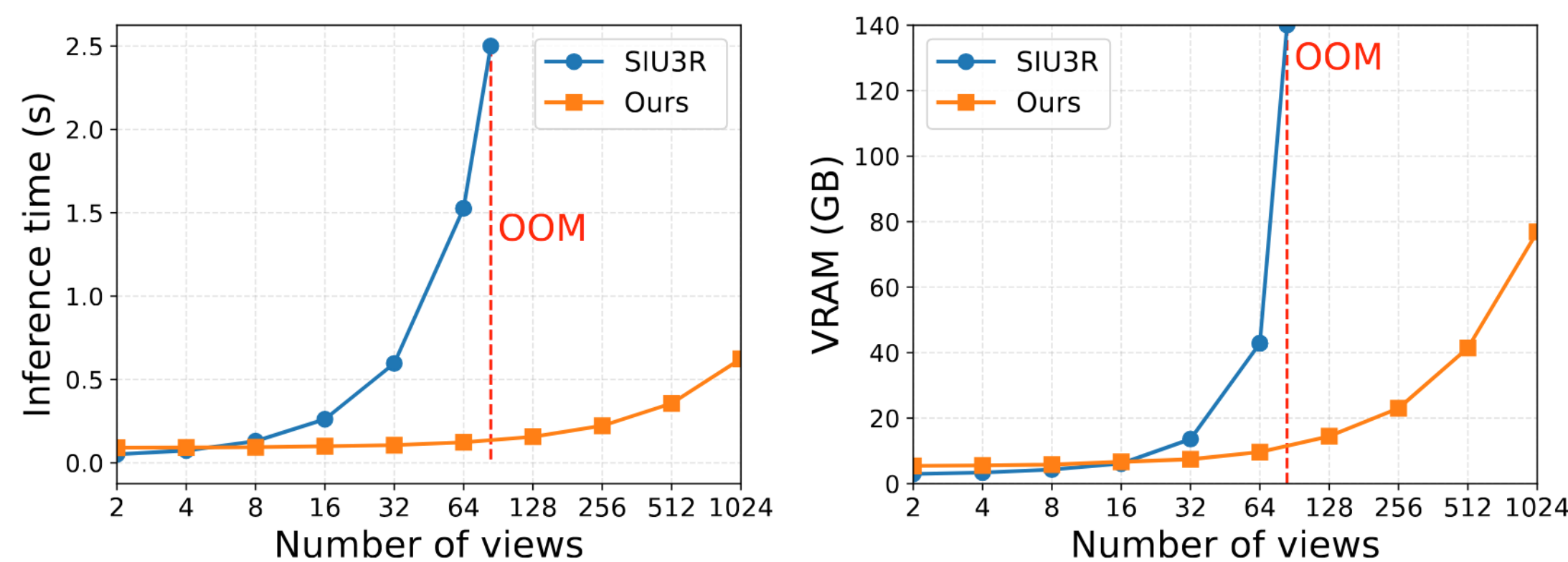


Motivation:

Existing offline-global methods repeatedly recompute interactions over all previous frames, causing runtime and memory to grow rapidly with sequence length. Even on a 140 GB H200 GPU, SIU3R runs out of memory after about 80 frames. This motivates an incremental streaming approach that avoids reprocessing history.



Results:

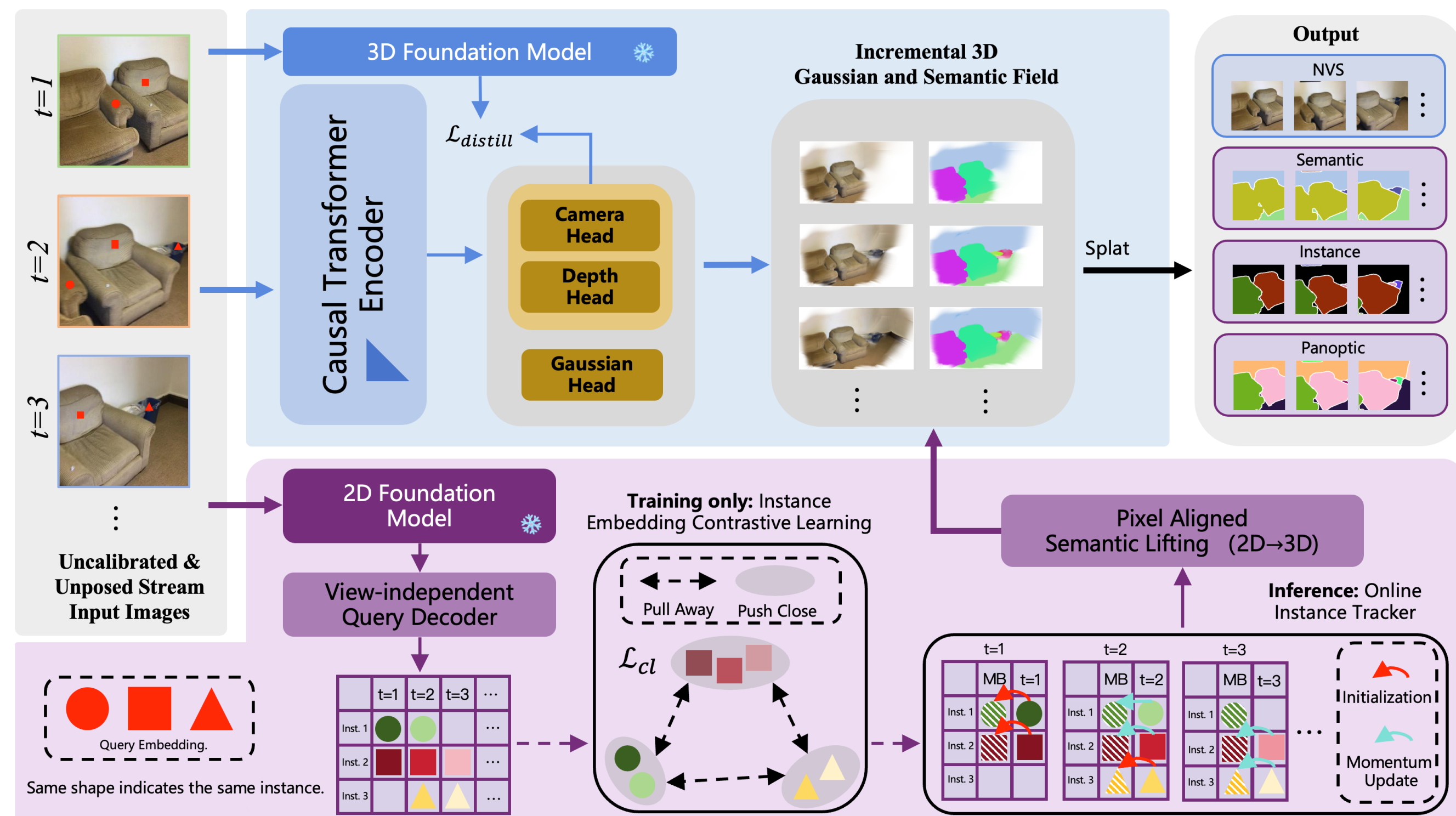
Method	2 views					8 views				
	PSNR↑	SSIM↑	mIoU↑	T-mIoU↑	T-SR↑	PSNR↑	SSIM↑	mIoU↑	T-mIoU↑	T-SR↑
• pixelSplat (Charatan et al., 2024)	24.76	0.804	-	-	-	-	-	-	-	-
• MVSPat (Chen et al., 2024)	23.63	0.784	-	-	-	-	-	-	-	-
• NoPoSplat (Ye et al., 2024)	25.27	0.811	-	-	-	-	-	-	-	-
† Mask2Former (Cheng et al., 2022)	-	-	47.32	42.11	<u>90.17</u>	-	-	45.85	<u>31.93</u>	<u>66.55</u>
† LSeg (Li et al., 2022)	-	-	33.27	-	-	-	-	32.87	-	-
* LSM (Fan et al., 2024)	22.38	0.714	31.31	-	-	-	-	-	-	-
* Uni3R (Sun et al., 2025)	<u>25.44</u>	<u>0.812</u>	32.18	-	-	18.16	0.627	33.75	-	-
* SIU3R (Xu et al., 2025)	25.79	0.819	47.83	44.25	85.07	<u>19.74</u>	<u>0.653</u>	44.78	29.41	62.93
* S2GS (Ours)	24.90	0.810	52.35	44.89	93.73	20.83	0.685	49.53	33.34	82.49

Method	14 views					32 views				
	PSNR↑	SSIM↑	mIoU↑	T-mIoU↑	T-SR↑	PSNR↑	SSIM↑	mIoU↑	T-mIoU↑	T-SR↑
† Mask2Former (Cheng et al., 2022)	-	-	<u>43.32</u>	25.43	55.96	-	-	41.59	25.15	40.91
† LSeg (Li et al., 2022)	-	-	29.17	-	-	-	-	30.46	-	-
* Uni3R (Sun et al., 2025)	16.36	0.583	31.31	-	-	16.74	0.593	32.17	-	-
* SIU3R (Xu et al., 2025)	<u>17.29</u>	<u>0.591</u>	37.38	<u>27.37</u>	50.18	<u>17.82</u>	<u>0.629</u>	39.98	<u>29.39</u>	<u>41.24</u>
* S2GS (Ours)	19.68	0.645	46.64	30.19	76.50	19.92	0.665	48.95	30.01	62.39

Views	SIU3R		Ours	
	Time(s) ↓	PGM(GB) ↓	Time(s) ↓	PGM(GB) ↓
16	0.26	6.14	0.10	6.68
64	1.52	42.89	0.12	9.66
128	OOM	OOM	0.15	14.48
512	OOM	OOM	0.35	41.45
1024	OOM	OOM	0.62	76.88

Method	Iters.	PSNR↑	Time↓
Splat-SLAM ₃₂	1/1/1	9.91	~115s
Splat-SLAM ₃₂	100/100/100	20.88	~345s
S2GS ₃₂	-	19.92	~4s
Splat-SLAM ₂₅₆	1/1/1	8.82	~725s
Splat-SLAM ₂₅₆	100/100/100	14.01	~2505s
S2GS ₂₅₆	-	16.37	~50s

Methods:



- S2GS performs online joint modeling through a geometry-semantic decoupled architecture. A causal geometry stream incrementally maintains geometry and appearance, while an independent semantic stream extracts stable multi-scale representations from a frozen 2D foundation model.
- Streaming-specific stabilization mechanisms, including query contrastive alignment and online instance memory, improve temporal identity consistency and enable open-vocabulary semantic understanding.
- Unlike offline-global methods, S2GS avoids reprocessing historical frames, achieving strong long-horizon scalability.

Visual comparison:

