

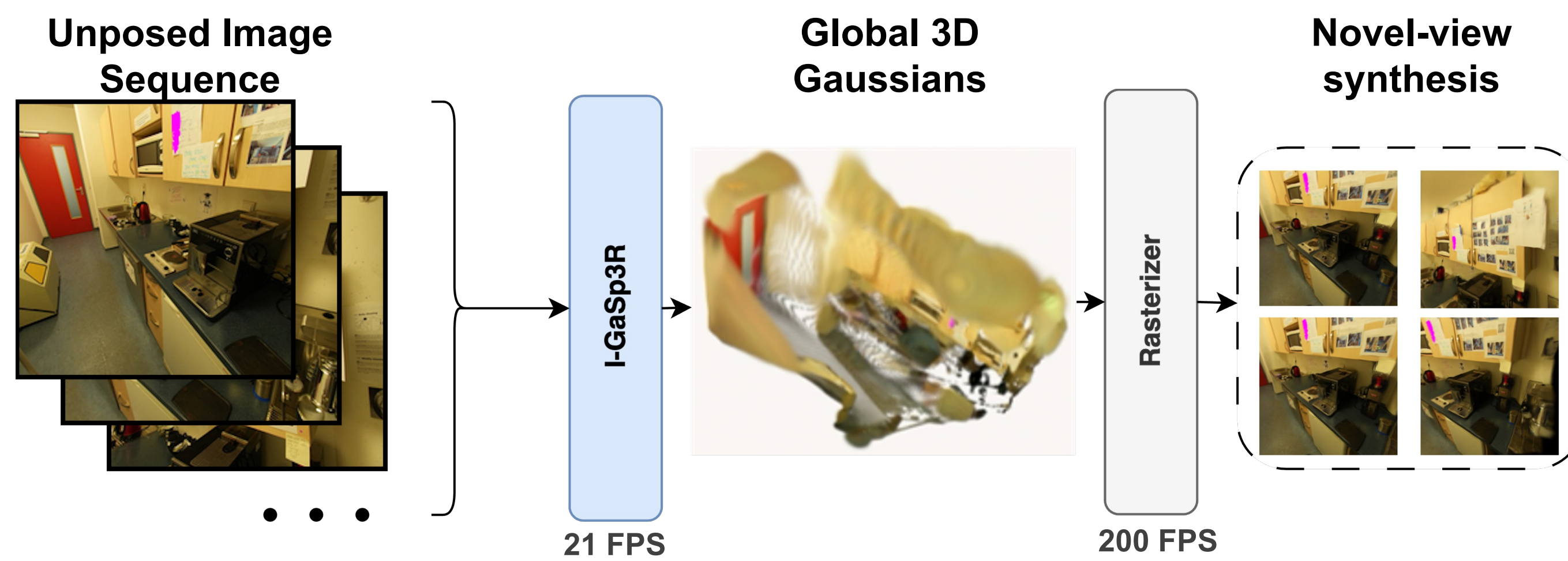
Instant Gaussians Splatting 3D Reconstruction

Osama Salem, Tobias Steinle



Motivation

3D reconstruction often relies on camera poses, limiting adaptability. **I-GaSp3R** removes this requirement with a **pose-free, feed-forward** pipeline that generates Gaussian splats for **novel view reconstruction** from **long image sequences**, enabling high-quality.



Key Contributions

Real-time Feed-forward 3D Gaussian Splatting

- Combines the **spatial memory network** module from Spann3R with the **Gaussian head** from Splatt3R to achieve **real-time feed-forward** 3D Gaussian Splatting reconstruction at 21 FPS on RTX 3080.

