Grid4D: 4D Decomposed Hash Encoding for High-Fidelity Dynamic Gaussian Splatting

Jiawei Xu [1], Zexin Fan [1], Jin Xie [2 3] and Jian Yang [1]

[1] PCA Lab, VCIP, College of Computer Science, Nankai University

[2] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

[3] School of Intelligence Science and Technology, Nanjing University, Suzhou, China

Problem Definition

Dynamic Scenes Rendering [1]

- Input: **Part view, Part time** images
- Output: **Any view, Any time** rendering

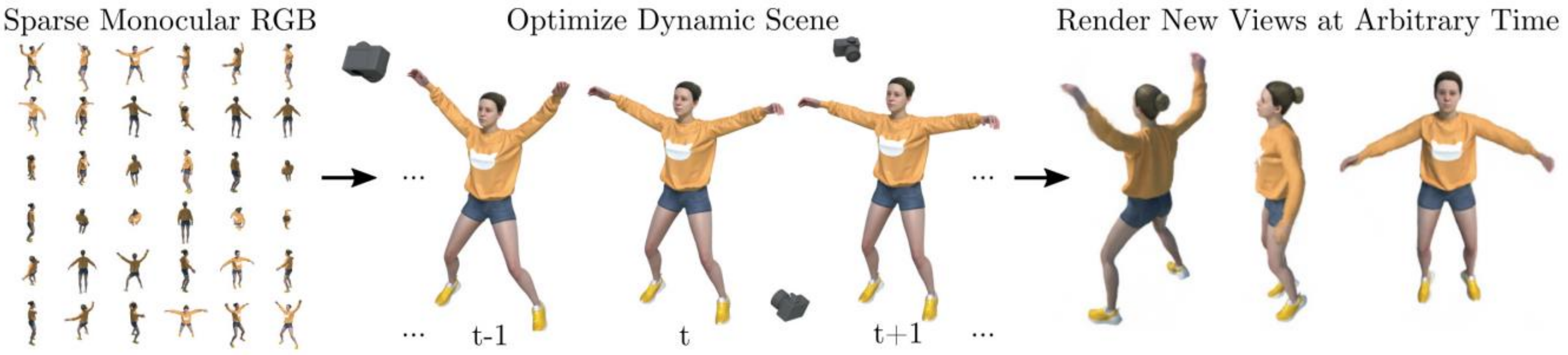
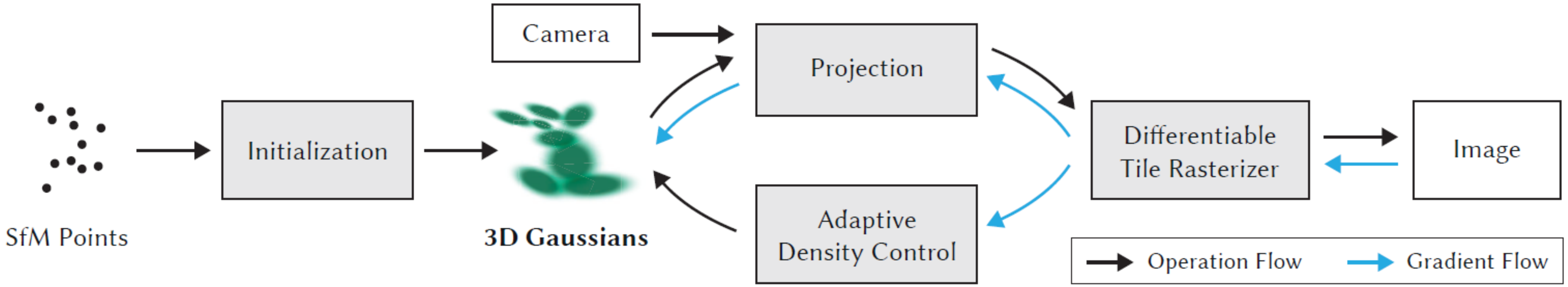


Figure 2: **Problem Definition.** Given a sparse set of images of a dynamic scene moving non-rigidly and being captured by a monocular camera, we aim to design a deep learning model to implicitly encode the scene and synthesize novel views at an arbitrary time. Here, we visualize a subset of the input training frames paired with accompanying camera parameters, and we show three novel views at three different time instances rendered by the proposed method.

Preliminary

Gaussian Splatting [2]

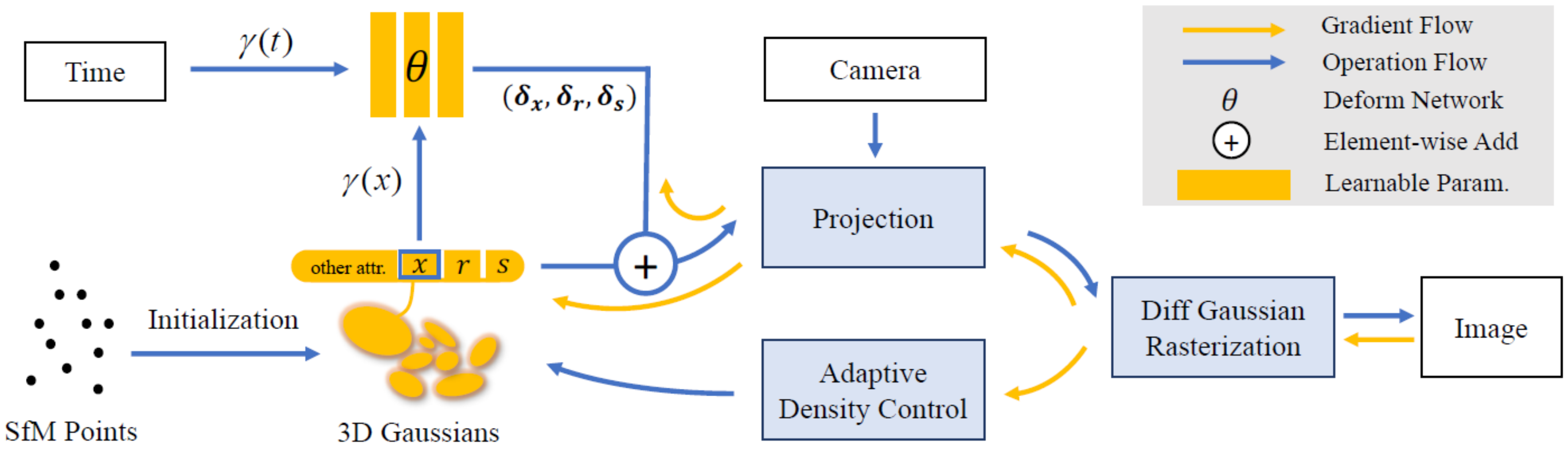
- Modeling static scenes with 3D Gaussians.
- Splitting, cloning and pruning Gaussians during optimization.



Related Work

Deformable 3D Gaussian (DeformGS) [3]

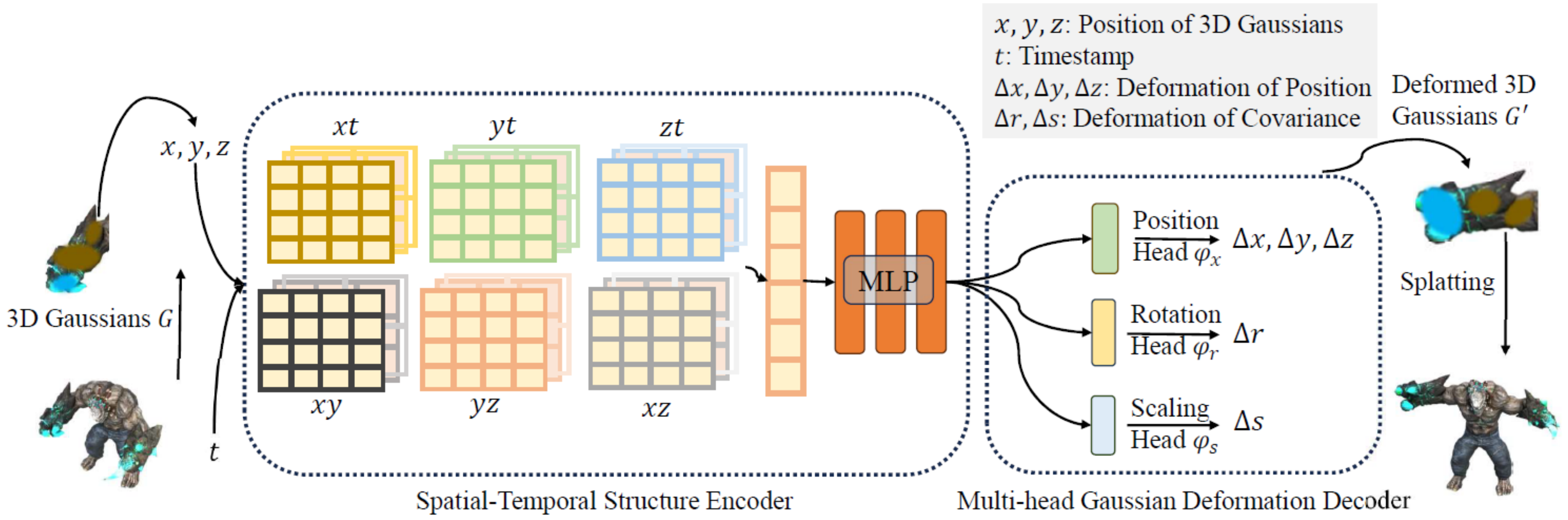
- ➢ Deforming Gaussians with fully MLP-based implicit neural network.
- ➢ Problem: over-smooth inherent property.



Related Work

4D-GS [4]

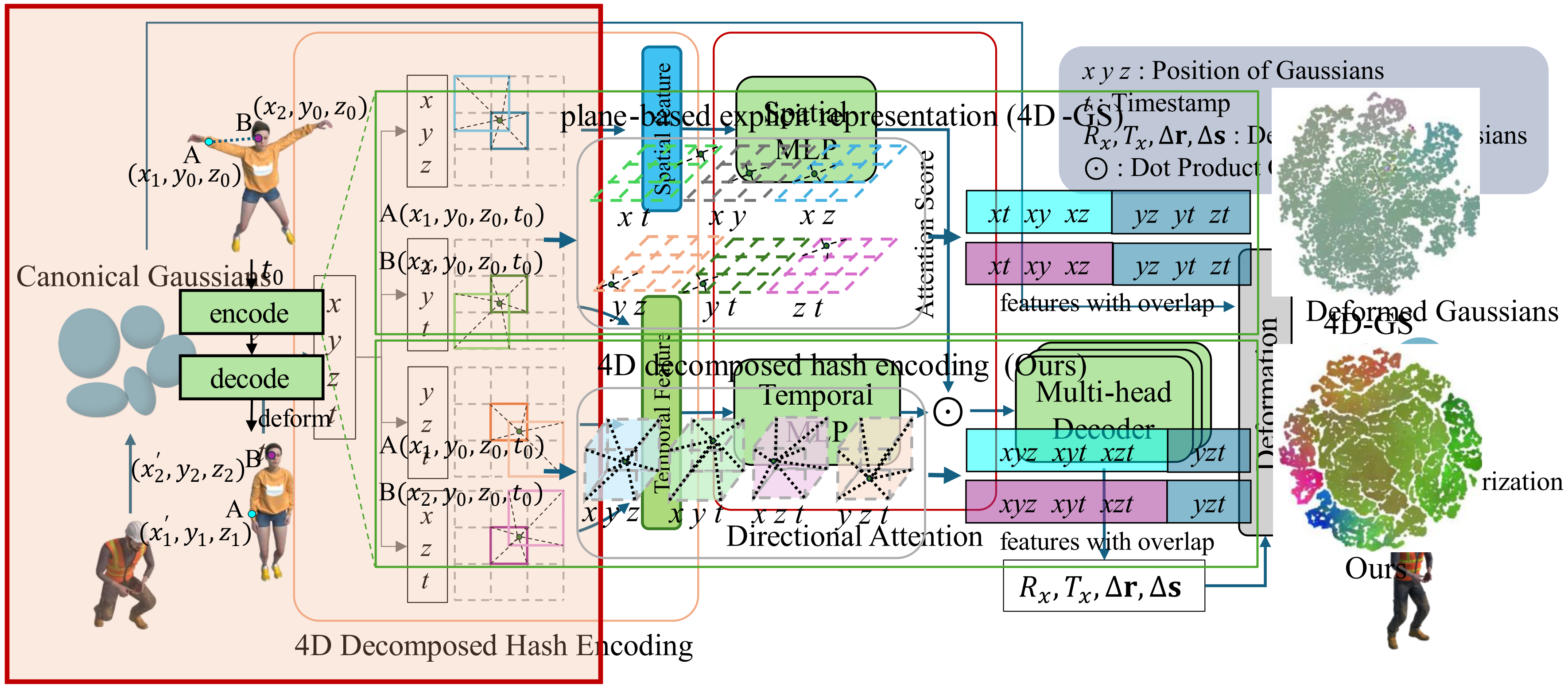
- Deforming Gaussians with plane-based explicit neural network.
- Problem: unsuitable low-rank assumption.



