MuDG: Taming Multi-modal Diffusion with Gaussian Splatting for Urban Scene Reconstruction

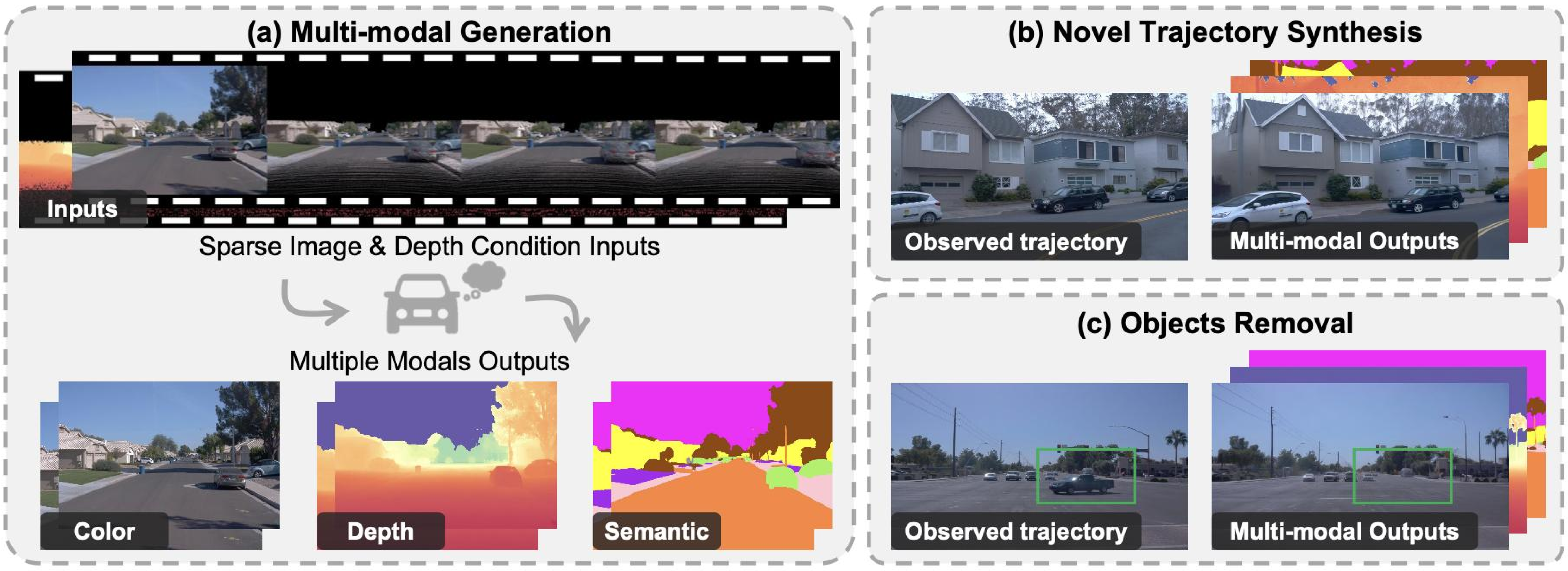
BMVC 2025

Yingshuang Zou*[1,2], Yikang Ding*[1,2], Chuanrui Zhang[1,2], Jiazhe Guo[1,2], Bohan Li[2,4], Xiaoyang Lyu[5], Feiyang Tan[3], Xiaojuan Qi[5], Haoqian Wang†[1]

[1]Tsinghua University [2]Megvii Technology [3]Mach Drive [4]Shanghai Jiao Tong [5]University University of Hong Kong

* Indicates equal contribution

What can MuDG do?