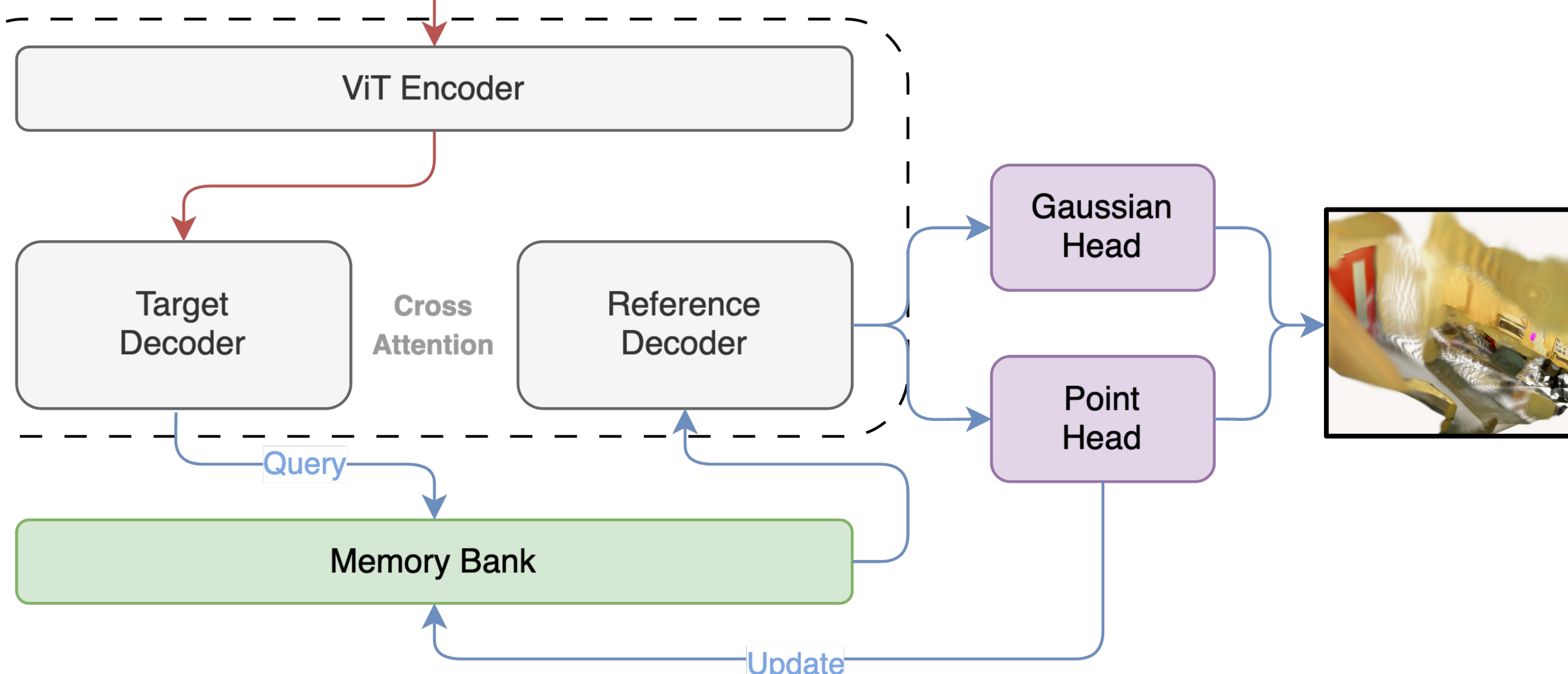
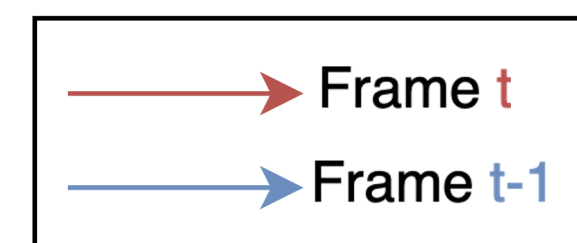
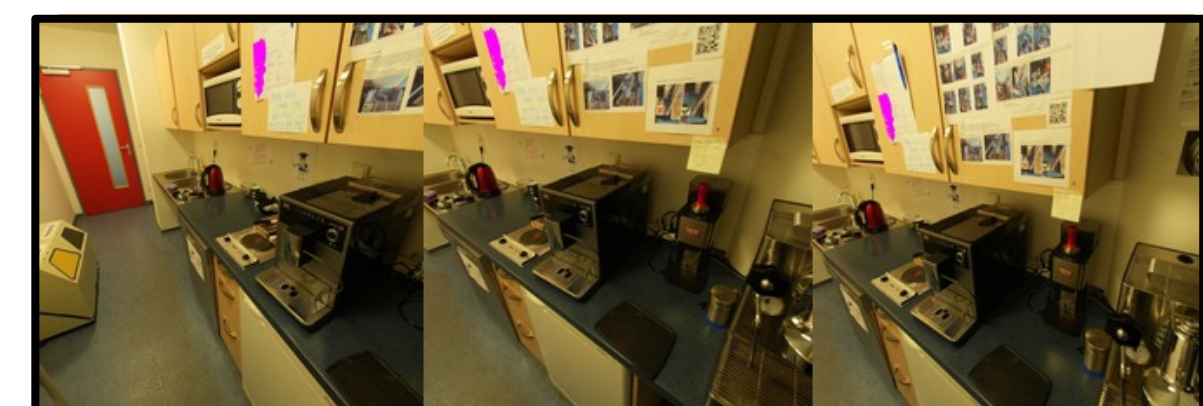
Temporal Consistency

- Introduces a **temporal consistency** loss to ensure **alias-free, stable reconstructions** in long-sequence, feed-forward Gaussian splatting.

Training Stability

- Improves robustness with **masking techniques** (frustum overlap & depth matching) and enhances convergence through **novel regularization and gradient weighting**.

Architecture

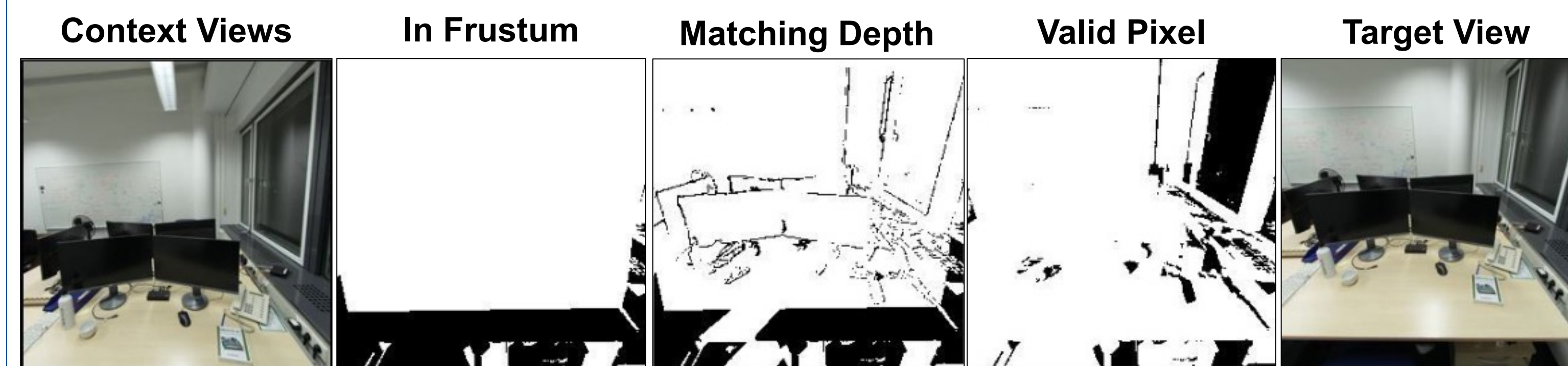


Training & Optimization

Robust Supervision & Optimization

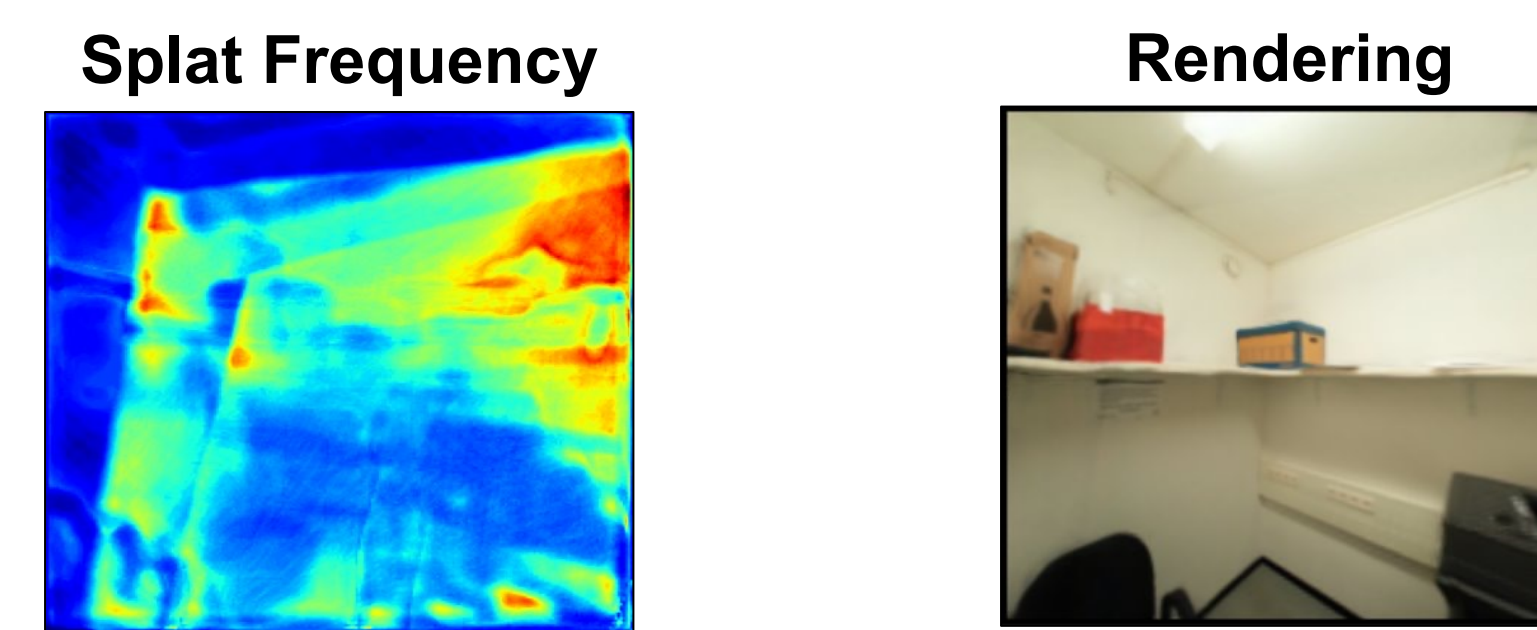
- Photometric** and **perceptual** loss to ensure accurate 3D reconstruction.

$$\mathcal{L}_{\text{MLPIPS}} = \mathcal{L}_{\text{LPIPS}}(M \cdot I_t, M \cdot \hat{I}_t) \quad \mathcal{L}_{\text{MMSE}} = \mathcal{L}_{\text{MSE}}(M \cdot I_t, M \cdot \hat{I}_t)$$



Gradient Weighting

- Balance contributions from different regions to improve convergence.