Methods

4D Decomposed Hash Encoding

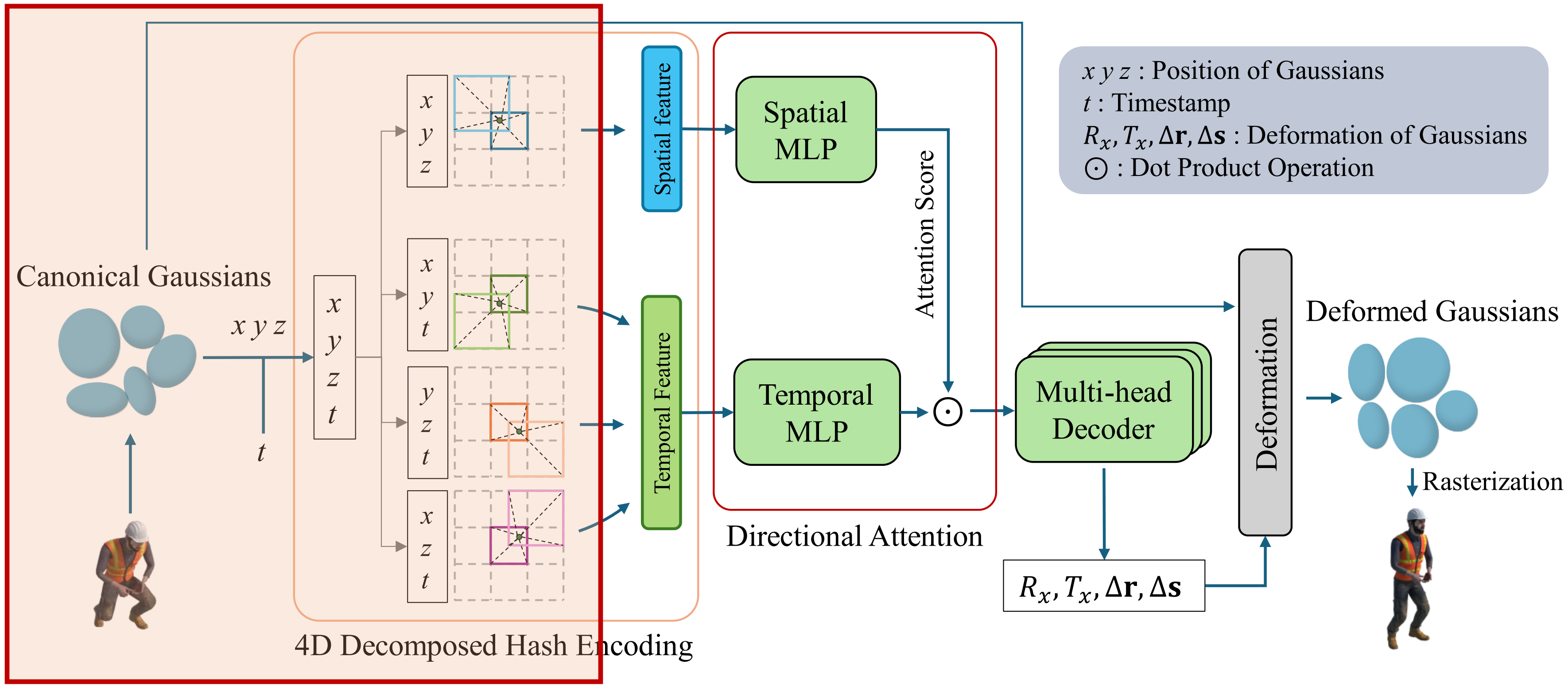
➢ Tri-axial 4D Decomposition.



