

SpeeDe3DGS: Speedy Deformable 3D Gaussian Splatting with Temporal Pruning and Motion Grouping

Allen Tu*

Haiyang Ying*

Alex Hanson

Yonghan Lee

Tom Goldstein

Matthias Zwicker

University of Maryland, College Park



Figure 1. Our SpeeDe3DGS framework achieves $9.88\times$ faster rendering, $11.37\times$ fewer Gaussians, and $2.87\times$ shorter training on the HyperNeRF [37] chicken scene while preserving the image quality of DeformableGS [53] through Temporal Sensitivity Pruning (TSP) and Sampling (TSS). Applying our GroupFlow method on top of pruning accelerates rendering and training by $33.13\times$ and $4.24\times$, respectively.

Abstract

Dynamic extensions of 3D Gaussian Splatting (3DGS) achieve high-quality reconstructions through neural motion fields, but per-Gaussian neural inference makes these models computationally expensive. Building on DeformableGS, we introduce Speedy Deformable 3D Gaussian Splatting (SpeeDe3DGS), which bridges this efficiency-fidelity gap through three complementary modules: Temporal Sensitivity Pruning (TSP) removes low-impact Gaussians via temporally aggregated sensitivity analysis, Temporal Sensitivity Sampling (TSS) perturbs timestamps to suppress floaters and improve temporal coherence, and GroupFlow distills the learned deformation field into shared $SE(3)$ transformations for efficient groupwise motion. On the 50 dynamic scenes in MonoDyGauBench, integrating TSP and TSS into DeformableGS accelerates rendering by $6.78\times$ on average while maintaining neural-field fidelity and using $10\times$ fewer primitives. Adding GroupFlow culminates in $13.71\times$ faster rendering and $2.53\times$ shorter training, surpassing all baselines in speed while preserving superior image quality. Project page: <https://speede3dgs.github.io>

* denotes equal contribution.

1. Introduction

Neural Radiance Fields (NeRFs) [33] have transformed novel-view synthesis by representing static scenes as continuous volumetric functions optimized through differentiable rendering. While capable of photorealistic reconstruction, NeRFs rely on dense per-ray MLP inference, resulting in slow training and rendering. 3D Gaussian Splatting (3DGS) [20] replaces volumetric integration with differentiable rasterization of point-based Gaussian primitives, achieving real-time rendering with high fidelity and rapid convergence. This efficiency has made 3DGS a foundation for modern approaches to 3D reconstruction.

Dynamic Gaussian Splatting extends this framework to time-varying scenes by coupling each Gaussian with a learned motion field. Methods such as DeformableGS [53] and 4DGS [49] model temporal variation using neural deformation fields or spatiotemporal grids, enabling detailed dynamic reconstructions. However, these approaches require per-Gaussian neural inference at every frame, leading to substantial computational cost. Monocular Dynamic Gaussian Splatting Benchmark (MonoDyGauBench) [25] presents the first systematic comparison across dynamic 3DGS variants, revealing that neural motion fields achieve the highest and

most stable reconstruction quality but render several times slower than analytic or non-neural representations.

This efficiency–fidelity gap motivates methods that preserve the expressiveness of neural motion fields while approaching the speed of analytic models. In this paper, we propose **Speedy Deformable 3D Gaussian Splatting (SpeeDe3DGS)**, a framework that bridges this gap by reducing redundant neural inference through three complementary modules. **Temporal Sensitivity Pruning (TSP)** eliminates redundant or low-impact Gaussians to reduce deformation and rendering cost, while its companion, **Temporal Sensitivity Sampling (TSS)**, stabilizes pruning by probing temporally jittered timestamps to suppress floaters and improve temporal coherence. Finally, **GroupFlow** distills the learned neural deformation field into shared rigid SE(3) transformations, grouping Gaussians with similar motion so that each cluster shares a single motion representation. Together, these modules reduce redundant primitives, stabilize temporal pruning, and dramatically accelerate dynamic reconstruction without sacrificing neural-field fidelity.

On MonoDyGauBench, integrating TSP and TSS into **DeformableGS** accelerates rendering by $6.78\times$ and training by $2.18\times$, maintaining neural-field fidelity while achieving frame rates comparable to non-neural baselines and using $10\times$ fewer primitives. Adding **GroupFlow** further increases rendering and training performance to $13.71\times$ and $2.53\times$, respectively, surpassing all baselines in speed while still outperforming every non-neural method in image quality.

In summary, we propose the following contributions:

1. **Temporal Sensitivity Pruning (TSP)**: Temporally aggregated gradient-based pruning that reduces neural inference cost by removing low-impact Gaussians.
2. **Temporal Sensitivity Sampling (TSS)**: Temporal perturbation during pruning to improve stability and suppress floaters in noisy or dynamic scenes.
3. **GroupFlow**: Grouped SE(3) motion distillation that replaces per-Gaussian deformation with shared rigid motion for efficient, temporally coherent rendering.

2. Related Work

Neural Radiance Fields (NeRF) [33] represent scenes as continuous volumetric functions optimized via differentiable volume rendering. Dynamic extensions incorporate time either through explicit temporal modeling [4, 9, 44] or by learning deformation fields that map a canonical space to time-varying observations [36, 37, 39]. Grid-based and factorized encodings [3, 7, 8, 11, 13, 29, 32, 38, 45] improve training and inference efficiency, but rendering remains slow due to per-ray MLP evaluation.

3D Gaussian Splatting (3DGS) [20] replaces volumetric integration with differentiable rasterization of Gaussian primitives, enabling real-time rendering for static scenes. Dynamic extensions incorporate explicit or implicit motion

models to represent time-varying geometry. Dy3DGS [31] fits 6-DoF trajectories to individual Gaussians, while DeformableGS [53] and 4DGS [49] learn neural deformation fields using MLPs or HexPlane grids [3]. RTGS [54] encodes temporal variation implicitly within a 4D Gaussian distribution. MonoDyGauBench [25] unifies these formulations and identifies five representative motion parameterizations – EffGS [19], STG [24], DeformableGS [53], 4DGS [49], and RTGS [54] – each trading off reconstruction fidelity and efficiency through distinct models of motion.

2.1. Gaussian Pruning

Redundancy in Gaussian-based reconstructions has motivated extensive pruning research to improve efficiency and reduce memory overhead [5, 6, 14, 15, 34]. Most approaches either learn per-Gaussian pruning masks [22, 56] or compute heuristic importance scores to remove low-contribution Gaussians [1, 2, 5, 6, 12, 26, 30, 34, 35]. Recent extensions adapt these strategies to dynamic scenes by incorporating temporal variation into pruning metrics [17, 28, 55]. PUP 3D-GS [15] introduced a Hessian-based sensitivity score to quantify each Gaussian’s contribution to reconstruction, while Speedy-Splat [14] refined this for pruning during training, achieving over $6\times$ faster rendering with minimal fidelity loss. Building on these sensitivity-based principles, our approach generalizes gradient-based importance analysis to dynamic 3D Gaussian Splatting, enabling pruning that accounts for motion-dependent redundancy across time.

2.2. Motion Analysis and Group-Based Modeling

Another strategy for compressing dynamic scenes is to reduce temporal complexity by decomposing motion into static and dynamic regions. Since most real-world scenes are largely static, several methods identify dynamic regions via motion cues [40, 47, 50, 51, 57] or cluster Gaussians using self-supervised or segmentation-based methods [47, 51], restricting expensive 4D modeling to dynamic areas.

However, static–dynamic decomposition alone is insufficient for complex non-rigid motion. Group-based motion representations address this by jointly modeling correlated motion across sets of Gaussians. SC-GS [16], SPGaussian [43], and MoSca [23] model motion using control points with linear blend skinning or dual quaternion blending, often combined with MLP deformation. Methods such as DynMF [21], Gaussian-Flow [27], and Shape of Motion [46] instead learn low-dimensional motion bases or flow fields. While effective for improving visual quality, these methods require costly neural inference [16, 43], complex deformation computation [23], or guidance from 2D priors [23, 46]. In contrast, our approach distills neural motion into efficient grouped SE(3) transformations, clustering Gaussians by trajectory similarity to share rigid motion and reduce per-Gaussian inference while maintaining high fidelity.













Baseline	Pruning (TSP + TSS)	GroupFlow	Pruning + GroupFlow
 112K Gaussians 51.86 FPS	 8.9K Gaussians 387.09 FPS	 115K Gaussians 418.65 FPS	 9.0K Gaussians 522.29 FPS
 145K Gaussians 48.45 FPS	 12.5K Gaussians 374.21 FPS	 145K Gaussians 387.38 FPS	 12.4K Gaussians 510.42 FPS
 97.6K Gaussians 65.87 FPS	 7.0K Gaussians 455.66 FPS	 95.6K Gaussians 321.76 FPS	 7.0K Gaussians 536.57 FPS

Figure 2. **Visual comparison of the baseline DeformableGS [53] and our Speede3DGS methods.** Pruning (TSP + TSS) and GroupFlow deliver vastly faster results. Top: *as* from NeRF-DS [52]. Middle: *basin* from NeRF-DS. Bottom: *trex* from D-NeRF [39].

3. Background

3.1. 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [20] represents scenes as parametric, point-based models composed of 3D Gaussians. Given a set of ground truth training images $\mathcal{I}_{gt} = \{I_i \in \mathbb{R}^{H \times W}\}_{i=1}^M$, the scene is initialized using Structure from Motion (SfM) to produce image pose estimates and a sparse point cloud that serves as the initial means for the 3D Gaussians. Each image is paired with its corresponding pose in $\mathcal{P}_{gt} = \{\phi_i \in \mathbb{R}^{3 \times 4}\}_{i=1}^M$ and used to optimize the scene.

Each 3D Gaussian primitive \mathcal{G}_i is parameterized by three geometry parameters – mean $\mu_i \in \mathbb{R}^3$, scale $s_i \in \mathbb{R}^3$, and rotation $r_i \in \mathbb{R}^4$ – and two appearance parameters – view-dependent spherical harmonic color $h_i \in \mathbb{R}^{16 \times 3}$ and opacity $\sigma_i \in \mathbb{R}$. The set of all parameters can be described as

$$\mathcal{G} = \{\mathcal{G}_i = \{\mu_i, s_i, r_i, h_i, \sigma_i\}\}_{i=1}^N, \quad (1)$$

where N is the number of Gaussians. Given camera pose ϕ , the scene is rendered by projecting all Gaussians to image space and compositing them via alpha blending. The value of the 2D projection of Gaussian \mathcal{G}_i at pixel p is given by:

$$g_i = e^q, \quad q = -\frac{1}{2}(p - \mu_{i2D})\Sigma_{i2D}^{-1}(p - \mu_{i2D})^T, \quad (2)$$

where μ_{i2D} is the projection of μ_i onto image space and Σ_{i2D}^{-1} is the inverse of the 2D covariance computed via the EWA Splatting approximation [58] of the perspective projected 3D Gaussian. The model is optimized via stochastic gradient descent on an image reconstruction loss

$$L(\mathcal{G}|\phi, I_{gt}) = \|I_{\mathcal{G}}(\phi) - I_{gt}\|_1 + L_{D-SSIM}(I_{\mathcal{G}}(\phi), I_{gt}), \quad (3)$$

where $I_{\mathcal{G}}(\phi)$ is the rendered image for pose ϕ . During training, the scene is periodically densified by cloning and splitting uncertain Gaussians and pruned by removing large or transparent ones. While this formulation effectively models static geometry, it cannot capture the non-rigid or time-varying motion present in dynamic scenes.

3.2. Dynamic Gaussian Splatting

Dynamic Gaussian Splatting extends 3DGS by coupling canonical Gaussians \mathcal{G} with a deformation function \mathcal{D} that predicts their evolution over time:

$$(\mu + \Delta\mu_t, r + \Delta r_t, s + \Delta s_t) = \mathcal{D}(\mu, r, s, t), \quad (4)$$

where μ , r , and s denote the canonical mean, rotation, and scale, and $(\Delta\mu_t, \Delta r_t, \Delta s_t)$ are time-dependent offsets. At each timestep t , the deformed set \mathcal{G}_t is rendered by the same differentiable rasterizer as static 3DGS, with each training image paired with its corresponding pose ϕ and timestamp t .

Recent dynamic Gaussian Splatting methods differ in how they parameterize \mathcal{D} , balancing motion expressiveness against computational efficiency. MonoDyGauBench [25] categorizes representative formulations into five major types:

1. **EffGS** [19] models per-Gaussian motion using second-order Fourier and first-order polynomial bases.
2. **SpaceTimeGaussians (STG)** [24] combines radial basis and polynomial bases to represent locally supported motion trajectories.
3. **DeformableGS** [53] learns a shared MLP deformation field conditioned on spatial position and time.
4. **4D Gaussian Splatting (4DGS)** [49] replaces the MLP with a factorized HexPlane grid [3] for more efficient spatiotemporal interpolation.
5. **RTGS** [54] is the only approach without an explicit motion model, encoding temporal variation implicitly within a 4D Gaussian distribution.

Among these, **DeformableGS** and **4DGS** represent the canonical *neural motion field* formulations. They consistently achieve the highest reconstruction quality among existing methods but at substantial computational cost, as their continuous deformation fields require neural inference for every Gaussian at every frame. This tradeoff between fidelity and efficiency motivates techniques that preserve high visual quality while reducing per-Gaussian inference overhead.

4. Method

Dynamic Gaussian Splatting methods that rely on *neural motion fields* effectively capture complex non-rigid motion, but their high computational cost limits real-time performance. To overcome this bottleneck, we introduce two complementary strategies that significantly improve efficiency while maintaining high visual fidelity: **Temporal Sensitivity Pruning (TSP)** and **GroupFlow**. TSP eliminates redundant or unstable Gaussians through temporally aggregated sensitivity analysis, while its submodule, **Temporal Sensitivity Sampling (TSS)**, introduces temporally jittered sampling to stabilize pruning and suppress floating artifacts. GroupFlow further distills the learned neural motion field into grouped SE(3) transformations that share rigid motion within coherent regions. We integrate these components into **Speedy Deformable 3D Gaussian Splatting (SpeeDe3DGS)**, a unified pipeline that reduces redundant primitives, stabilizes temporal pruning, and lowers deformation inference cost – achieving neural reconstruction quality at rendering speeds comparable to non-neural motion representations.

4.1. Temporal Sensitivity Pruning

Recent studies have shown that 3D Gaussian Splatting models are heavily over-parameterized – comparable visual fidelity can often be achieved with far fewer primitives [15]. Because the computational cost of deformable 3DGS scales

linearly with the number of Gaussians, reducing redundant primitives is a direct path to accelerating both training and rendering. Building on the gradient-based pruning method of Speedy-Splat [14], which achieves state-of-the-art efficiency in static 3DGS, we extend this concept to dynamic reconstruction. Our formulation introduces a **temporal sensitivity pruning score** that quantifies each Gaussian’s cumulative influence on reconstruction across both space and time.

For Gaussian \mathcal{G}_i , we compute the second-order sensitivity of the L_2 reconstruction loss to perturbations of its projected contribution g_i across all training views and timesteps:

$$\nabla_{g_i}^2 L_2 = \sum_{\phi, t \in \mathcal{P}_{gt}} \left((\nabla_{g_i} I_{\mathcal{G}_t}(\phi))^2 + (I_{\mathcal{G}_t}(\phi) - I_{g_t}) \nabla_{g_i}^2 I_{\mathcal{G}_t}(\phi) \right). \quad (5)$$

As training converges and the residual term diminishes, this simplifies to our temporal sensitivity pruning score:

$$\tilde{U}_{\mathcal{G}_i} \approx \nabla_{g_i}^2 L_2 \approx \sum_{\phi, t \in \mathcal{P}_{gt}} (\nabla_{g_i} I_{\mathcal{G}_t}(\phi))^2. \quad (6)$$

Here, \mathcal{P}_{gt} is the set of training poses and timestamps (ϕ, t) , and $I_{\mathcal{G}_t}(\phi)$ is the rendered image for the deformed Gaussians \mathcal{G}_t at time t and pose ϕ . g_i is the 2D projection of Gaussian \mathcal{G}_i as defined in Equation 2, and $\nabla_{g_i} I_{\mathcal{G}_t}(\phi)$ is the image-space gradient readily available from the renderer’s backward pass. Since deformation parameters $(\Delta\mu_t, \Delta r_t, \Delta s_t)$ vary across timesteps, these gradients inherently encode temporal motion coupling, making $\tilde{U}_{\mathcal{G}_i}$ sensitive not only to static appearance but also to dynamic contribution.

Our **Temporal Sensitivity Pruning (TSP)** approach periodically removes Gaussians with low temporal sensitivity pruning scores, extending sensitivity-based pruning to dynamic 3DGS. By aggregating second-order sensitivities over both spatial and temporal domains, TSP identifies motion-dependent redundancies and eliminates them in a principled, temporally aware manner. This reduces training and inference cost while preserving visual fidelity by concentrating computation on the most informative Gaussians.

4.1.1. Temporal Sensitivity Sampling

While TSP effectively identifies redundant primitives, it evaluates gradients only at the observed training timesteps. This limitation mirrors prior pruning methods for *static* 3DGS, which compute sensitivities from fixed viewpoints and single frames. In dynamic scenes, however, a Gaussian’s contribution can vary over time, and gradient estimates confined to training frames may overlook temporally unstable primitives. These unstable Gaussians – often manifesting as *floaters* – exhibit weak gradients on observed views but introduce visible artifacts at unseen timesteps.

Previous gradient-based analyses [15, 18, 48] have shown that gradient magnitudes can implicitly capture Gaussian sensitivity or stability. We extend this concept to dynamic 3DGS

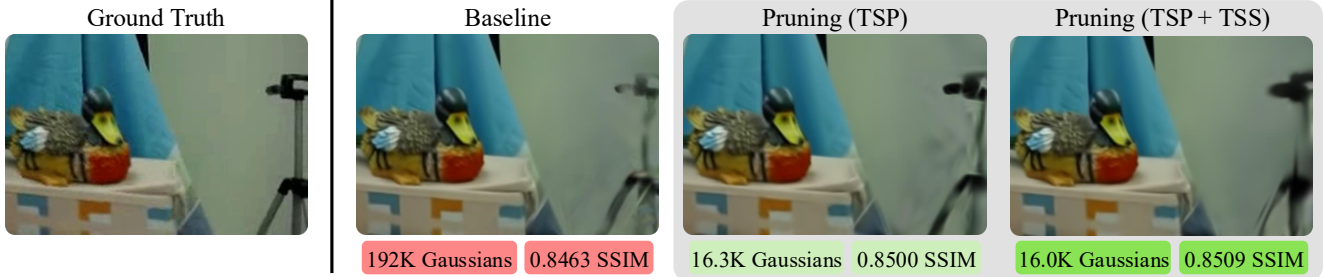


Figure 3. **Comparison of our pruning methods on the real-world NeRF-DS [52] bell scene.** Our proposed Temporal Sensitivity Pruning (TSP) and Temporal Sensitivity Sampling (TSS) methods achieve higher SSIM than the baseline DeformableGS [53] model while using $11\times$ fewer Gaussians. The left regions of the renderings appear visually identical, while the right regions show that combining TSP with TSS significantly reduces temporal flicker and floating artifacts compared to both standard pruning and the unpruned baseline.

by interpreting low sensitivities as indicators of temporal inconsistency – Gaussians that appear stable under observed motion but drift when extrapolated to unseen deformation. To expose such instability, we introduce **Temporal Sensitivity Sampling (TSS)**, a mechanism that probes the temporal dimension during sensitivity estimation.

TSS injects a linearly annealed Gaussian perturbation into the timestamp input of the deformation function, enabling the model to evaluate sensitivities over nearby motion states:

$$(\mu + \Delta\mu, r + \Delta r, s + \Delta s) = \mathcal{D}(\mu, r, s, t + \mathcal{X}(i)), \quad (7)$$

$$\mathcal{X}(i) = \mathcal{N}(0, 1) \cdot \beta \cdot \Delta t \cdot (1 - i/\tau)$$

where \mathcal{D} is the deformation function, $\mathcal{N}(0, 1)$ is standard normal noise, $\beta=0.1$ controls perturbation magnitude, Δt is the frame interval, and $\tau=20,000$ is the annealing period. In our TSS approach, we extend our TSP method by evaluating sensitivities at perturbed motion states defined by Equation 7. Since Equation 6 requires no ground-truth supervision, TSS can compute these sensitivities in a fully self-supervised manner by sampling any temporally jittered motion state. Appendix A.4 ablates our choice of ($\beta=0.1, \tau=20,000$).

This temporal perturbation reveals Gaussians that appear well-behaved on training frames but behave inconsistently under small motion shifts – a hallmark of floaters. Such primitives receive lower sensitivity scores and are pruned, improving spatial compactness and temporal coherence. Stronger perturbations early in training encourage temporal exploration and the removal of floaters during densification, while annealing the noise to zero in later iterations focuses optimization on precise reconstruction of the observed frames. This schedule balances temporal robustness with reconstruction accuracy, yielding stable and deterministic evaluations.

Empirically, TSS improves pruning stability, suppresses floating artifacts, and enhances temporal smoothness. When combined with TSP, it produces sharper, more temporally consistent reconstructions than standard gradient-based pruning approaches, as shown in Figure 3. The impact of TSP and TSS is further analyzed in Section 5.1, demonstrating their complementary effects on efficiency and fidelity.

4.2. GroupFlow

While pruning substantially reduces computation, dynamic Gaussian Splatting still incurs high inference cost because the deformation network must predict motion for every Gaussian. A more efficient strategy is to distill learned neural motion into SE(3) transformations that encode rigid motion. However, assigning a distinct SE(3) transformation to each Gaussian would be prohibitively expensive in both memory and parameter count, making this approach impractical for large or complex scenes.

To overcome this limitation, we introduce **GroupFlow**, illustrated in Figure 4, which clusters Gaussians with similar motion trajectories and assigns each cluster a shared SE(3) transformation. Many dynamic objects in real scenes exhibit locally rigid motion, making groupwise modeling both natural and interpretable. By representing coherent motion with a single transformation rather than individual per-Gaussian predictions, GroupFlow reduces the number of motion trajectories from N (one per Gaussian) to J (one per group), balancing reconstruction fidelity and computational efficiency. This neural-to-rigid distillation yields a compact motion representation that accelerates rendering while maintaining temporal coherence and high visual quality.