Framework of MuDG

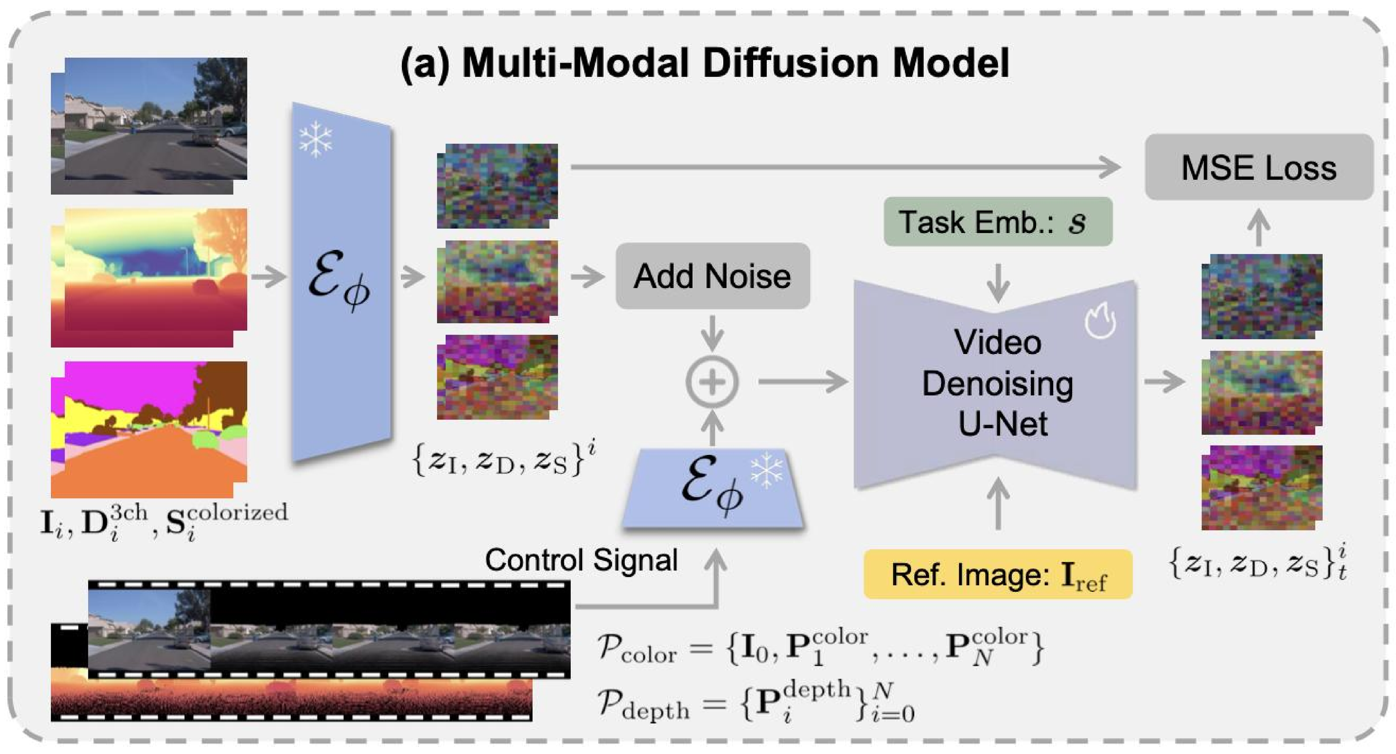
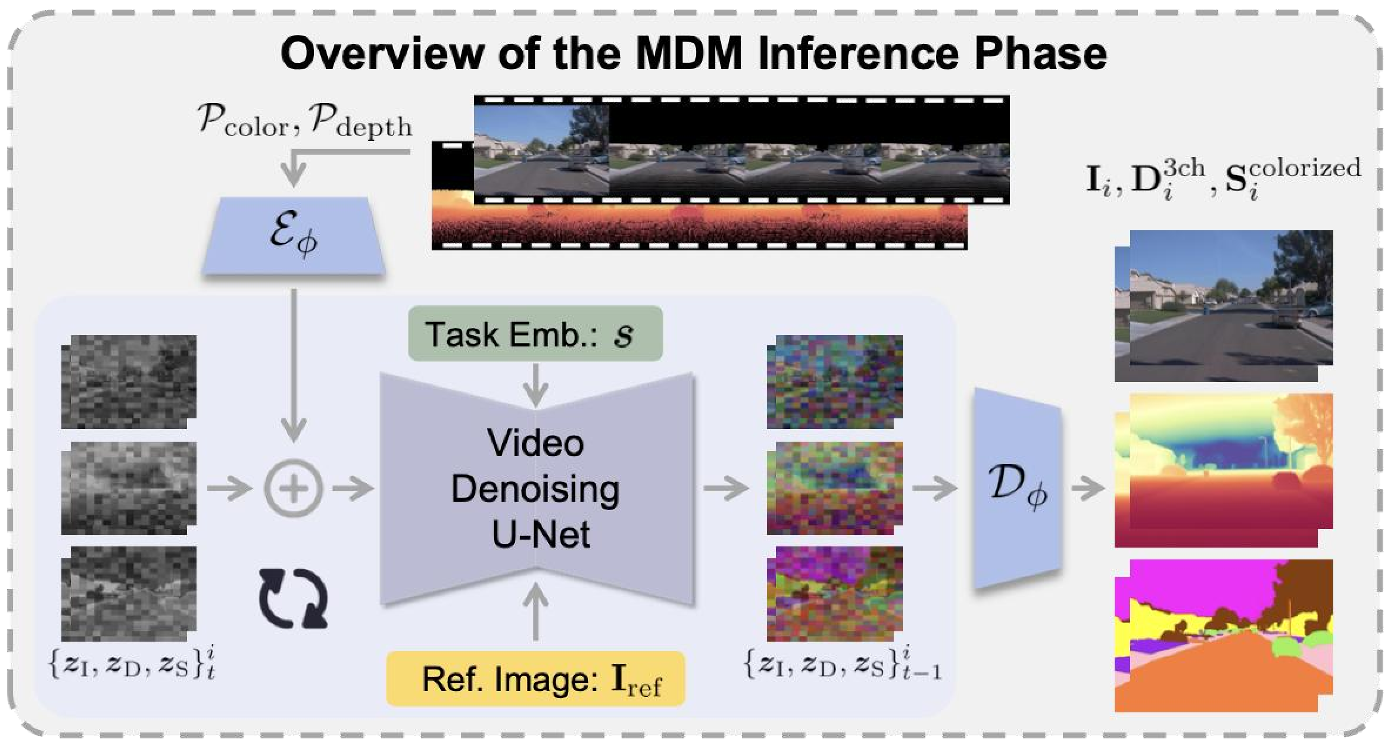