Methods

Directional Attention

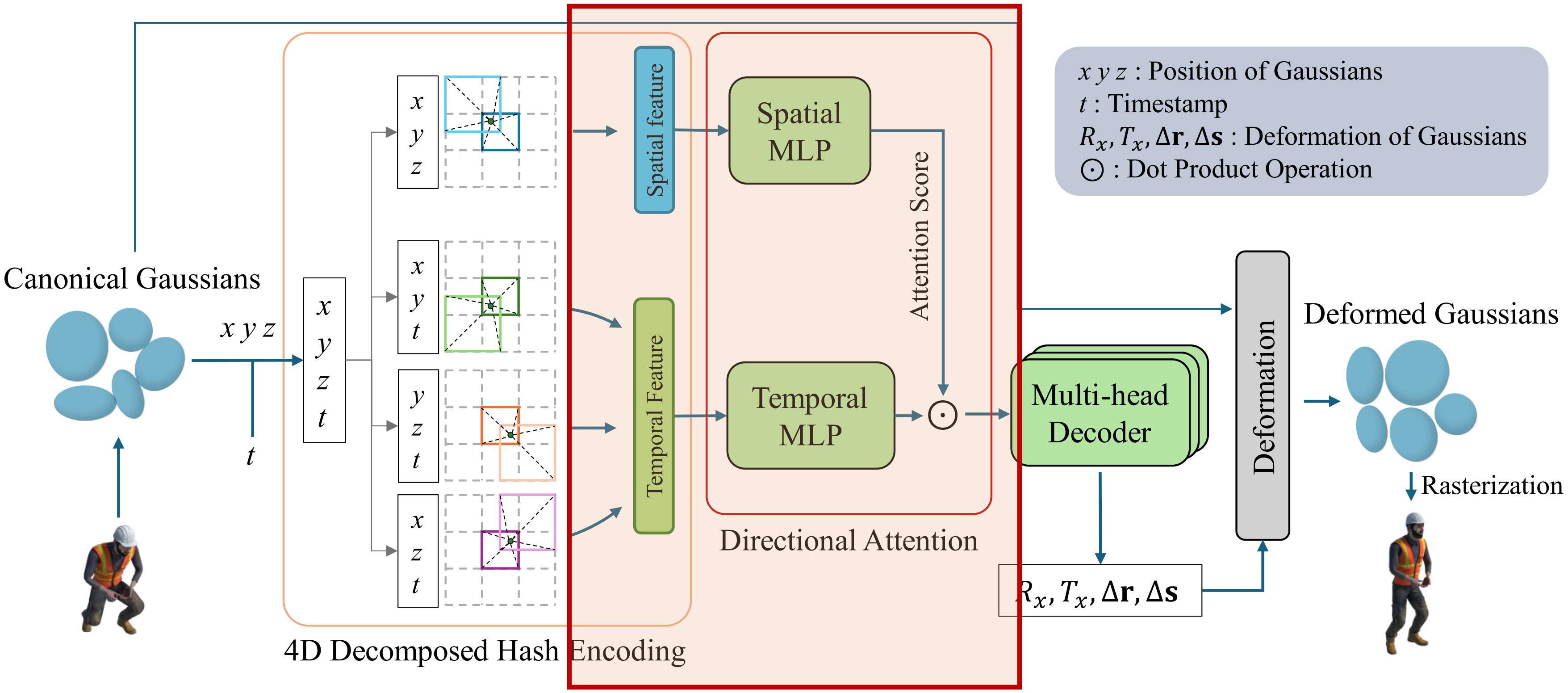
➢ Spatial features choose and direct temporal features.



