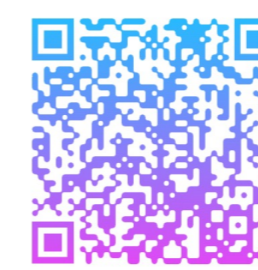


DIFFSPLAT: Repurposing Image Diffusion Models for Scalable 3D Gaussian Splat Generation

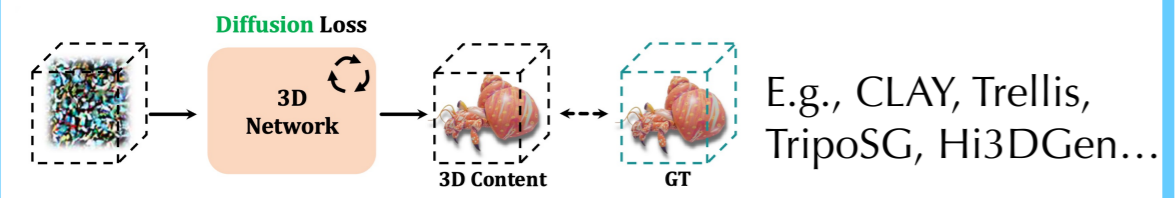


Chenguo Lin¹, Panwang Pan², Bangbang Yang², Zeming Li², Yadong Mu¹

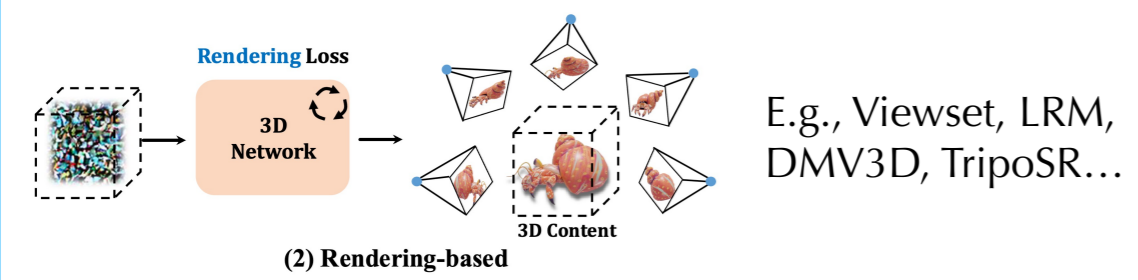
Code: <https://github.com/chenguolin/DiffSplat>



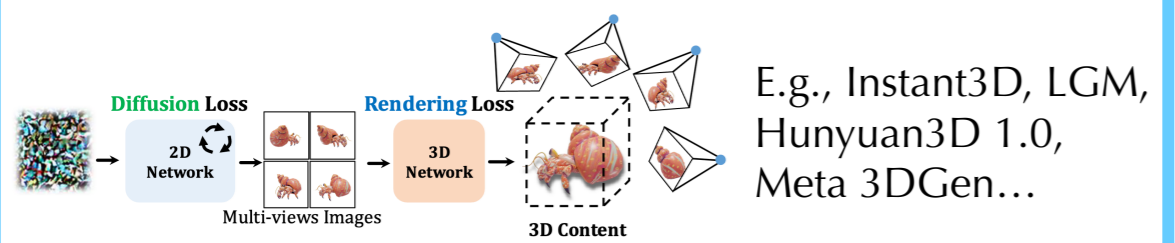
Motivation



Pros: current (2025.04) SOTA, high-quality, 3D coherent
Cons: data limitation, costly training, absence of 2D priors

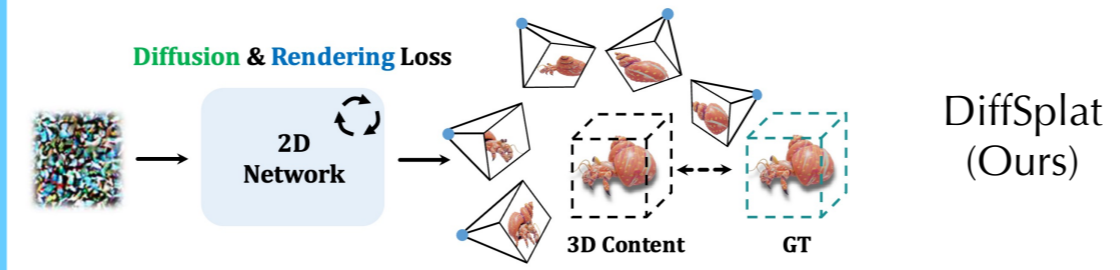


Pros: only need multi-view images for supervision
Cons: unstable training, same as (1)'s cons