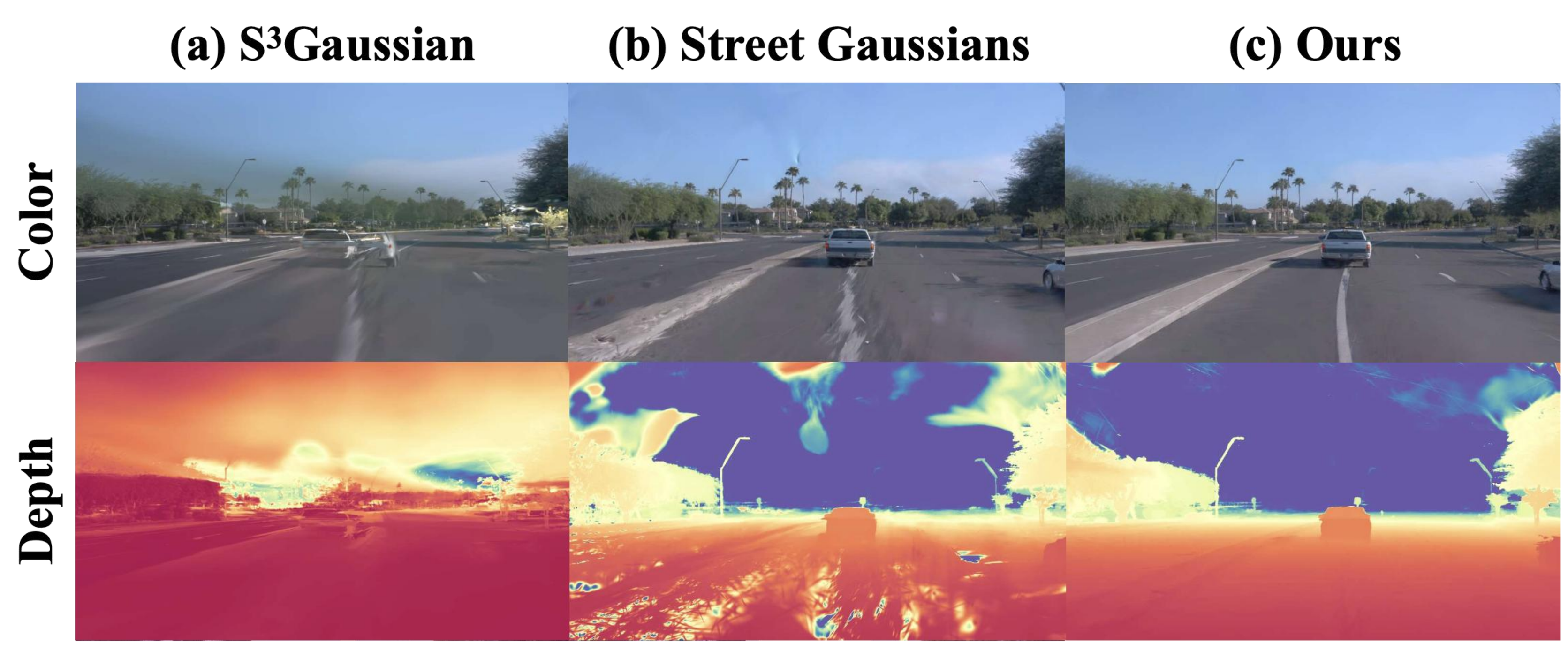
Illustration of Inference Phase



Quantitative Comparison of NVS on the Waymo[1]

Model Outputs

Depth Results



Render Results

	Shift 2m		Shift 3m		Shift 4m	
Method	FID ↓	FVD ↓	FID ↓	FVD ↓	FID ↓	FVD ↓
S^3Gaussian [10]	169.08	2031.97	191.38	2543.47	206.99	2663.56
3DGS* [11]	101.63	1062.57	117.36	1379.68	135.10	1675.16
EmerNeRF [30]	73.58	645.90	93.09	983.94	110.94	1250.56
StreetGS [28]	69.79	<u>563.00</u>	87.71	898.05	100.05	1269.99
FreeVS [22]	92.08	652.09	95.91	784.74	117.68	<u>801.20</u>
Ours-R	**50.02**	671.24	**49.42**	<u>777.78</u>	**56.52**	**770.39**
Ours-S	<u>54.14</u>	**509.34**	<u>58.85</u>	**659.27**	<u>65.44</u>	808.85

[1] Sun P, Kretzschmar H, Dotiwalla X, et al. Scalability in perception for autonomous driving: Waymo open dataset[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2446-2454.