Methods

Multi-head Decoder

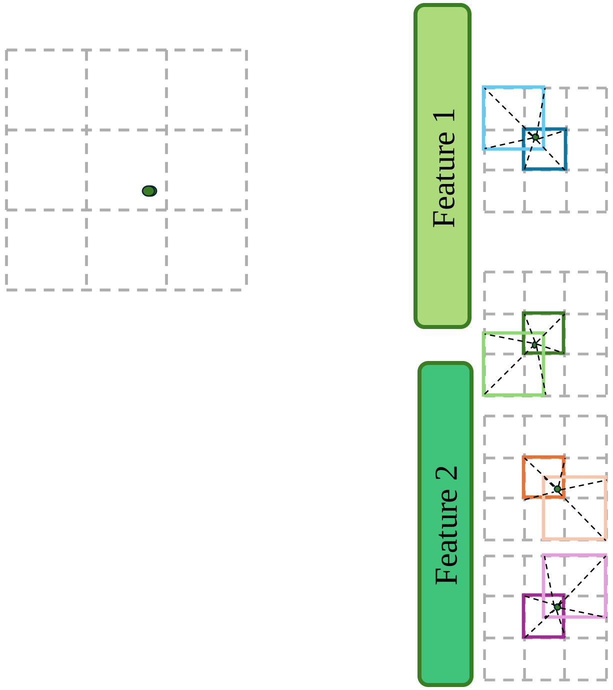
➢ Decode features for deformation



Methods

Smooth Regularization

$$L_r = \left\| G_{xyzt}(x, y, z, t) - G_{xyzt}\left(x + \epsilon_x, y + \epsilon_y, z + \epsilon_z, t + \epsilon_t\right) \right\|^2$$



