
(Supplementary Materials)

DGS-LRM: Real-Time Deformable 3D Gaussian Reconstruction From Monocular Videos

1 We provide video results reconstructed with our DGS-LRM on DAVIS and DyCheck. Note that, due
2 to institutional policies, we do not provide visual samples that contain human identities.

3 1 DyCheck

4 **Comparison.** Both D3DGS and PGDVS create unrealistic deformations. In addition, D3DGS often
5 has significant errors in the scene scale, while PGDVS has severe warping artifacts and discontinuity
6 in motion. Our DGS-LRM better recovers the scene scale when the input video has sufficient
7 movements, and reconstructs better and intuitive deformations.

8 Column Definitions.

- 9 • We compare with D3DGS and PGDVS.
- 10 • The first column is the input video.
- 11 • The second column is the ground-truth video in the novel view.
- 12 • The third column is our DGS-LRM output in the novel view.
- 13 • The fourth column is D3DGS results in the novel view.
- 14 • The fifth column is PGDVS results in the novel view.

15 **Improper Covisibility.** DyCheck uses the entire input video sequence to label the covisibility
16 in the novel view, which annotates the regions that are visible in any frame in the input video.
17 Using such a covisibility mask for evaluation is disadvantageous for our DGS-LRM, since we only
18 reconstruct regions based on the input clip (in the left-most column of the video). In comparison,
19 both D3DGS and PGDVS use the entire video sequence for reconstruction. Consequently, they have
20 more information than DGS-LRM, leading to better reconstructions in these regions. Here are some
21 example regions that are not visible in the input clip:

- 22 • In the *Apple* sequence, the right leg of the person.
- 23 • In the *Teddy* sequence, the regions between the human and the teddy bear.
- 24 • In the *Paperwindmill* sequence, the bottom-right side of the pot
- 25 • In the *Brick* sequence, the left-most Lego brick.

26 Additional Notes.

- 27 • We do not include L4GM since the visibility is significantly misaligned.
- 28 • Frames without ground-truth image and covisibility mask are blacked out.

29 2 DAVIS

30 Column Definitions.

- 31 • The first column shows the input video.
- 32 • In the second column, we show the view synthesis rendering (first row) and 3D scene flow
33 (second row) results in the input camera poses.

34 • In the last column, we show the results in novel views with spiral camera motion. We repeat the
35 same motion twice. The spiral frequency is different from the z-direction oscillation frequency,
36 so that the two repetitive sequences have different view coverages.

37 **Scene Flow Visualization.** For 3D scene flow visualization, we first downsample the per-pixel
38 scene flow for visual clarity. Additionally, due to the motion of dynamic objects, certain pixels will
39 correspond to several tracking points. For visual clarity, we exclude the stationary tracking points in
40 the background. This is done by applying the object mask provided by DAVIS.

41 **Reconstructing Only The Visible Area.** DGS-LRM is dedicated to the reconstruction task. There-
42 fore, the invisible area is expected to remain unreconstructed. The two types of invisible areas:

- 43 • For the static geometry, the regions that are never visible to the camera throughout the video.
- 44 • For the dynamic geometry, it further includes the dynamic parts that have moved outside the
45 camera frustum at the current timestamp.