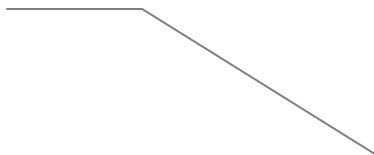


# Dynamic 3D Scene Reconstruction

## Final Presentation

Quang Minh Dinh - 301449800



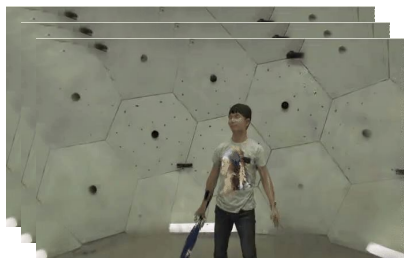
• • • • • • • •  
• • • • • • • •  
• • • • • • • •  
• • • • • • • •

# Input

*Monocular*



*Multi-views*

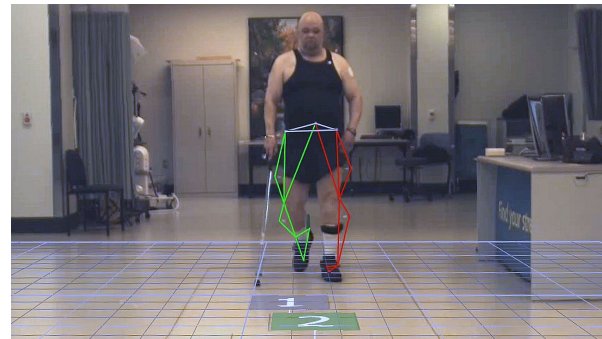


---

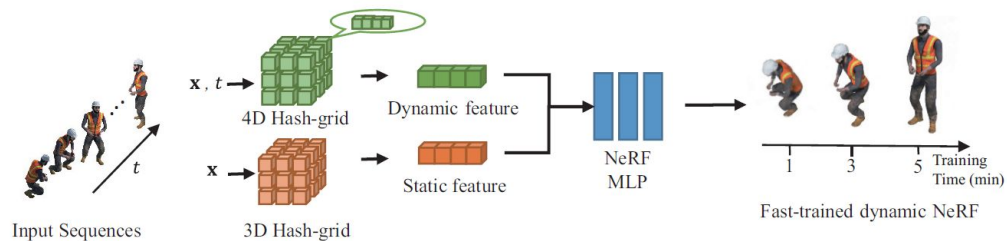
# Output



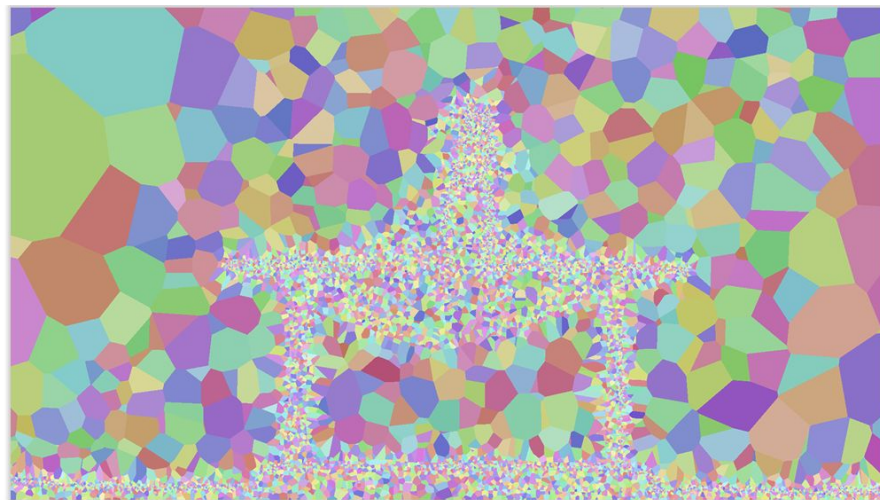
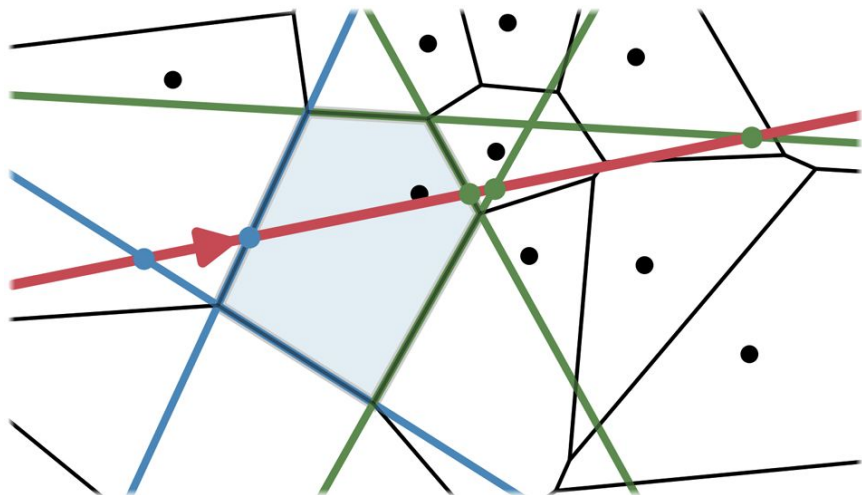
# Motivation



# Related works



# Radiant Foam: Real-Time Differentiable Ray Tracing



# Temporal Modelling

- Density:

$$\sigma_i(t) = \sigma_i^s \exp\left(-\frac{|t/T - \mu_i^\tau|^2}{\exp(2s_i^\tau)}\right)$$

- Motion:

$$\mu_i(t) = \sum_{k=1}^{n_p} b_{i,k} (t/T - \mu_i^\tau)^k$$

- Spherical harmonic coefficients:

$$sh_i(t) = \sum_{j=1}^m \cos\left(\frac{2\pi j}{T}t\right) sh_i^{s,j}$$



# Structure Collapse



# Temporal Perturbations

$$t' \sim \mathcal{N}(t, \Delta t/4), \quad \begin{cases} t' = t + \Delta t/2 & \text{if } t' > t + \Delta t/2 \\ t' = t - \Delta t/2 & \text{if } t' < t - \Delta t/2 \end{cases}$$



# Color Gradient Dependency



*Generated view*

*Ground-truth*

*Error map*





# Structural Similarity Index Measure Loss

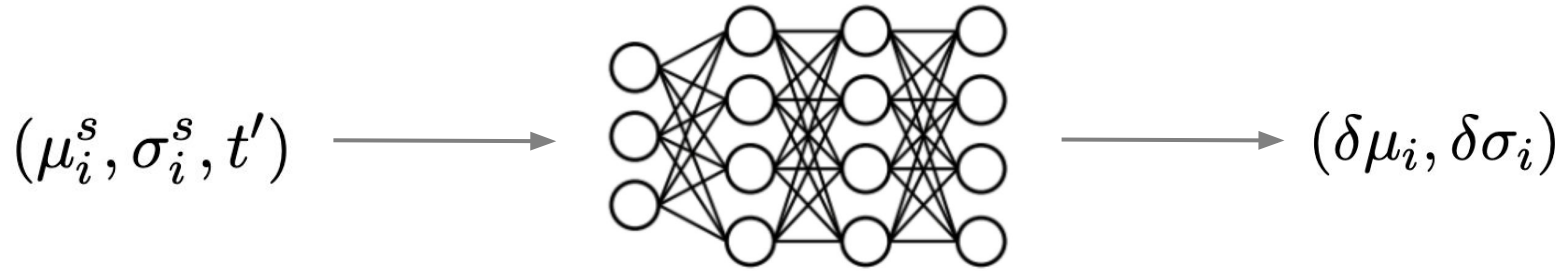
---

$$\text{SSIM}_{\text{loss}}(x, y) = 1 - \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

- $\mu, \sigma$  : pixel sample mean and variance.



# Ablation: Deformation Network





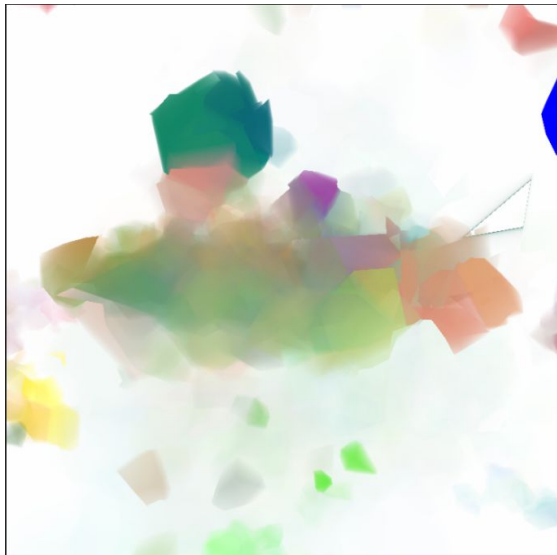
# Timeline

---

- **Week 1 (17/2 - 23/2):** Read Radiant Foam's codebase, replicated results. Implemented dataloaders for D-NeRF & Neural 3D Video.
- **Week 2 & 3 (24/2 - 9/3):** Pre-computed COLMAP sparse reconstructions. Added temporal modelling and ran experiments on monocular videos.
- **Week 4 (10/3 - 16/3):** Modified the training pipeline for multiviews videos.
- **Week 5 (17/3 - 23/3):** Fine-tuned multiviews pipeline, experimented with different temporal functions, losses...
- **Week 6 (24/3 - 30/3):** Implemented deformation network, experimented on the monocular pipeline.
- **Week 7 (31/3 - 6/3):** Conducted final evaluations, wrote report.



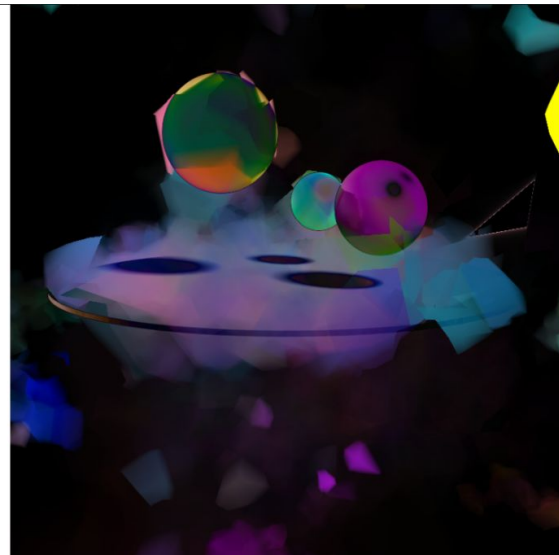
# Monocular Results



*Generated view*



*Ground-truth*



*Error map*



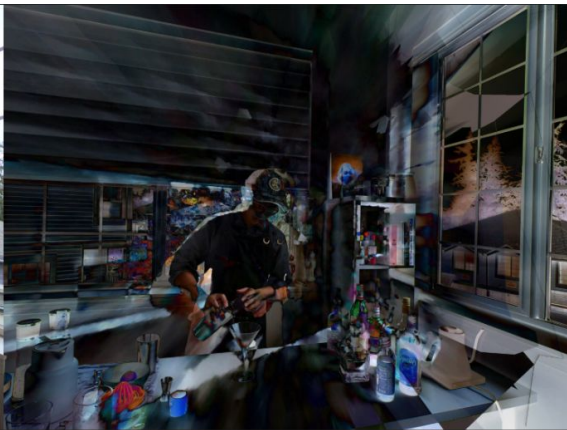
# Monocular Results



*Generated view*



*Ground-truth*



*Error map*



# Multi-views Results



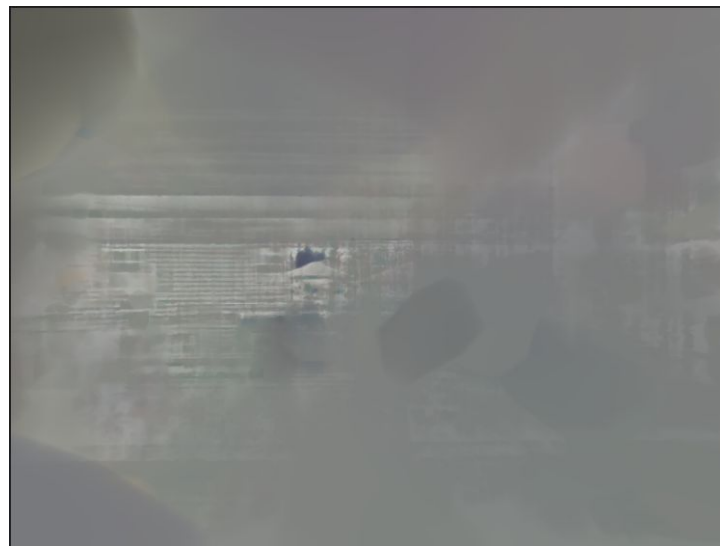
*Generated view*

*Ground-truth*

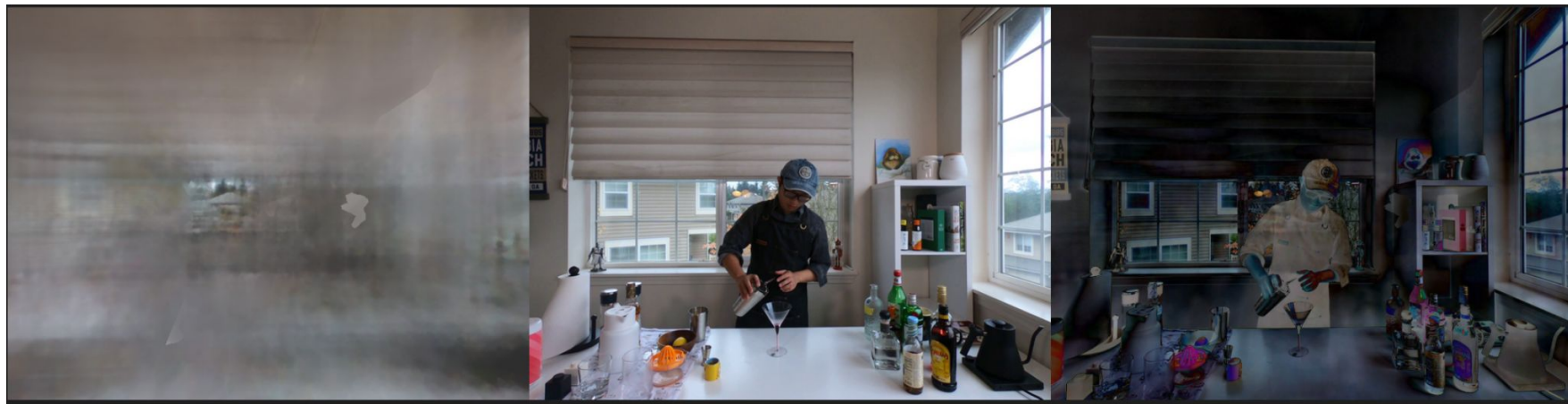
*Error map*



# Effects of Temporal Perturbations



# Effects of SSIM Loss



*Generated view*

*Ground-truth*

*Error map*





# Conclusion

---

- Temporal radial basis functions do not work well on monocular videos (wrong angle).
- Dynamic reconstruction with Radiant Foam is very sensitive to pruning.
- Point density is very sensitive to changes.



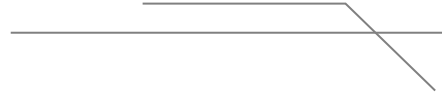


# Next Steps

---

- Implement dynamic AABB tree for faster training.
- Find better ways to represent dynamic adjacency list.
- Pruning and densification strategies for dynamic reconstruction.
- Find better methods to model complex motions.





# References



[1] Govindarajan, S., Rebain, D., Yi, K.M., & Tagliasacchi, A. (2025). Radiant Foam: Real-Time Differentiable Ray Tracing.

[2] Li, Z., Chen, Z., Li, Z., & Xu, Y. (2023). Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8508-8520.

[3] Huang, Y., Sun, Y., Yang, Z., Lyu, X., Cao, Y., & Qi, X. (2023). SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4220-4230.

[4] Schönberger, J.L., Zheng, E., Frahm, J., & Pollefeys, M. (2016). Pixelwise View Selection for Unstructured Multi-View Stereo. European Conference on Computer Vision.

[5] Schönberger, J.L., & Frahm, J. (2016). Structure-from-Motion Revisited. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4104-4113.