4.2.1. Flow Grouping via Motion Analysis

To model coherent motion compactly, GroupFlow analyzes the trajectories of deformable Gaussians to identify clusters with similar motion. Starting from a dense deformable 3D Gaussian model \mathcal{G} , the motion \mathcal{M}_i of each Gaussian \mathcal{G}_i is represented as a sequence of means $\mu_i^t \in \mathbb{R}^3$ and rotation quaternions $r_i^t \in \mathbb{S}^3$ across F timesteps:

$$\mathcal{M} = \{\mathcal{M}_i\}_{i=1}^N, \quad \mathcal{M}_i = \{\mu_i^t, r_i^t\}_{t=0}^{F-1}, \quad (8)$$

where $N = |\mathcal{G}|$ is the total number of Gaussians. This unified 4D Gaussian representation enables similarity comparisons across trajectories.

We designate the first timestep $t = 0$ as the canonical frame to ensure temporal consistency. To form motion-based groups, we select J Gaussian means $h_j^t \in \mathbb{R}^3$ as control

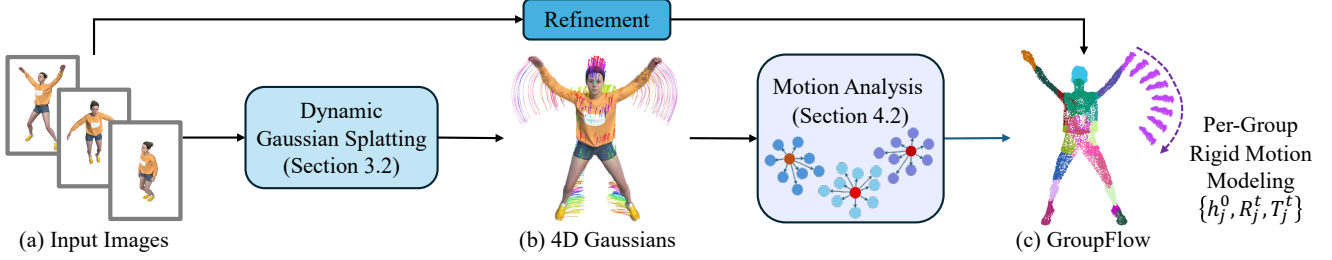


Figure 4. **Overview of our GroupFlow method.** Given a dynamic Gaussian Splatting model \mathcal{G} , we identify a subset of Gaussians as control points and assign each Gaussian to the control point h_j with the most similar motion trajectory. The motion of each group is then estimated via a rigid transformation $[R_j^t|T_j^t]$ at each timestep, reducing inference from per-Gaussian to per-group deformation.

points via farthest point sampling on the set of means at $t = 0$. All Gaussians \mathcal{G}_i are then assigned to the control point h_j with the most similar trajectory obtained via:

$$\arg \min_{j \in \{1, \dots, J\}} S_{i,j}, \quad (9)$$

$$S_{i,j} = \lambda_r \text{std}_t(\|\mu_i^t - h_j^t\|) + (1 - \lambda_r) \text{mean}_t(\|\mu_i^t - h_j^t\|),$$

where $S_{i,j}$ is the trajectory similarity score between μ_i and control point h_j , $\text{std}_t(\cdot)$ and $\text{mean}_t(\cdot)$ denote the standard deviation and mean of time-varying residual $\mu_i^t - h_j^t \in \mathbb{R}^3$, and $\lambda_r = 0.5$ is an empirically selected weighting ratio. All means assigned to the control point h_j form a partitioning group \mathcal{M}^j on \mathcal{M} .

Next, we estimate a time-varying rigid transformation for each group to capture its motion across time. For each group $j \in \{1, \dots, J\}$ and time $t > 0$, we estimate an SE(3) rigid transformation $[R_j^t|T_j^t]$ that maps this group from the canonical frame at time $t = 0$ to time t using Umeyama alignment [41, 42]. Specifically, we randomly sample a subset of means $\mathcal{M}_{\text{samp}}^j$ ($|\mathcal{M}_{\text{samp}}^j| = \min\{N_j, N_{\text{max}}\}$) from each group \mathcal{M}^j and estimate their rigid transformation via:

$$\arg \min_{R_j^t, T_j^t} \sum_{\mu_i \in \mathcal{M}_{\text{samp}}^j} \|\mu_i^t - (R_j^t(\mu_i^0 - h_j^0) + h_j^0 + T_j^t)\|^2, \quad (10)$$

where μ_i^t is the mean at time t , μ_i^0 is the mean in the canonical frame at time $t = 0$, $N_j = |\mathcal{M}^j|$ is the number of Gaussians in group \mathcal{M}^j , and $N_{\text{max}} = 100$ is an empirically selected threshold representing the maximum number of means sampled per group. These group-wise motion parameters $\{h_j^0, R_j^t, T_j^t\}$ represent the shared flow of the mean μ_i – and thus Gaussian \mathcal{G}_i – within group \mathcal{M}^j over time.

4.2.2. Flow Group Training and Inference

To predict the motion parameters $\{\mu_i, r_i\} \in \mathcal{M}^j$ at timestep t , we apply an SE(3) rigid transformation $[R_j^t|T_j^t]$ to the canonical μ_i^0 , relative to its assigned control point h_j^0 :

$$\mu_i^t = R_j^t(\mu_i^0 - h_j^0) + h_j^0 + T_j^t. \quad (11)$$

We also apply group-wise rotation R_j^t to per-Gaussian rotation at canonical state r_i^0 :

$$r_i^t = \text{quat}(R_j^t \text{mat}(r_i^0)), \quad (12)$$

where $\text{mat}(\cdot)$ converts quaternions to rotation matrices and $\text{quat}(\cdot)$ performs the inverse. We set the shared flow $\{h_j^0, R_j^t, T_j^t\}$ of each group \mathcal{M}^j as learnable parameters.

Our **GroupFlow** approach reduces the number of transformations predicted per timestep from N (one per Gaussian) to J (one per group) by enabling all Gaussians within a group to share the same motion parameters. In Appendix A.3, we ablate J and select $J = 2048$ groups to balance visual fidelity with total model size.

By distilling neural motion into grouped rigid transformations, GroupFlow significantly reduces deformation inference cost while maintaining high-quality motion reconstruction. Beyond efficiency, GroupFlow also provides a regularizing effect on motion learning. As discussed in Section 5.2, it can stabilize reconstruction in scenes with noisy or inconsistent motion and even improve visual fidelity over the baseline despite using fewer motion parameters. These results suggest that grouping similar motion trajectories not only compresses the deformation field but also enhances robustness to imperfect motion supervision.

4.3. Speedy Deformable 3D Gaussian Splatting

We integrate our methods into DeformableGS [53] to form **Speedy Deformable 3D Gaussian Splatting (SpeeDe3DGS)**, an efficient pipeline for dynamic scene reconstruction. During training, we apply **Temporal Sensitivity Pruning (TSP)** and **Temporal Sensitivity Sampling (TSS)** to progressively adapt model capacity to scene complexity. This two-phase strategy – soft pruning during densification and hard pruning afterward – removes redundant Gaussians and reduces deformation overhead, enabling real-time rendering while preserving the visual fidelity of neural motion models. After densification, we apply **GroupFlow** to further accelerate inference by distilling the learned deformation field into grouped SE(3) transformations. These shared groupwise motions replace per-Gaussian neural predictions, yielding rendering speeds that surpass prior dynamic Gaussian Splatting methods while maintaining high reconstruction quality and compact model size. We ablate the components of SpeeDe3DGS and compare them with other dynamic Gaussian Splatting approaches in Section 5.

Table 1. **Results on the seven scenes in the real-world NeRF-DS dataset [52] with our SpeeDe3DGS framework.** *TSP*, *TSS*, and *GF* denote Temporal Sensitivity Pruning, Sampling, and GroupFlow, respectively. *Size* measures the combined deformation network and point cloud storage. Each experiment is run three times and averaged to reduce training variance. The **best** and **second-best** results are highlighted. FPS and Train Time are measured on RTX 3090 and RTX A5000 GPUs, respectively. Appendix A.6 reports per-scene results.

TSP	TSS	GF	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Size (MB) \downarrow	# Gaussians \downarrow	Train Time (s) \downarrow
			23.80	0.8503	0.1781	54.37 (1.00 \times)	33.21 (1.00 \times)	132.22K (1.00 \times)	1523.83 (1.00 \times)
	✓		23.78	0.8507	0.1863	346.96 (6.38 \times)	4.52 (7.35 \times)	10.90K (12.13 \times)	741.66 (2.05 \times)
	✓	✓	23.81	0.8515	0.1853	345.24 (6.35 \times)	4.55 (7.29 \times)	11.06K (11.95 \times)	750.69 (2.03 \times)
		✓	23.54	0.8433	0.1892	406.21 (8.58 \times)	51.00 (0.65 \times)	132.32K (1.00 \times)	826.75 (1.84 \times)
✓	✓	✓	23.66	0.8487	0.1901	505.60 (10.68 \times)	21.40 (1.55 \times)	11.10K (11.91 \times)	625.48 (2.44 \times)

5. Experiments

We evaluate our approach in two complementary settings: (i) our **Speedy Deformable 3D Gaussian Splatting (SpeeDe3DGS) framework**, which extends DeformableGS [53] for detailed ablation and analysis, and (ii) the **Monocular Dynamic Gaussian Splatting Benchmark (MonoDyGauBench) [25]**, which standardizes comparisons across motion representations. Both settings integrate the complete SpeeDe3DGS pipeline – Temporal Sensitivity Pruning (TSP), Temporal Sensitivity Sampling (TSS), and GroupFlow – into dynamic Gaussian Splatting.

For consistent runtime comparison, all FPS metrics are measured on an NVIDIA RTX 3090 GPU. Unless otherwise specified, experiments follow each framework’s default hyperparameters and train for a total of 30,000 iterations. Starting at iteration 6,000, we apply TSP every 3,000 iterations, soft pruning 60% of Gaussians at a time during densification and hard pruning 30% after densification. The impact of different soft and hard pruning percentages is analyzed in Appendix A.5. Temporal perturbations for TSS are annealed using $\beta=0.1$ and $\tau=20,000$, and GroupFlow is initialized after densification at iteration 15,000 with $J=2048$ motion groups. The influence of group count on performance and model size is examined in Appendix A.3. This configuration enables detailed ablations within SpeeDe3DGS and ensures fair cross-method comparisons under MonoDyGauBench.

5.1. SpeeDe3DGS Results

We first evaluate the components of SpeeDe3DGS on the real-world NeRF-DS dataset [52] in Table 1. Applying **TSP** alone accelerates rendering by 6.38 \times , reduces the number of Gaussians by 12.13 \times , and shortens training time by 2.05 \times on average, while maintaining the image quality of DeformableGS [53]. Adding **TSS** further improves PSNR and SSIM beyond the baseline, as temporal perturbations act as a regularizer that suppresses floaters and stabilizes motion. Qualitative comparisons in Figure 3 show that TSP and TSS together effectively remove floaters and improve temporal coherence without sacrificing sharpness.