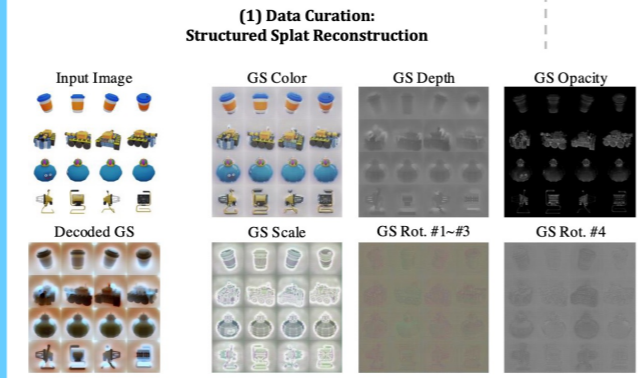
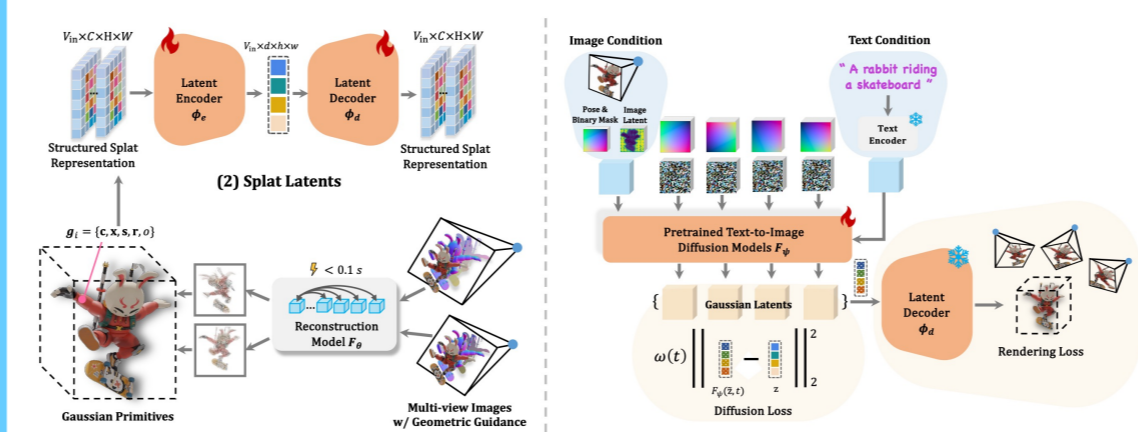


Pros: most popular, leverage 2D priors by multi-view images
Cons: independent two stages, 3D inconsistency

Method



TL;DR: DiffSplat is fine-tuned from any text-to-image diffusion model to directly generate 3DGS attributes.



Visualize 3DGS attributes: a "special image style" !

Applications

Text-conditioned				Image-conditioned	
<i>A beautiful rainbow fish</i>	<i>A bright red fire hydrant</i>	<i>A brown horse in a green pasture</i>	<i>A colorful camping tent in a patch of grass</i>		
<i>A fluffy, orange cat</i>	<i>A green enameled watering can</i>	<i>A green frog on a lily pad</i>	<i>A human skull</i>		
<i>A jar of homemade jam</i>	<i>A lighthouse on a rocky shore</i>	<i>A plush velvet armchair</i>	<i>A red cardinal on a snowy branch</i>		
ControlNet					
Original Object	Input Control	A steampunk robot with brass gears and steam pipes		A cute cartoon robot with oversized eyes	
		<i>A Santa festive plush bear toy</i>	<i>An adorable baby panda</i>		
Text-conditioned Reconstruction					
Input Image	A sculpture			A mask	