Stability Enhancements

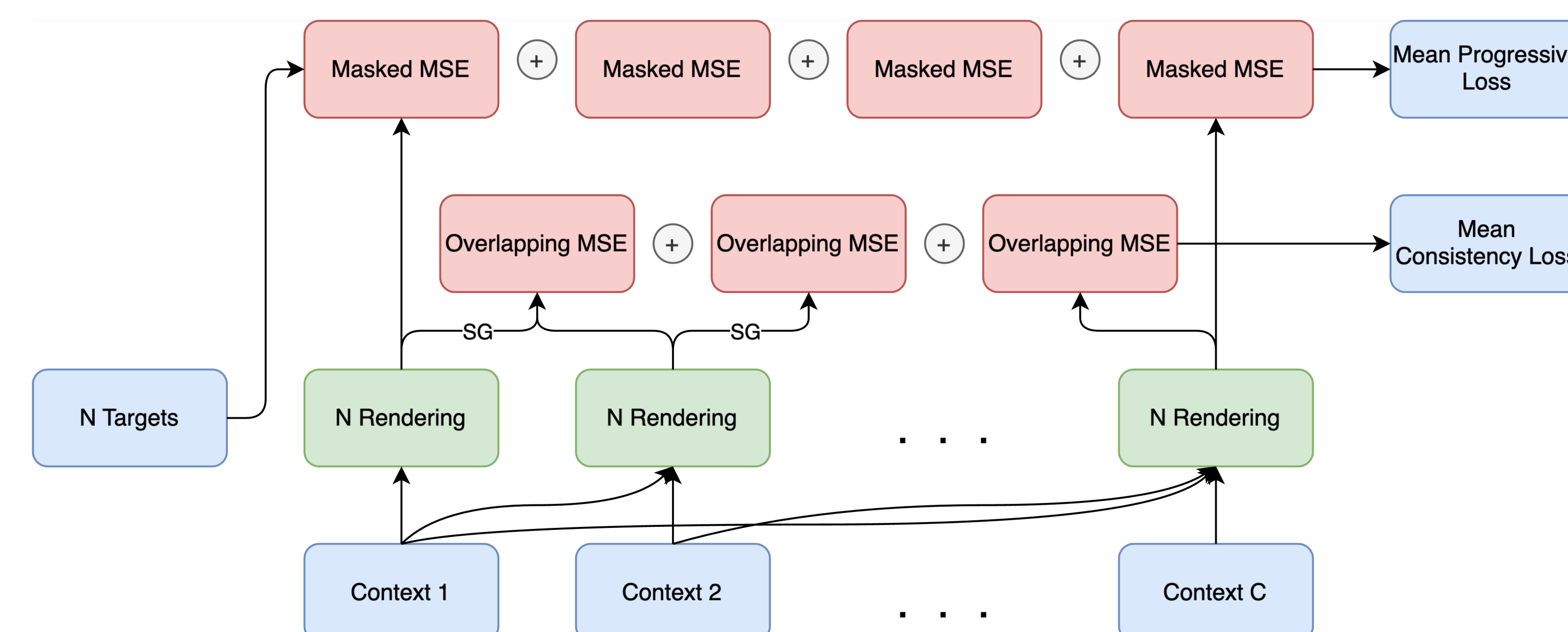
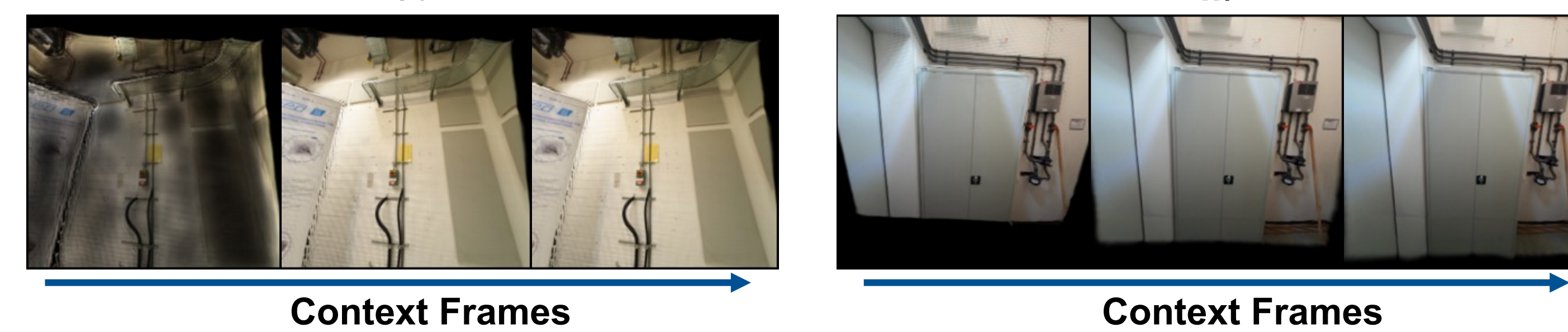
- Applies regularization to prevent **floaters** and promote **sparsity**.

$$\mathcal{L}_{\text{conf}} = \sum_{t=1}^T \sum_i^{H \times W} C_t^i \ell_{\text{regr}}(t, i) + \alpha \log(C_t^i)$$

$$\mathcal{L}_{\text{opacity}} = \frac{1}{N} \sum_{i=1}^N |o_i| \quad \mathcal{L}_{\text{scale}} = \frac{1}{N} \|s_i\|_2^2$$

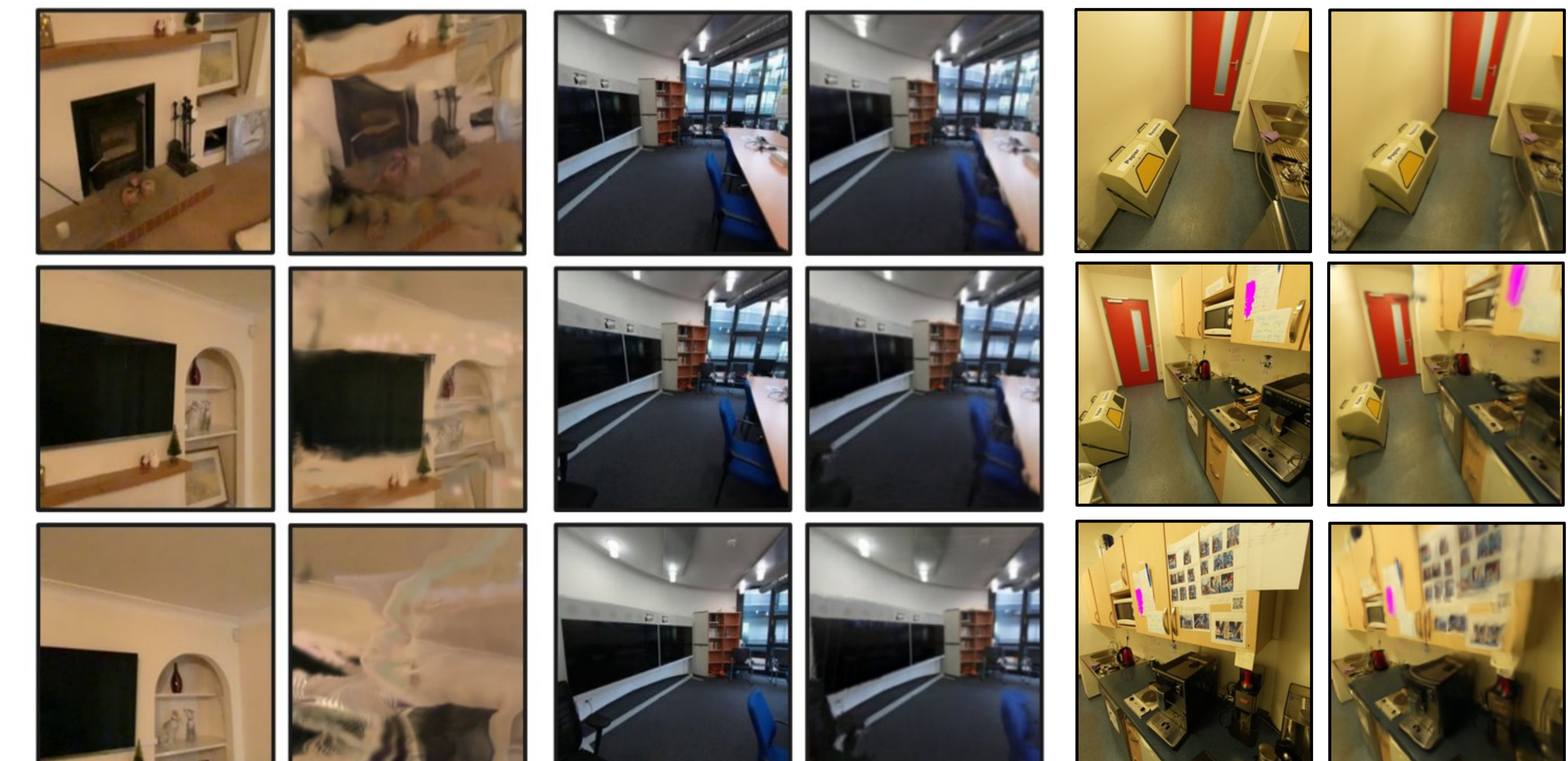
Temporal Consistency & Progressive Loss

- Adds contextual consistency to enforce **causality** (i.e. **completion & refinement**).



Evaluation

Results on Scannet++

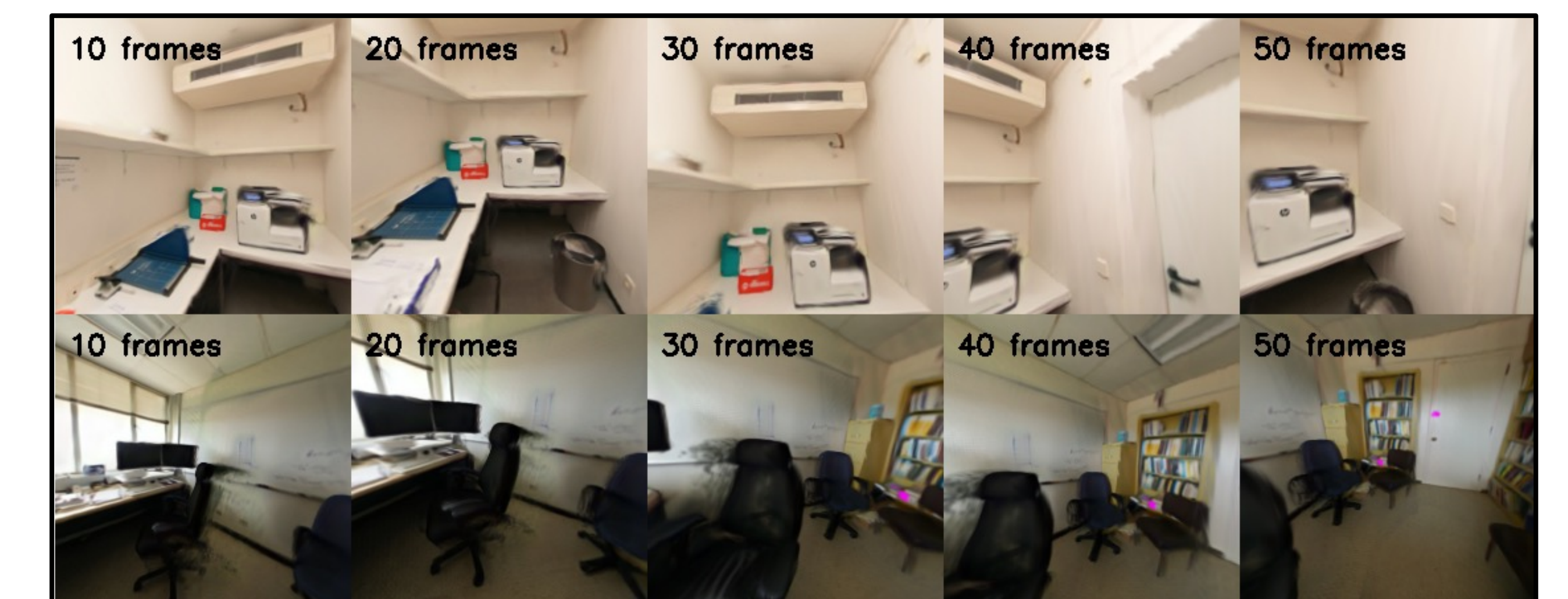


Method	Context: 2, Distance: 5			Context: 10, Distance: 3			Context: 50, Distance: 2		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
MVSpLat	23.67	0.811	0.180	19.31	0.697	0.334	-	-	-
PREF3R	22.83	0.800	0.124	22.60	0.793	0.128	20.38	0.702	0.206
Ours	26.46	0.912	0.126	24.90	0.880	0.151	20.81	0.749	0.286

Ablations

Long-Tail Sequence

- Our method can generate clear reconstructions, even when additional context frames are added.



Memory Footprint

- Our incremental model uses only **linearly bounded** GPU memory, unlike MVSpLat's **square growth**.