Using **GroupFlow** without pruning yields an 8.58 \times ren-

dering speedup with a modest drop in image quality, reflecting the trade-off of representing motion through grouped rigid transformations rather than full neural deformation. When combined with pruning (**TSP + TSS + GroupFlow**), rendering accelerates by 10.68 \times and training time shortens by 2.44 \times . Notably, the pruned representation improves GroupFlow’s motion grouping, achieving higher PSNR and SSIM than GroupFlow alone. Even with additional GroupFlow parameters, the full SpeeDe3DGS configuration remains 1.55 \times more compact than the baseline model.

Results on the synthetic D-NeRF [39] and real-world HyperNeRF [37] datasets are provided in Appendix A.2. On HyperNeRF, TSP + TSS and GroupFlow individually achieve 9.37 \times and 15.66 \times rendering speedups, respectively. When used together, they reach 29.21 \times faster rendering, 12.18 \times Gaussian reduction, and 3.74 \times shorter training time, demonstrating the synergy of our methods. Training times for NeRF-DS and D-NeRF are measured on RTX A5000 GPUs (24 GB). We use RTX A6000 GPUs (48 GB) to train HyperNeRF due to the baseline memory requirements; however, pruning enables efficient training on 24 GB of VRAM. Per-scene metrics for all datasets are provided in Appendix A.6.

5.2. MonoDyGauBench Evaluation

We evaluate our approach on the 50 scenes in **MonoDyGauBench [25]**, extending the high-fidelity yet computationally heavy DeformableGS [53] and 4DGS [49] baselines with our **Pruning (TSP + TSS)** and **GroupFlow** modules. Table 2 summarizes aggregate results across all scenes, while Appendix A.1 provides per-dataset breakdowns. Qualitative results are presented in Appendix A.7. Although MonoDyGauBench experiments are nominally conducted on an RTX 3090 GPU, several baseline scenes exceed its 24 GB memory limit – a trend also noted in Section 5.1. Our pruning integration markedly reduces memory usage, enabling training within 24 GB on an RTX A5000 GPU.

As shown in Table 2, pruning makes **4DGS** on par with the fastest non-neural baselines while maintaining high image quality across metrics. Adding GroupFlow on top of pruning achieves the highest rendering speed – over 100 FPS faster than the fastest baseline – but with some loss of im-

Table 2. **Results on Monocular Dynamic Gaussian Splatting Benchmark (MonoDyGauBench) [25]**. Quantitative results averaged across five datasets and 50 scenes for all methods in Section 3.2. We *cumulatively* apply our SpeeDe3DGS methods to the DeformableGS [53] and 4DGS [49] baselines, keeping the original neural variants with **low FPS** for reference, but excluding them from comparisons to focus on real-time methods. Pruning is performed using TSP and TSS. Each experiment is repeated three times and averaged. The **best** and **second-best** results are highlighted; improvements over corresponding baselines are **bolded**. FPS and baseline Train Time are measured on an RTX 3090 GPU, while our Train Time* is measured on an RTX A5000 (both 24 GB). Per-dataset results are reported in Appendix A.1.

Method	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Train Time (s) \downarrow
EffGS [19]	21.84	0.672	0.725	0.347	177.21	3757.81
STG-decoder [24]	21.81	0.678	0.742	0.352	109.42	5980.64
STG [24]	19.51	0.583	0.643	0.475	181.70	5359.56
RTGS [54]	21.61	0.663	0.720	0.350	143.37	7352.52
4DGS [49]	23.55	0.708	0.765	0.277	62.99 (1.00 \times)	8628.89 (1.00 \times)
+ Pruning (Ours)	22.44	0.689	0.737	0.334	179.64 (2.85\times)	4358.17* (1.47\times)
+ GroupFlow (Ours)	21.00	0.667	0.705	0.380	290.21 (4.61\times)	4176.49* (2.07\times)
DeformableGS [53]	24.07	0.694	0.755	0.283	20.20 (1.00 \times)	6227.43 (1.00 \times)
+ Pruning (Ours)	23.86	0.694	0.749	0.295	137.01 (6.78\times)	2850.60* (2.18\times)
+ GroupFlow (Ours)	23.52	0.709	0.771	0.313	276.91 (13.71\times)	2461.14* (2.53\times)

age quality, suggesting that per-Gaussian inference in DeformableGS better preserves fidelity under grouped motion.

Integrating pruning into **DeformableGS** accelerates rendering by 6.78 \times , reaching frame rates comparable to RTGS [26] while maintaining high visual fidelity across all datasets. It surpasses all non-neural baselines and our 4DGS variants in overall image quality, reinforcing DeformableGS as a strong backbone for neural motion modeling. Adding GroupFlow on top of pruning further increases rendering speed to 13.71 \times , making it nearly 100 FPS faster than all baseline methods and on-par with our 4DGS+Pruning+GroupFlow variant. Training time is also the shortest, reduced by 2.53 \times . Despite these speedups, reconstruction quality remains high: SSIM and MS-SSIM exceed the original DeformableGS baseline, suggesting a mild regularizing effect from GroupFlow. As discussed in Appendix A.1, this effect is most pronounced in scenes with unstable pose estimation. Overall, our SpeeDe3DGS strategies achieve an effective balance between speed, fidelity, and efficiency for dynamic Gaussian Splatting.

6. Limitations

While SpeeDe3DGS achieves substantial acceleration with minimal fidelity loss, some trade-offs remain. Minor quality degradation can occur under extreme pruning ratios, though this can be mitigated by adjusting pruning hyperparameters to balance speed and fidelity. Similarly, GroupFlow preserves quality in most scenes through locally rigid motion, but highly deformable regions may lose fidelity when the number of motion groups is limited. We analyze these trade-offs through ablations of pruning percentages and number of groups in Appendices A.5 and A.3. While our methods intro-

duce mild regularizing effects that improve temporal stability, they do not incorporate explicit motion priors or learned dynamics. They are therefore complementary to prior-driven or motion-aware 3DGS frameworks and can accelerate such methods. In practice, TSP and TSS can be applied to any dynamic 3DGS framework, while GroupFlow extends to those with an explicit motion model. Appendix A.8 further discusses baseline and motion model limitations. For a broader analysis of dynamic Gaussian Splatting methods and their limitations, we refer readers to MonoDyGauBench [25].

7. Conclusion

We introduced **Speedy Deformable 3D Gaussian Splatting (SpeeDe3DGS)**, a framework that bridges the efficiency–fidelity gap in dynamic 3D Gaussian Splatting by reducing redundant neural inference through pruning and motion distillation. Our **Temporal Sensitivity Pruning (TSP)** module removes redundant Gaussians via temporally aggregated sensitivity analysis, while **Temporal Sensitivity Sampling (TSS)** stabilizes this process by perturbing timestamps to suppress floaters and enhance temporal coherence. Complementing pruning, **GroupFlow** distills neural motion into shared rigid SE(3) transformations, enabling efficient groupwise deformation and faster convergence. Across the 50 dynamic scenes in MonoDyGauBench [25], pruning accelerates DeformableGS by 6.78 \times in rendering speed with minimal fidelity loss and 10 \times fewer primitives, and adding GroupFlow further improves performance to 13.71 \times faster rendering and 2.53 \times shorter training while preserving superior image quality. Together, these modules demonstrate that temporally aware pruning and groupwise motion distillation can deliver neural-field fidelity at real-time speeds.

Acknowledgments

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 140D0423C0076. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Commercial support was provided by the Amazon Research Awards program and Open Philanthropy. Further support was provided by DARPA TIAMAT and the NSF TRAILS Institute (2229885).

References

- [1] Muhammad Salman Ali, Maryam Qamar, Sung-Ho Bae, and Enzo Tartaglione. Trimming the fat: Efficient compression of 3d gaussian splats through pruning. *arXiv preprint arXiv:2406.18214*, 2024. 2
- [2] Muhammad Salman Ali, Sung-Ho Bae, and Enzo Tartaglione. Elms: Enhancing memory and computation scalability through compression for 3d gaussian splatting. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2591–2600. IEEE, 2025. 2
- [3] Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 130–141, 2023. 2, 4
- [4] Yilun Du, Yan Zhang, Hong-Xing Yu, Joshua B Tenenbaum, and Jiajun Wu. Neural radiance flow for 4d view synthesis and video processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14324–14334, 2021. 2
- [5] Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, Zhangyang Wang, et al. Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. *Advances in neural information processing systems*, 37:140138–140158, 2024. 2
- [6] Guangchi Fang and Bing Wang. Mini-splatting: Representing scenes with a constrained number of gaussians. In *European Conference on Computer Vision*, pages 165–181. Springer, 2024. 2
- [7] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, 2022. 2
- [8] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 2
- [9] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 2
- [10] Hang Gao, Ruilong Li, Shubham Tulsiani, Bryan Russell, and Angjoo Kanazawa. Monocular dynamic view synthesis: A reality check. In *NeurIPS*, 2022. 12, 13, 17
- [11] Quankai Gao, Qiangeng Xu, Hao Su, Ulrich Neumann, and Zexiang Xu. Strivec: Sparse tri-vector radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17569–17579, 2023. 2
- [12] Sharath Girish, Kamal Gupta, and Abhinav Shrivastava. Eagles: Efficient accelerated 3d gaussians with lightweight encodings. In *European Conference on Computer Vision*, pages 54–71. Springer, 2024. 2
- [13] Xiang Guo, Guanying Chen, Yuchao Dai, Xiaoqing Ye, Jiada Sun, Xiao Tan, and Errui Ding. Neural deformable voxel grid for fast optimization of dynamic view synthesis. In *Proceedings of the Asian Conference on Computer Vision*, pages 3757–3775, 2022. 2
- [14] Alex Hanson, Allen Tu, Geng Lin, Vasu Singla, Matthias Zwicker, and Tom Goldstein. Speedy-splat: Fast 3d gaussian splatting with sparse pixels and sparse primitives. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 21537–21546, 2025. 2, 4
- [15] Alex Hanson, Allen Tu, Vasu Singla, Mayuka Jayawardhana, Matthias Zwicker, and Tom Goldstein. Pup 3d-gs: Principled uncertainty pruning for 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 5949–5958, 2025. 2, 4
- [16] Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4220–4230, 2024. 2
- [17] Saqib Javed, Ahmad Jarrar Khan, Corentin Dumery, Chen Zhao, and Mathieu Salzmann. Temporally compressed 3d gaussian splatting for dynamic scenes, 2024. 2
- [18] Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherf: Active view selection and uncertainty quantification for radiance fields using fisher information. *arXiv*, 2023. 4
- [19] Kai Katsumata, Duc Minh Vo, and Hideki Nakayama. A compact dynamic 3d gaussian representation for real-time dynamic view synthesis. In *European Conference on Computer Vision*, pages 394–412. Springer, 2024. 2, 4, 8, 12, 13
- [20] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 1, 2, 3
- [21] Agelos Kratimenos, Jiahui Lei, and Kostas Daniilidis. Dynmf: Neural motion factorization for real-time dynamic view synthesis with 3d gaussian splatting. In *European Conference on Computer Vision*, pages 252–269. Springer, 2024. 2
- [22] Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21719–21728, 2024. 2
- [23] Jiahui Lei, Yijia Weng, Adam Harley, Leonidas Guibas, and Kostas Daniilidis. Mosca: Dynamic gaussian fusion

- from casual videos via 4d motion scaffolds. *arXiv preprint arXiv:2405.17421*, 2024. 2
- [24] Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8508–8520, 2024. 2, 4, 8, 12, 13
- [25] Yiqing Liang, Mikhail Okunev, Mikaela Angelina Uy, Runfeng Li, Leonidas Guibas, James Tompkin, and Adam W Harley. Monocular dynamic gaussian splatting: Fast, brittle, and scene complexity rules. *Transactions on Machine Learning Research*, 2025. Survey Certification. 1, 2, 4, 7, 8, 12, 13, 14, 16, 17
- [26] Weikai Lin, Yu Feng, and Yuhao Zhu. Rtgs: Enabling real-time gaussian splatting on mobile devices using efficiency-guided pruning and foveated rendering. *arXiv preprint arXiv:2407.00435*, 2024. 2, 8
- [27] Youtian Lin, Zuozhuo Dai, Siyu Zhu, and Yao Yao. Gaussian flow: 4d reconstruction with dynamic 3d gaussian particle. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21136–21145, 2024. 2
- [28] Hengyu Liu, Yifan Liu, Chenxin Li, Wuyang Li, and Yixuan Yuan. Lgs: A light-weight 4d gaussian splatting for efficient surgical scene reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 660–670. Springer, 2024. 2
- [29] Jia-Wei Liu, Yan-Pei Cao, Weijia Mao, Wenqiao Zhang, David Junhao Zhang, Jussi Keppo, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. Devrf: Fast deformable voxel radiance fields for dynamic scenes. *Advances in Neural Information Processing Systems*, 35:36762–36775, 2022. 2
- [30] Xiangrui Liu, Xinju Wu, Pingping Zhang, Shiqi Wang, Zhu Li, and Sam Kwong. Compgs: Efficient 3d scene representation via compressed gaussian splatting. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2936–2944, 2024. 2
- [31] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *3DV*, 2024. 2
- [32] Marko Mihajlovic, Sergey Prokudin, Marc Pollefeys, and Siyu Tang. Resfields: Residual neural fields for spatiotemporal signals. *International Conference on Learning Representations*, 2024. 2
- [33] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2
- [34] Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, and Federico Tombari. Radsplat: Radiance field-informed gaussian splatting for robust real-time rendering with 900+ fps. In *International Conference on 3D Vision 2025*, 2025. 2
- [35] Panagiotis Papantonakis, Georgios Kopanas, Bernhard Kerbl, Alexandre Lanvin, and George Drettakis. Reducing the memory footprint of 3d gaussian splatting. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–17, 2024. 2
- [36] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5865–5874, 2021. 2, 12, 13, 17
- [37] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph.*, 40(6), 2021. 1, 2, 7, 12, 13, 14, 15, 16, 17
- [38] Sida Peng, Yunzhi Yan, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Representing volumetric videos as dynamic mlp maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4252–4262, 2023. 2
- [39] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2, 3, 7, 12, 13, 14, 15, 16
- [40] Sara Sabour, Lily Goli, George Kopanas, Mark Matthews, Dmitry Lagun, Leonidas Guibas, Alec Jacobson, David Fleet, and Andrea Tagliasacchi. Spotlessplats: Ignoring distractors in 3d gaussian splatting. *ACM Transactions on Graphics*, 44(2):1–11, 2025. 2
- [41] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, pages 109–116. Citeseer, 2007. 6
- [42] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(04):376–380, 1991. 6
- [43] Diwen Wan, Ruijie Lu, and Gang Zeng. Superpoint gaussian splatting for real-time high-fidelity dynamic scene reconstruction. In *Proceedings of the 41st International Conference on Machine Learning*, pages 49957–49972, 2024. 2
- [44] Chaoyang Wang, Ben Eckart, Simon Lucey, and Orazio Gallo. Neural trajectory fields for dynamic novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 2
- [45] Feng Wang, Zilong Chen, Guokang Wang, Yafei Song, and Huaping Liu. Masked space-time hash encoding for efficient dynamic scene reconstruction. *Advances in neural information processing systems*, 2023. 2
- [46] Qianqian Wang, Vickie Ye, Hang Gao, Weijia Zeng, Jake Austin, Zhengqi Li, and Angjoo Kanazawa. Shape of motion: 4d reconstruction from a single video. In *International Conference on Computer Vision (ICCV)*, 2025. 2
- [47] Rui Wang, Quentin Lohmeyer, Mirko Meboldt, and Siyu Tang. Degauss: Dynamic-static decomposition with gaussian splatting for distractor-free 3d reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6294–6303, 2025. 2
- [48] Joey Wilson, Marcelino Almeida, Sachit Mahajan, Martin Labrie, Maani Ghaffari, Omid Ghasemalizadeh, Min Sun,

- Cheng-Hao Kuo, and Arnab Sen. Pop-gs: Next best view in 3d-gaussian splatting with p-optimality. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3646–3655, 2025. [4](#)
- [49] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20310–20320, 2024. [1](#), [2](#), [4](#), [7](#), [8](#), [12](#), [13](#)
- [50] Jiahao Wu, Rui Peng, Zhiyan Wang, Lu Xiao, Luyang Tang, Jinbo Yan, Kaiqiang Xiong, and Ronggang Wang. Swift4d: Adaptive divide-and-conquer gaussian splatting for compact and efficient reconstruction of dynamic scene. *arXiv preprint arXiv:2503.12307*, 2025. [2](#)
- [51] Kai Xu, Tze Ho Elden Tse, Jizong Peng, and Angela Yao. Das3r: Dynamics-aware gaussian splatting for static scene reconstruction. *arXiv preprint arxiv:2412.19584*, 2024. [2](#)
- [52] Zhiwen Yan, Chen Li, and Gim Hee Lee. Nerf-ds: Neural radiance fields for dynamic specular objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8285–8295, 2023. [3](#), [5](#), [7](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [53] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20331–20341, 2024. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#)
- [54] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2024. [2](#), [4](#), [8](#), [12](#), [13](#)
- [55] Xinjie Zhang, Zhening Liu, Yifan Zhang, Xingtong Ge, Dailan He, Tongda Xu, Yan Wang, Zehong Lin, Shuicheng Yan, and Jun Zhang. Mega: Memory-efficient 4d gaussian splatting for dynamic scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 27828–27838, 2025. [2](#)
- [56] Zhaoliang Zhang, Tianchen Song, Yongjae Lee, Li Yang, Cheng Peng, Rama Chellappa, and Deliang Fan. Lp-3dgs: Learning to prune 3d gaussian splatting. In *Advances in Neural Information Processing Systems*, pages 122434–122457. Curran Associates, Inc., 2024. [2](#)
- [57] Hongyu Zhou, Jiahao Shao, Lu Xu, Dongfeng Bai, Weichao Qiu, Bingbing Liu, Yue Wang, Andreas Geiger, and Yiyi Liao. Hugs: Holistic urban 3d scene understanding via gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21336–21345, 2024. [2](#)
- [58] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002. [3](#)

A. Appendix

A.1. Per-Dataset Metrics for MonoDyGauBench

We analyze results on Monocular Dynamic Gaussian Splatting Benchmark (MonoDyGauBench) [25], which evaluates dynamic 3DGS variants across five datasets. For consistency, we directly use the baseline tables from the manuscript. Both **4DGS** and **DeformableGS** – the neural motion-field baselines – are extended with our pruning (TSP+TSS) and GroupFlow modules. Performance differences with our SpeeDe3DGS framework arise because MonoDyGauBench standardizes all methods within a unified training wrapper, integrating every motion representation and our components under common hyperparameters. Although MonoDyGauBench experiments are nominally conducted on an RTX 3090 GPU, baseline scenes in the HyperNeRF, Nerfies, and iPhone datasets exceed its 24 GB memory limit. Our pruning integration markedly reduces memory usage, enabling training within 24 GB on an RTX A5000 GPU. Qualitative results are presented in Appendix A.7.

On **D-NeRF** [39], Table 3 shows that pruning and GroupFlow preserve the high neural fidelity of both baselines, surpassing all non-neural methods by a wide margin. Interestingly, adding GroupFlow slightly decreases rendering and training speed for DeformableGS – a behavior unique to D-NeRF within MonoDyGauBench and not observed in our SpeeDe3DGS framework results in Appendix A.2.

On **HyperNeRF** [37], Table 4 shows that adding pruning to DeformableGS improves PSNR and SSIM beyond the baseline with an $8.00\times$ rendering speedup, while adding GroupFlow further enhances SSIM and MS-SSIM, culminating in a $23.84\times$ speedup. Efficiency improvements with our SpeeDe3DGS framework on HyperNeRF are reported in Appendix A.2.

On **NeRF-DS** [52], Table 5 shows that adding TSP and TSS yields comparable or improved image quality over the DeformableGS baseline, while adding GroupFlow provides a $10.49\times$ rendering acceleration with minimal fidelity loss. These trends are consistent with our findings in Section 5.1. The 4DGS baseline performs comparatively poorly, consistent with MonoDyGauBench’s broader analysis of motion–representation trade-offs.

On **Nerfies** [36], Table 6 shows that pruning improves all quality metrics for DeformableGS, demonstrating the stabilizing effects of TSP+TSS. Adding GroupFlow further increases efficiency, achieving a $19.50\times$ rendering speedup and the highest FPS in the table, while maintaining higher image quality than all baselines.

On **iPhone** [10], Table 7 shows that adding GroupFlow significantly improves PSNR, SSIM, and MS-SSIM while achieving a $26.34\times$ rendering speedup. By enforcing locally consistent motion through shared rigid SE(3) transformations, GroupFlow acts as a strong structural regularizer that

Table 3. **Results for the eight scenes in the synthetic D-NeRF dataset [39] with MonoDyGauBench [25].** Quantitative results for all methods in Section 3.2. We *cumulatively* apply our SpeeDe3DGS methods to the DeformableGS [53] and 4DGS [49] baselines, keeping the original neural variants with **low FPS** for reference, but excluding them from comparisons to focus on real-time methods. Pruning is performed using TSP and TSS. Each experiment is repeated three times and averaged. The **best** and **second-best** results are highlighted; improvements over corresponding baselines are **bolded**. FPS and baseline Train Time are measured on an RTX 3090 GPU, while our Train Time* is measured on an RTX A5000 (both 24 GB). Results with our SpeeDe3DGS framework are reported in Table 8.

Method	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Train Time (s) \downarrow
EffGS [19]	30.52	0.96	0.98	0.04	289.47	1042.04
STG-decoder [24]	25.89	0.91	0.90	0.17	160.32	3462.29
STG [24]	17.09	0.88	0.66	0.29	208.98	4889.50
RTGS [54]	28.78	0.96	0.96	0.05	192.37	1519.60
4DGS [49]	33.27	0.98	0.99	0.02	134.13 (1.00 \times)	1781.46 (1.00 \times)
+ Pruning (Ours)	32.99	0.98	0.99	0.02	238.98 (1.78 \times)	1263.30* (1.41 \times)
+ GroupFlow (Ours)	31.70	0.97	0.99	0.03	264.61 (1.97 \times)	1296.21* (1.37 \times)
DeformableGS [53]	37.14	0.99	0.99	0.01	50.78 (1.00 \times)	2048.38 (1.00 \times)
+ Pruning (Ours)	36.25	0.99	0.99	0.01	294.44 (5.80 \times)	1241.56* (1.65 \times)
+ GroupFlow (Ours)	35.18	0.99	0.99	0.01	275.32 (5.42 \times)	1164.55* (1.76 \times)

Table 4. **Results for the 17 scenes in the real-world HyperNeRF dataset [37] with MonoDyGauBench [25].** Results with our SpeeDe3DGS framework are reported in Table 9.

Method	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Train Time (s) \downarrow
EffGS [19]	22.24	0.70	0.79	0.37	138.79	4119.66
STG-decoder [24]	23.92	0.73	0.83	0.34	66.72	7423.41
STG [24]	22.92	0.70	0.78	0.41	183.21	5729.80
RTGS [54]	22.99	0.71	0.79	0.35	104.64	11507.80
4DGS [49]	25.70	0.79	0.89	0.23	37.46 (1.00 \times)	10170.86 (1.00 \times)
+ Pruning (Ours)	25.47	0.78	0.88	0.28	157.91 (4.22 \times)	5623.29* (1.81 \times)
+ GroupFlow (Ours)	23.98	0.75	0.85	0.31	269.57 (7.20 \times)	5871.93* (1.73 \times)
DeformableGS [53]	24.58	0.74	0.83	0.27	10.91 (1.00 \times)	8855.46 (1.00 \times)
+ Pruning (Ours)	24.83	0.75	0.82	0.31	87.27 (8.00 \times)	3668.02* (2.41 \times)
+ GroupFlow (Ours)	24.13	0.75	0.85	0.32	260.05 (23.84 \times)	3361.70* (2.63 \times)

Table 5. **Results for the seven scenes in the real-world NeRF-DS dataset [52] with MonoDyGauBench [25].** Results with our SpeeDe3DGS framework are reported in Table 1.

Method	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Train Time (s) \downarrow
EffGS [19]	21.28	0.78	0.77	0.25	307.70	1597.90
STG-decoder [24]	21.73	0.80	0.81	0.20	212.24	2131.48
STG [24]	20.13	0.69	0.72	0.39	302.89	2214.62
RTGS [54]	19.88	0.75	0.73	0.30	259.83	3116.21
4DGS [49]	20.08	0.73	0.71	0.25	100.22 (1.00 \times)	4075.43 (1.00 \times)
+ Pruning (Ours)	20.27	0.74	0.73	0.24	255.48 (2.55 \times)	2787.93* (1.46 \times)
+ GroupFlow (Ours)	20.14	0.73	0.72	0.25	323.86 (3.23 \times)	2712.69* (1.50 \times)
DeformableGS [53]	23.42	0.84	0.88	0.16	30.27 (1.00 \times)	2885.07 (1.00 \times)
+ Pruning (Ours)	23.49	0.84	0.89	0.15	255.48 (8.44 \times)	1488.77* (1.94 \times)
+ GroupFlow (Ours)	23.36	0.84	0.88	0.15	317.47 (10.49 \times)	1556.77* (1.85 \times)

enhances temporal coherence and mitigates drift under noisy or unstable camera trajectories.

A.2. Additional SpeeDe3DGS Results

We extend our analysis to the synthetic D-NeRF [39] and real-world HyperNeRF [37] datasets using the **DeformableGS** [53] codebase. Rendering FPS is measured on an RTX 3090 GPU with 24 GB of VRAM for consistency with MonoDyGauBench [25].

On **D-NeRF** [39], both pruning and GroupFlow individually provide around a $3\times$ increase in rendering speed and a

Table 6. Results for the four scenes in the real-world Nerfies dataset [36] with MonoDyGauBench [25].

Method	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Train Time (s) \downarrow
EffGS [19]	20.13	0.43	0.61	0.65	159.17	3249.33
STG-decoder [24]	20.93	0.47	0.67	0.58	90.46	6050.42
STG [24]	20.55	0.46	0.66	0.62	142.30	5264.58
RTGS [54]	20.06	0.42	0.62	0.61	153.38	9059.54
4DGS [49]	21.84	0.50	0.72	0.47	35.70 (1.00 \times)	9797.50 (1.00 \times)
+ Pruning (Ours)	21.79	0.49	0.71	0.55	143.74 (4.03 \times)	5760.44* (1.70 \times)
+ GroupFlow (Ours)	20.59	0.46	0.66	0.58	280.14 (7.85 \times)	5776.24* (\times)
DeformableGS [53]	21.26	0.43	0.66	0.54	14.99 (1.00 \times)	6950.94 (1.00 \times)
+ Pruning (Ours)	21.76	0.46	0.69	0.51	85.21 (5.68 \times)	3935.12* (1.77 \times)
+ GroupFlow (Ours)	21.35	0.46	0.67	0.58	284.86 (19.00 \times)	3001.68* (2.32 \times)

Table 7. Results for the 14 scenes in the real-world iPhone dataset [10] with MonoDyGauBench [25].

Method	PSNR \uparrow	SSIM \uparrow	MS-SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Train Time (s) \downarrow
EffGS [19]	16.82	0.47	0.50	0.47	99.64	6275.33
STG-decoder [24]	16.85	0.47	0.52	0.49	86.18	7694.83
STG [24]	15.60	0.40	0.45	0.57	114.95	6629.64
RTGS [54]	15.36	0.39	0.44	0.52	101.33	8484.05
4DGS [49]	15.13	0.42	0.43	0.52	42.53 (1.00 \times)	14205.48 (1.00 \times)
+ Pruning (Ours)	14.00	0.44	0.43	0.56	144.45 (3.40 \times)	4974.91* (2.86 \times)
+ GroupFlow (Ours)	11.83	0.42	0.38	0.68	315.97 (7.43 \times)	4038.46* (\times)
DeformableGS [53]	16.56	0.46	0.48	0.45	10.47 (1.00 \times)	7278.28* (1.00 \times)
+ Pruning (Ours)	16.40	0.45	0.47	0.45	74.26 (7.09 \times)	3148.54* (2.31 \times)
+ GroupFlow (Ours)	16.82	0.51	0.53	0.48	275.75 (26.34 \times)	2406.27* (3.02 \times)

1.5 \times reduction in training time, measured on an RTX A5000 GPU (24 GB). Pruning slightly reduces reconstruction quality more than GroupFlow, likely because D-NeRF’s synthetic scenes are compact, noise-free, and already sparsely parameterized, leaving limited redundancy to remove. Nevertheless, pruning achieves a 13.79 \times reduction in the number of Gaussians. Since these scenes contain no pose noise or floaters, TSP and TSP+TSS perform nearly identically, showing that TSS adds stability without affecting clean synthetic settings. GroupFlow performs particularly well due to the simple, rigid motions of D-NeRF scenes. When combined, pruning and GroupFlow achieve a 4.11 \times rendering speedup and a 1.93 \times reduction in training time.

For HyperNeRF [37], we report results on the same eight real-world scenes used in DeformableGS – *espresso*, *americano*, *cookie*, *chicken*, *torchocolate*, *lemon*, *hand*, and *printer*. Following DeformableGS, image-quality metrics are not reported because most scenes lack evaluation splits; we provide full results on MonoDyGauBench in Table 4. Pruning alone yields a 9.37 \times rendering speedup, an 11.88 \times reduction in Gaussians, and a 2.67 \times reduction in training time. GroupFlow independently achieves a 15.66 \times rendering speedup and a 2.10 \times shorter training time. Together, they reach 29.21 \times faster rendering and 3.74 \times shorter training, demonstrating strong synergy between our SpeeDe3DGS components. Baseline HyperNeRF models require an RTX A6000 GPU with 48 GB of VRAM for training, whereas our pruned configurations fit comfortably within 24 GB on an RTX A5000. All inference metrics, including FPS, are collected on an RTX 3090 GPU (24 GB), matching the MonoDyGauBench setup and underscoring that baseline models exceed 24 GB memory during training despite being evaluated on an RTX 3090 for inference.

A.3. Ablation on Number of Groups

We ablate the number of motion groups J for GroupFlow on the NeRF-DS [52] and D-NeRF [39] datasets using the DeformableGS [53] codebase. Table 10 reports quantitative results averaged over three independent runs to reduce training variance. All FPS and training time measurements are collected on RTX 3090 and RTX A5000 GPUs.

Increasing J improves reconstruction quality by allowing finer motion grouping, though the number of learnable parameters grows linearly with J . We select $J=2048$ as it provides the best balance between fidelity and compactness – the combined size of the pruned DeformableGS and GroupFlow model is smaller than that of the unpruned baseline and its 1.94 MB deformation network. Smaller configurations, such as $J=128$, produce even lighter GroupFlow models than the deformation network but incur a minor drop in quality. Conversely, $J=4096$ slightly degrades performance, suggesting that GroupFlow introduces a mild regularizing effect on SE(3) transformations, as nearby Gaussians naturally share coherent motion. Rendering FPS remains nearly constant across all J due to the efficiency of batched SE(3) transformations. Notably, combining pruning with GroupFlow yields better results than GroupFlow alone across all J values on the NeRF-DS dataset, likely because pruning suppresses noise that would otherwise hinder motion grouping.

A.4. Ablation on TSS Hyperparameters

Table 11 ablates the perturbation magnitude β and annealing period τ for Temporal Sensitivity Sampling (TSS) on the NeRF-DS [52] dataset using the SpeeDe3DGS codebase. TSS improves pruning performance across all configurations, and its hyperparameters are not sensitive because the changes between timesteps in real-time video are small. We adopt the same $\beta=0.1$ and $\tau=20,000$ as DeformableGS [53].

A.5. Ablation on Pruning Percentages

Figure 5 illustrates parameter sweeps over soft (densification-stage) and hard (post-densification) pruning percentages for the NeRF-DS [52] and D-NeRF [39] datasets using the SpeeDe3DGS codebase. Each configuration is evaluated in 5% intervals and repeated three times to reduce variance, with results averaged across all runs. All experiments are performed without Temporal Sensitivity Sampling (TSS) or GroupFlow to isolate the effect of Temporal Sensitivity Pruning (TSP) alone. FPS and training time are collected on an RTX A5000 GPU. We empirically select the (60%, 30%) soft–hard pruning ratio as it provides the best balance between rendering speed and visual quality.

A.6. Per-Scene Metrics for SpeeDe3DGS

Tables 12, 13, and 14 report per-scene PSNR, SSIM, and LPIPS, respectively, using our standalone SpeeDe3DGS codebase on NeRF-DS [52] and D-NeRF [39]. Tables 15 and

Table 8. **Results on the eight scenes in the synthetic D-NeRF dataset [39] with our SpeeDe3DGS framework.** *TSP*, *TSS*, and *GF* denote Temporal Sensitivity Pruning, Sampling, and GroupFlow, respectively. *Size* measures the combined deformation network and point cloud storage. Each experiment is run three times and averaged to reduce training variance. The **best** and **second-best** results are highlighted. FPS and Train Time are measured on RTX 3090 and RTX A5000 GPUs, respectively. Per-scene results are provided in Appendix A.6. Results with MonoDyGauBench [25] are reported in Table 3.

TSP	TSS	GF	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Size (MB) \downarrow	# Gaussians \downarrow	Train Time (s) \downarrow
			38.92	0.9892	0.0143	127.47 (1.00 \times)	16.88 (1.00 \times)	62.92K (1.00 \times)	940.52 (1.00 \times)
✓			36.19	0.9793	0.0349	422.48 (3.31 \times)	3.08 (5.48\times)	4.56K (13.79\times)	524.52 (1.79 \times)
✓	✓		36.19	0.9792	0.0350	423.89 (3.33\times)	3.08 (5.48\times)	4.57K (13.78\times)	524.34 (1.79\times)
		✓	36.85	0.9862	0.0172	374.32 (2.94 \times)	28.17 (0.60 \times)	62.77K (1.00 \times)	636.39 (1.48 \times)
✓	✓	✓	35.07	0.9771	0.0365	524.19 (4.11\times)	13.96 (1.21 \times)	4.56K (13.79\times)	486.10 (1.93\times)

Table 9. **Results on the eight real-world scenes from the HyperNeRF dataset [37] in the DeformableGS [53] paper with our SpeeDe3DGS framework.** *TSP*, *TSS*, and *GF* denote Temporal Sensitivity Pruning, Sampling, and GroupFlow, respectively. *Size* measures the combined deformation network and point cloud storage. Each experiment is run three times and averaged to reduce training variance. The **best** and **second-best** results are highlighted. FPS and Train Time are measured on RTX 3090 and RTX A6000 GPUs, respectively. Per-scene results are reported in Appendix A.6. Results with MonoDyGauBench [25] are reported in Table 4.

TSP	TSS	GF	FPS \uparrow	Size (MB) \downarrow	# Gaussians \downarrow	Train Time (s) \downarrow
			14.48 (1.00 \times)	159.30 (1.00 \times)	665.35K (1.00 \times)	5248.91 (1.00 \times)
✓			141.11 (9.75 \times)	14.98 (10.64\times)	55.12K (12.07\times)	1992.39 (2.63 \times)
✓	✓		135.70 (9.37 \times)	15.18 (10.49\times)	56.00K (11.88 \times)	1968.72 (2.67\times)
		✓	226.77 (15.66\times)	179.79 (0.89 \times)	666.31K (1.00 \times)	2505.44 (2.10 \times)
✓	✓	✓	422.88 (29.21\times)	31.62 (5.04 \times)	54.63K (12.18\times)	1401.85 (3.74\times)

Table 10. **Ablation on group count J on the NeRF-DS [52] and D-NeRF [39] datasets with our SpeeDe3DGS framework.** $J=-$ indicates that GroupFlow is not used. Each experiment is repeated three times and averaged to reduce training variance. The **best** and **second best** results are highlighted. FPS and Train Time are measured on RTX 3090 and RTX A5000 GPUs, respectively.

Prune	J	NeRF-DS [52]							D-NeRF [39]						
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Size (MB) \downarrow	# Gaussians \downarrow	Train Time (s) \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FPS \uparrow	Size (MB) \downarrow	# Gaussians \downarrow	Train Time (s) \downarrow
-	-	23.80	0.8503	0.1781	54.37	33.21	132.22K	1523.83	38.92	0.9892	0.0143	127.47	16.88	62.92K	940.52
128	-	22.67	0.8181	0.2322	415.53	36.92	146.43K	905.16	35.91	0.9844	0.0185	376.31	16.22	63.13K	651.65
256	-	23.04	0.8322	0.2103	407.14	37.61	144.46K	916.39	36.28	0.9851	0.0179	373.42	16.97	62.92K	643.34
512	-	22.86	0.8238	0.2232	417.04	38.43	138.27K	855.18	36.68	0.9860	0.0170	375.79	18.66	63.27K	640.09
1024	-	23.48	0.8421	0.1919	406.31	41.75	132.73K	827.25	36.75	0.9861	0.0170	373.09	21.81	63.02K	639.20
2048	-	23.54	0.8433	0.1892	406.21	51.00	132.32K	826.75	36.85	0.9862	0.0172	374.32	28.17	62.77K	636.39
4096	-	22.64	0.7995	0.2195	416.92	69.89	133.14K	879.70	36.96	0.9861	0.0177	376.61	40.99	62.71K	645.95
✓	-	23.81	0.8515	0.1853	345.24	4.55	11.06K	750.69	36.19	0.9792	0.0350	423.89	3.08	4.57K	524.34
✓	128	23.29	0.8402	0.2012	514.64	3.86	11.01K	635.63	34.64	0.9767	0.0363	519.12	1.92	4.56K	496.64
✓	256	23.41	0.8440	0.1958	516.81	5.05	11.09K	665.38	34.96	0.9773	0.0358	515.71	2.73	4.58K	490.41
✓	512	23.55	0.8457	0.1930	514.06	7.40	11.14K	643.96	35.07	0.9775	0.0354	514.24	4.34	4.61K	488.84
✓	1024	23.64	0.8476	0.1903	511.18	12.07	11.17K	635.00	35.11	0.9776	0.0360	520.66	7.54	4.56K	485.39
✓	2048	23.66	0.8487	0.1901	505.60	21.40	11.10K	625.48	35.07	0.9771	0.0365	524.19	13.96	4.56K	486.10
✓	4096	23.03	0.8316	0.2230	504.66	39.96	10.56K	698.64	35.08	0.9769	0.0373	524.41	26.80	4.57K	492.66

Table 11. **Ablation on perturbation magnitude β and annealing period τ on the NeRF-DS [52] dataset with our SpeeDe3DGS framework.** $\beta=-$ and $\tau=-$ indicate that TSS is not used. Each experiment is repeated three times and averaged to reduce training variance.

Method	β	τ	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
TSP	-	-	23.78	0.8507	0.1863
TSP + TSS	0.05	20K	23.82	0.8513	0.1861
TSP + TSS	0.1	15K	23.82	0.8510	0.1858
TSP + TSS	0.1	20K	23.81	0.8515	0.1853
TSP + TSS	0.1	25K	23.80	0.8517	0.1853
TSP + TSS	0.2	20K	23.81	0.8514	0.1858

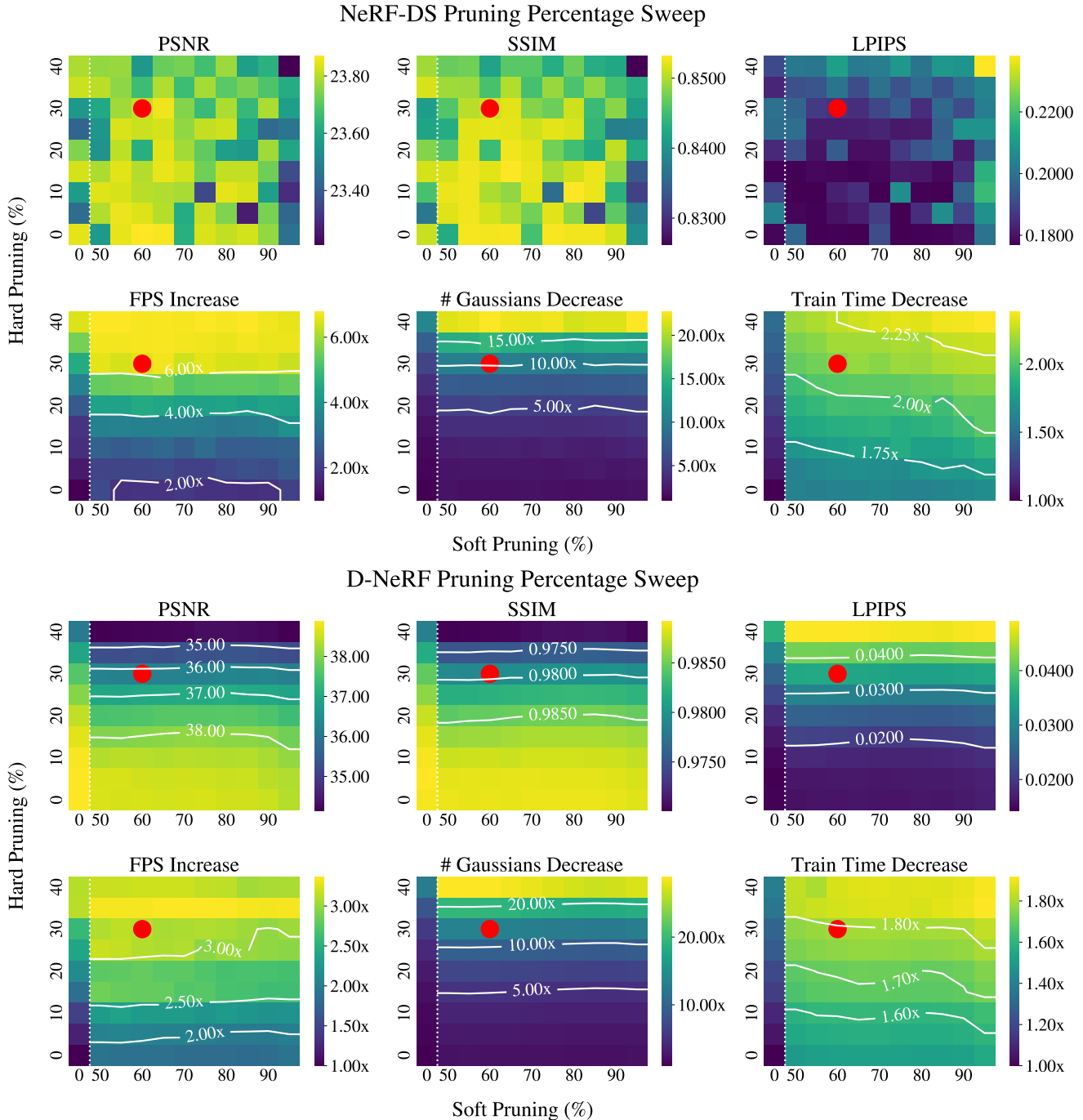


Figure 5. **Ablation on pruning percentages with our Speede3DGS framework.** We sweep soft (densification-stage) and hard (post-densification) pruning ratios in 5% increments for the NeRF-DS [52] and D-NeRF [39] datasets using the DeformableGS [53] codebase. Each configuration is run three times without TSS or GroupFlow, and results are averaged across all runs. (0%, 0%) corresponds to the unpruned baseline, while the first row and column show pruning in isolation. The red dot marks our selected (60%, 30%) soft-hard ratio. FPS and Train Time improvements are measured on an RTX A5000 GPU.

16 provide per-scene rendering FPS and training time in seconds for NeRF-DS, D-NeRF, and the eight HyperNeRF [37] scenes used in DeformableGS [53]. As in Appendix A.2, ren-

dering FPS is measured on an RTX 3090 GPU, while training time is measured on an RTX A5000 GPU for NeRF-DS and D-NeRF and on an RTX A6000 GPU for HyperNeRF.

Table 12. **PSNR \uparrow on each scene with our SpeeDe3DGS framework.** *TSP*, *TSS*, and *GF* denote Temporal Sensitivity Pruning, Sampling, and GroupFlow, respectively. Each experiment is run three times and averaged to reduce training variance. The **best** and **second best** results are highlighted.

TSP	TSS	GF	NeRF-DS [52]							D-NeRF [39]							
			<i>as</i>	<i>basin</i>	<i>bell</i>	<i>cup</i>	<i>plate</i>	<i>press</i>	<i>sieve</i>	<i>balls</i>	<i>warrior</i>	<i>hook</i>	<i>jacks</i>	<i>lego</i>	<i>mutant</i>	<i>standup</i>	<i>trex</i>
DeformableGS [53]			26.09	19.69	25.00	24.61	20.36	25.32	25.12	40.92	41.22	36.68	37.40	32.52	41.90	43.44	37.25
✓			26.15	19.54	25.20	24.56	20.39	25.49	25.08	38.24	39.47	34.18	35.54	30.58	41.02	39.08	35.44
✓	✓		26.10	19.67	25.46	24.63	20.36	25.47	24.87	38.32	39.44	34.16	35.43	30.47	36.96	39.11	35.62
		✓	25.78	19.44	25.06	24.04	20.22	25.05	25.21	38.44	40.10	35.12	35.06	31.31	39.81	40.03	34.97
✓	✓	✓	26.05	19.47	25.22	24.30	20.22	25.09	25.25	36.68	38.86	33.24	34.04	30.05	36.32	37.51	33.85

Table 13. **SSIM \uparrow on each scene with our SpeeDe3DGS framework.**

TSP	TSS	GF	NeRF-DS [52]							D-NeRF [39]							
			<i>as</i>	<i>basin</i>	<i>bell</i>	<i>cup</i>	<i>plate</i>	<i>press</i>	<i>sieve</i>	<i>balls</i>	<i>warrior</i>	<i>hook</i>	<i>jacks</i>	<i>lego</i>	<i>mutant</i>	<i>standup</i>	<i>trex</i>
DeformableGS [53]			0.8781	0.7935	0.8424	0.8895	0.8100	0.8616	0.8712	0.9955	0.9866	0.9850	0.9892	0.9766	0.9941	0.9939	0.9925
✓			0.1910	0.7912	0.8436	0.8881	0.8126	0.8636	0.8694	0.9925	0.9792	0.9721	0.9819	0.9598	0.9783	0.9849	0.9856
✓	✓		0.8838	0.7947	0.8498	0.8886	0.8116	0.8622	0.8676	0.9925	0.9793	0.9719	0.9817	0.9591	0.9781	0.9849	0.9858
		✓	0.8748	0.7837	0.8411	0.8790	0.8022	0.8562	0.8661	0.9932	0.9835	0.9813	0.9856	0.9727	0.9924	0.9907	0.9900
✓	✓	✓	0.8799	0.7890	0.8466	0.8864	0.8080	0.8620	0.8688	0.9905	0.9771	0.9695	0.9790	0.9567	0.9777	0.9828	0.9836

Table 14. **LPIPS \downarrow on each scene with our SpeeDe3DGS framework.**

TSP	TSS	GF	NeRF-DS [52]							D-NeRF [39]							
			<i>as</i>	<i>basin</i>	<i>bell</i>	<i>cup</i>	<i>plate</i>	<i>press</i>	<i>sieve</i>	<i>balls</i>	<i>warrior</i>	<i>hook</i>	<i>jacks</i>	<i>lego</i>	<i>mutant</i>	<i>standup</i>	<i>trex</i>
DeformableGS [53]			0.1861	0.1856	0.1611	0.1553	0.2231	0.1914	0.1475	0.0088	0.0262	0.0174	0.0144	0.0212	0.0069	0.0088	0.0107
✓			0.1914	0.1974	0.1770	0.1602	0.2299	0.2004	0.1582	0.0177	0.0525	0.0436	0.0291	0.0481	0.0360	0.0269	0.0254
✓	✓		0.1922	0.1957	0.1628	0.1599	0.2293	0.1996	0.1596	0.0178	0.0523	0.0441	0.0292	0.0489	0.0361	0.0269	0.0251
		✓	0.1961	0.2040	0.1654	0.1695	0.2337	0.1996	0.1564	0.0127	0.0309	0.0204	0.0172	0.0232	0.0088	0.0118	0.0126
✓	✓	✓	0.1971	0.2052	0.1653	0.1646	0.2356	0.2020	0.1610	0.0216	0.0551	0.0446	0.0312	0.0491	0.0360	0.0285	0.0258

Table 15. **FPS \uparrow on each scene with our SpeeDe3DGS framework.** Metrics are collected on an RTX 3090 GPU for consistency with MonoDyGauBench [25].

TSP	TSS	GF	NeRF-DS [52]							D-NeRF [39]							HyperNeRF [37]								
			<i>as</i>	<i>basin</i>	<i>bell</i>	<i>cup</i>	<i>plate</i>	<i>press</i>	<i>sieve</i>	<i>balls</i>	<i>warrior</i>	<i>hook</i>	<i>jacks</i>	<i>lego</i>	<i>mutant</i>	<i>standup</i>	<i>trex</i>	<i>americano</i>	<i>chicken</i>	<i>cookie</i>	<i>espresso</i>	<i>hand</i>	<i>lemon</i>	<i>printer</i>	<i>torchocolate</i>
DeformableGS [53]			51.64	40.32	34.82	49.22	48.58	50.65	51.62	85.36	277.08	98.70	182.66	45.71	80.71	185.68	65.87	8.89	13.24	11.34	25.39	5.37	17.13	21.40	13.07
✓			426.84	346.75	280.81	377.27	388.79	331.68	421.80	427.86	480.22	392.53	441.17	355.65	393.94	440.04	450.47	123.26	135.82	121.11	199.73	60.66	159.31	193.08	135.90
✓	✓		423.39	330.80	277.22	421.34	416.87	328.87	379.39	429.47	480.19	391.86	444.19	359.19	393.63	438.20	428.41	114.45	130.83	110.68	188.44	59.33	155.84	190.59	126.97
		✓	418.65	387.38	394.63	413.78	424.97	421.47	382.58	330.51	489.38	378.30	417.95	300.25	328.20	428.41	321.76	167.79	245.01	177.56	320.61	129.54	217.72	316.91	239.04
✓	✓	✓	532.29	510.42	499.90	510.30	517.44	500.52	478.33	484.77	546.52	527.00	548.85	492.38	511.32	546.11	536.37	408.62	438.86	405.26	427.20	408.48	401.04	451.23	442.34
			(1.01x)	(2.18x)	(2.05x)	(1.03x)	(1.06x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)	(1.01x)

Table 16. **Training time \downarrow in seconds on each scene with our SpeeDe3DGS framework.** Metrics are collected with an RTX A5000 GPU for NeRF-DS [52] and D-NeRF [39] and an RTX A6000 GPU for HyperNeRF [37]; see Appendix A.2 for more details.

TSP	TSS	GF	NeRF-DS [52]							D-NeRF [39]							HyperNeRF [37]								
			<i>as</i>	<i>basin</i>	<i>bell</i>	<i>cup</i>	<i>plate</i>	<i>press</i>	<i>sieve</i>	<i>balls</i>	<i>warrior</i>	<i>hook</i>	<i>jacks</i>	<i>lego</i>	<i>mutant</i>	<i>standup</i>	<i>trex</i>	<i>americano</i>	<i>chicken</i>	<i>cookie</i>	<i>espresso</i>	<i>hand</i>	<i>lemon</i>	<i>printer</i>	<i>torchocolate</i>
DeformableGS [53]			1407.08	1692.53	1924.65	1467.75	1378.45	1404.10	1386.70	1090.20	490.14	888.91	604.36	1604.89	1021.78	580.32	1243.60	6746.79	4138.62	6145.04	3156.94	10118.35	4482.59	2660.25	4542.70
✓			723.34	776.01	940.58	748.42	626.04	683.67	714.11	579.25	381.99	523.37	412.22	741.95	540.69	401.63	615.06	2426.49	1432.92	2216.52	1383.00	3867.42	1744.81	1040.99	1826.96
✓	✓		709.59	801.15	961.06	763.10	616.79	716.46	707.28	575.82	383.70	521.08	411.99	746.84	540.70	400.76	613.80	2409.60	1441.43	2185.75	1368.90	3646.07	1799.51	1072.80	1825.71
		✓	758.47	843.14	1011.61	824.64	729.46	800.62	819.34	798.95	403.14	577.25	498.40	936.23	660.42	445.82	770.69	3283.08	1794.64	2971.33	1644.66	4539.15	2179.02	1252.77	2378.86
✓	✓	✓	600.07	640.83	754.56	638.72	531.30	592.14	620.75	590.77	371.01	449.02	431.48	625.99	482.93	381.81	555.79	1756.25	976.64	1612.27	1119.53	2300.66	1321.41	810.68	1317.33
			(2.34x)	(2.64x)	(2.55x)	(2.3x)	(2.59x)	(2.37x)	(2.23x)	(1.85x)	(1.32x)	(1.98x)	(1.4x)	(2.56x)	(2.12x)	(1.52x)	(2.24x)	(3.84x)	(4.24x)	(3.81x)	(2.82x)	(4.40x)	(3.39x)	(3.28x)	(3.45x)



Figure 6. **Visual comparison of the baseline DeformableGS [53] and our Speede3DGS methods on MonoDyGauBench (MDGB) [25].** The baseline examples are reproduced through retraining, as the original MonoDyGauBench models are not available. For all other visualizations, we use our standalone Speede3DGS codebase, which produces consistently higher FPS and image quality than the MDGB wrapper. Top: *slice-banana* from HyperNeRF [37]. Middle: *curls* from Nerfies [36]. Bottom: *creeper* from iPhone [10].

A.7. Qualitative Results on MonoDyGauBench

Figure 6 presents qualitative comparisons between the DeformableGS [53] baseline and our Speede3DGS methods, using representative scenes from HyperNeRF [37], Nerfies [36], and iPhone [10] within the MonoDyGauBench framework [25]. Pruning (TSP+TSS) and GroupFlow significantly accelerate rendering while maintaining comparable or better visual fidelity than the baseline. As discussed in Appendix A.1, grouped SE(3) motion distillation also acts as a regularizer in real-world scenes – enhancing temporal coherence, mitigating drift under noisy or unstable camera trajectories, and improving overall image quality. This effect is especially pronounced on the *creeper* scene from iPhone, where the SSIM of the Pruning+GroupFlow model is twice that of both the baseline and Pruning-only variants. Note that our standalone Speede3DGS codebase achieves superior rendering speed and visual fidelity compared to the standardized MonoDyGauBench framework.

A.8. Baseline and Motion Model Limitations

Our methods inherit the limitations of the dynamic Gaussian Splatting frameworks into which they are integrated. In particular, highly deformable scenes remain challenging for current approaches. MonoDyGauBench [25] finds that the baseline methods struggle with non-rigid motion and noisy camera poses, so our approach faces similar limitations.

Nevertheless, Sections 5.2 and Appendices A.1 and A.3 demonstrate that GroupFlow can introduce regularizing effects that improve temporal stability in such scenarios. In practice, globally non-rigid scenes consist of clusters of small Gaussians whose motion is locally rigid due to cloning and splitting during densification. Increasing the number of groups J enables progressively finer motion modeling; in the extreme case $J=N$, GroupFlow distills per-Gaussian neural motion into efficient SE(3) transformations. Adaptive strategies that refine groups based on motion variation could further improve modeling of complex non-rigid motion.