Get similar

Experiments

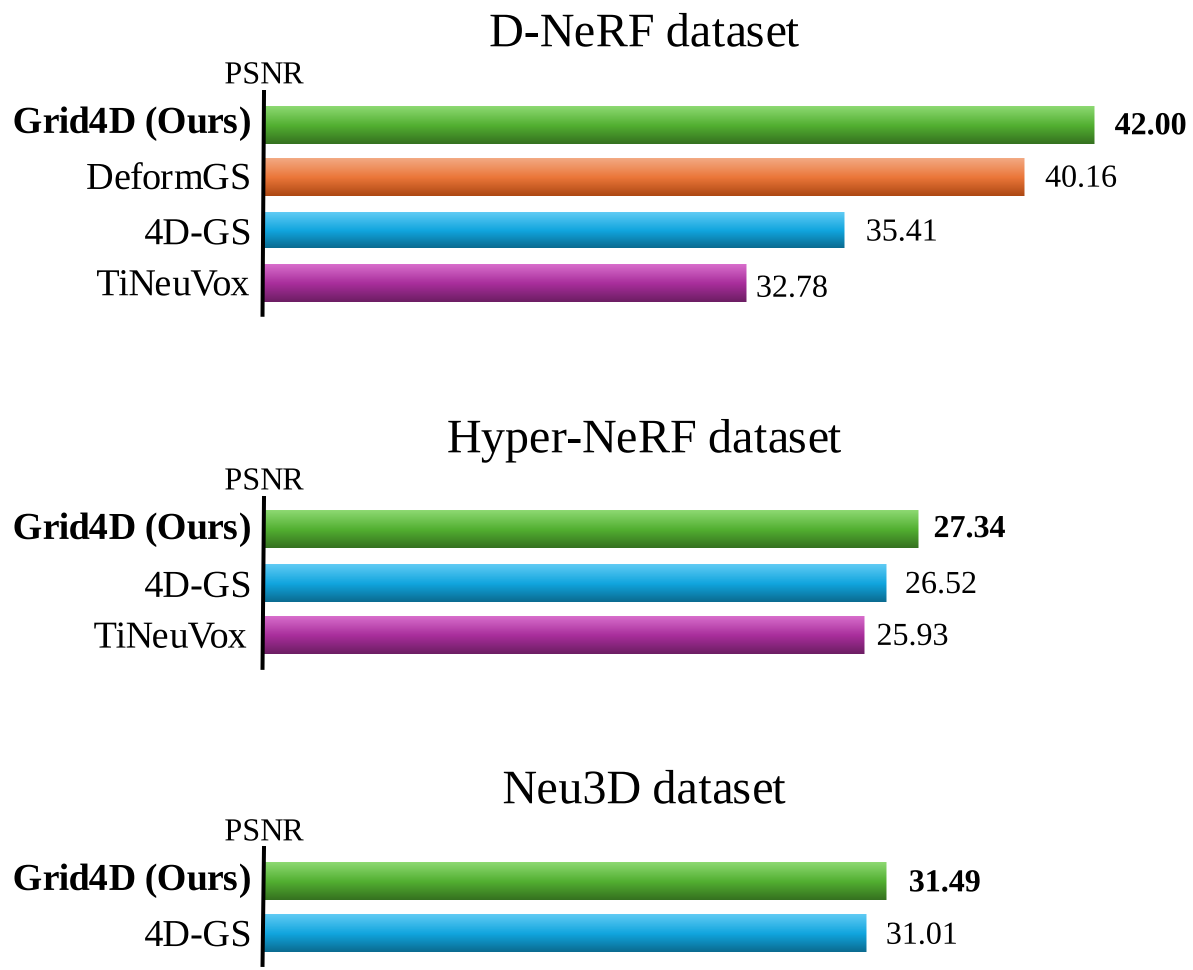
Dataset

- D-NeRF [1] synthetic dataset
- HyperNeRF [5] real-world dataset
- Neu3D [6] real-world dataset

Baseline

- TiNeuVox [7]
- 4D-GS [4]
- DeformGS (Deformable 3D Gaussian) [3]

Comparison


D-NeRF dataset
PSNR
Grid4D (Ours)
42.00
DeformGS
40.16
4D-GS
35.41
TiNeuVox
32.78
Hyper-NeRF dataset
PSNR
Grid4D (Ours)
27.34
4D-GS
26.52
TiNeuVox
25.93
Neu3D dataset
PSNR
Grid4D (Ours)
31.49
4D-GS
31.01

Comparison

- Comparisons on the synthetic D-NeRF dataset.

4D-GS Deform-GS **Grid4D(Ours)**

Comparison

- Comparisons on the real-world HyperNeRF dataset.

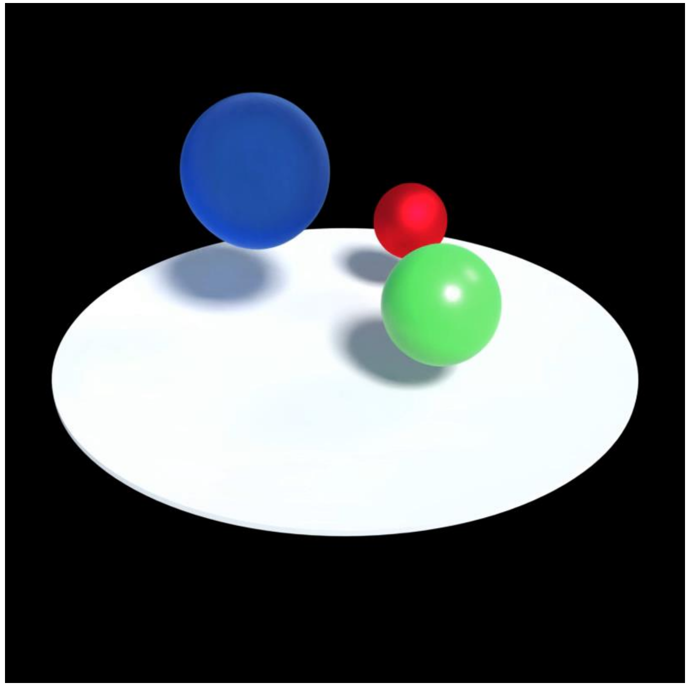
4D-GS

Deform-GS

Grid4D(Ours)

Visualization

- Qualitative results on the synthetic D-NeRF dataset.



Bouncing Balls

Mutant

Standup

Visualization

- Qualitative results on the real-world Neu3D dataset.

More details and experiments

Project page: https://jiaweixu8.github.io/Grid4D-web/

References

[1] Pumarola et al. D-NeRF: Neural radiance fields for dynamic scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[2] Kerbl et al. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4), 2023.

[3] Yang et al. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction. arXiv preprint. arXiv:2309.13101, 2023.

[4] Wu et al. 4D Gaussian splatting for real-time dynamic scene rendering. arXiv preprint arXiv:2310.08528, 2023.

[5] Park et al. HyperNeRF: A higher-dimensional representation for topologically varying neural radiance fields. arXiv preprint arXiv:2106.13228, 2021.

[6] Li et al. Neural 3D video synthesis from multi-view video. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5521–5531, 2022.

[7] Fang et al. Fast dynamic radiance fields with time-aware neural voxels. In SIGGRAPH Asia 2022 Conference Papers, pages 1–9, 2022.