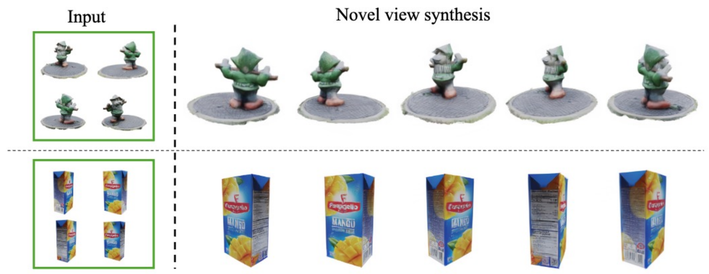
The crux is to generate the integrity of multi-view information while efficiently generating a sufficiently long sequence of Gaussians.

Methods



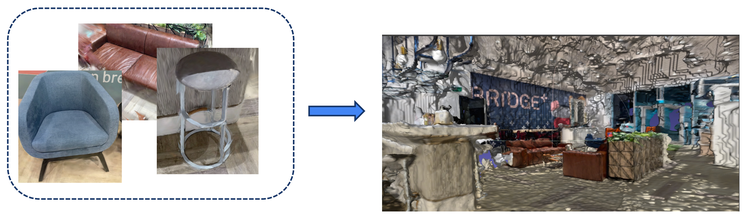
Experiement

Sparse-view Reconstruction



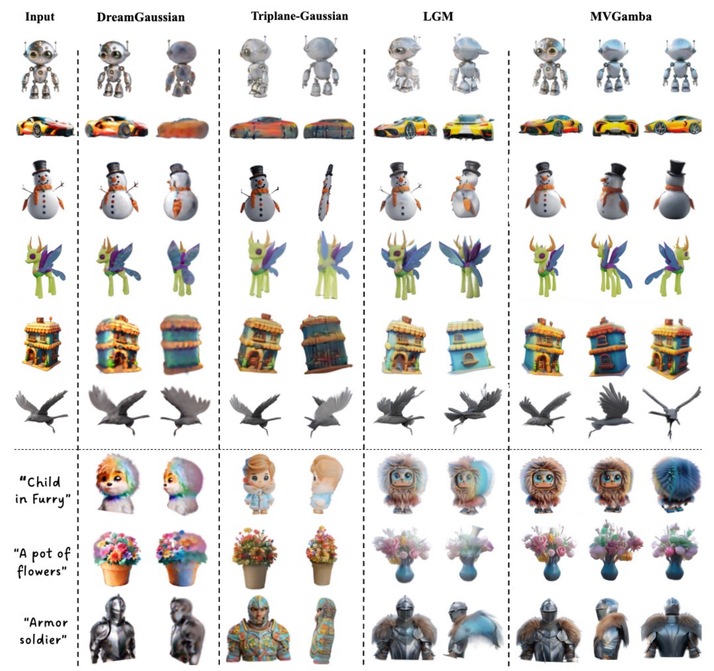
Method	#views	PSNR↑	LPIPS↓	SSIM↑	INF. Time↓	CD↓	VIoU↑
SparseGS [81]	16	22.19	0.162	0.775	34s	-	-
SparseNeuS [80]	16	23.17	0.130	0.814	6s	0.0566	0.3479
LGM [19]	4	24.20	0.112	0.845	0.07s	0.0198	0.4410
MVGamba	4	**26.25**	**0.069**	**0.881**	**0.03s**	**0.0132**	**0.4829**

Engine-friendly Interaction

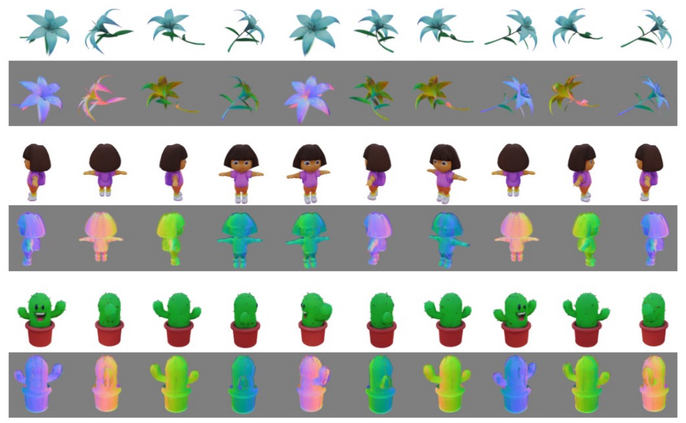


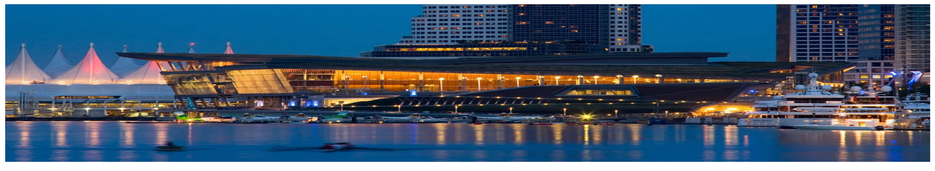
Experiement

Image / Text-to-3D Generation



Normal Map Generation





QR Code to Github Link:

QR Code to Paper Link: