

Figure 1: An updated illustration of multiple correct 3D scene configurations from a dynamic monocular video

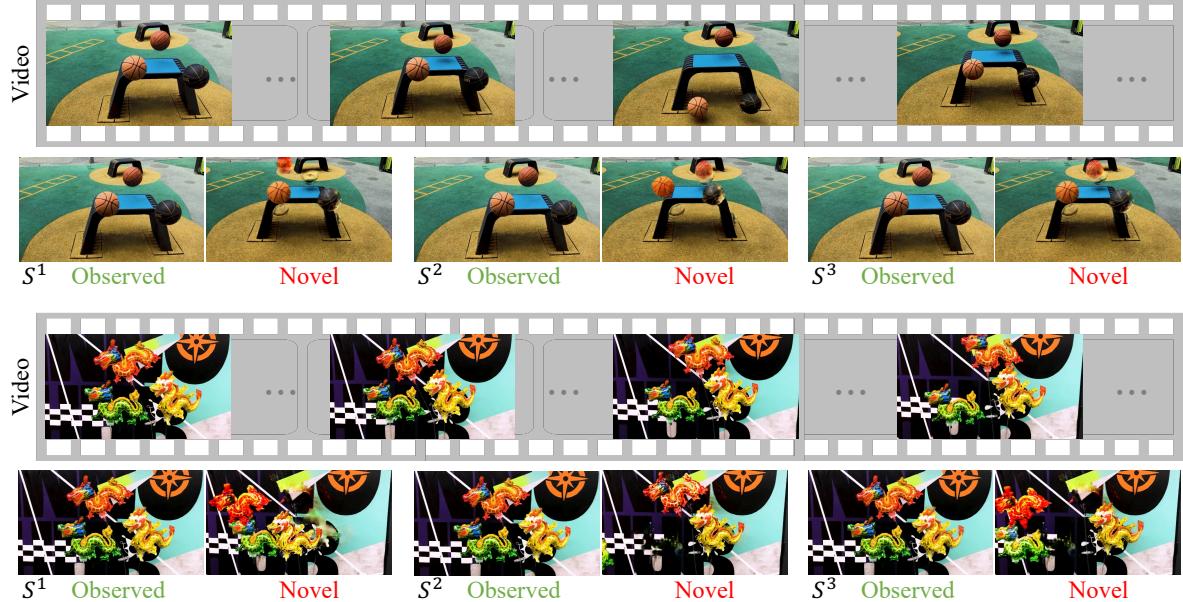


Figure 2: Qualitative results on two daily scenes captured by a mobile phone. For each scene we show the captured monocular video, and the rendering results from observed and novel views under 3 sampled scale combinations at a specific time step.

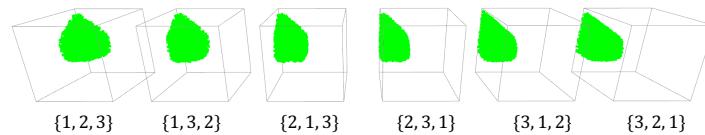


Figure 3: The distribution of valid scale combinations learned for "Gnome House" scene, with different permutation of object orders. Each 3D bounding box represents the normalized range for scale combinations of 3 objects, and the valid scale combinations out of the 100K sampled ones are marked as green. Object orders of different permutations are aligned for better visualization

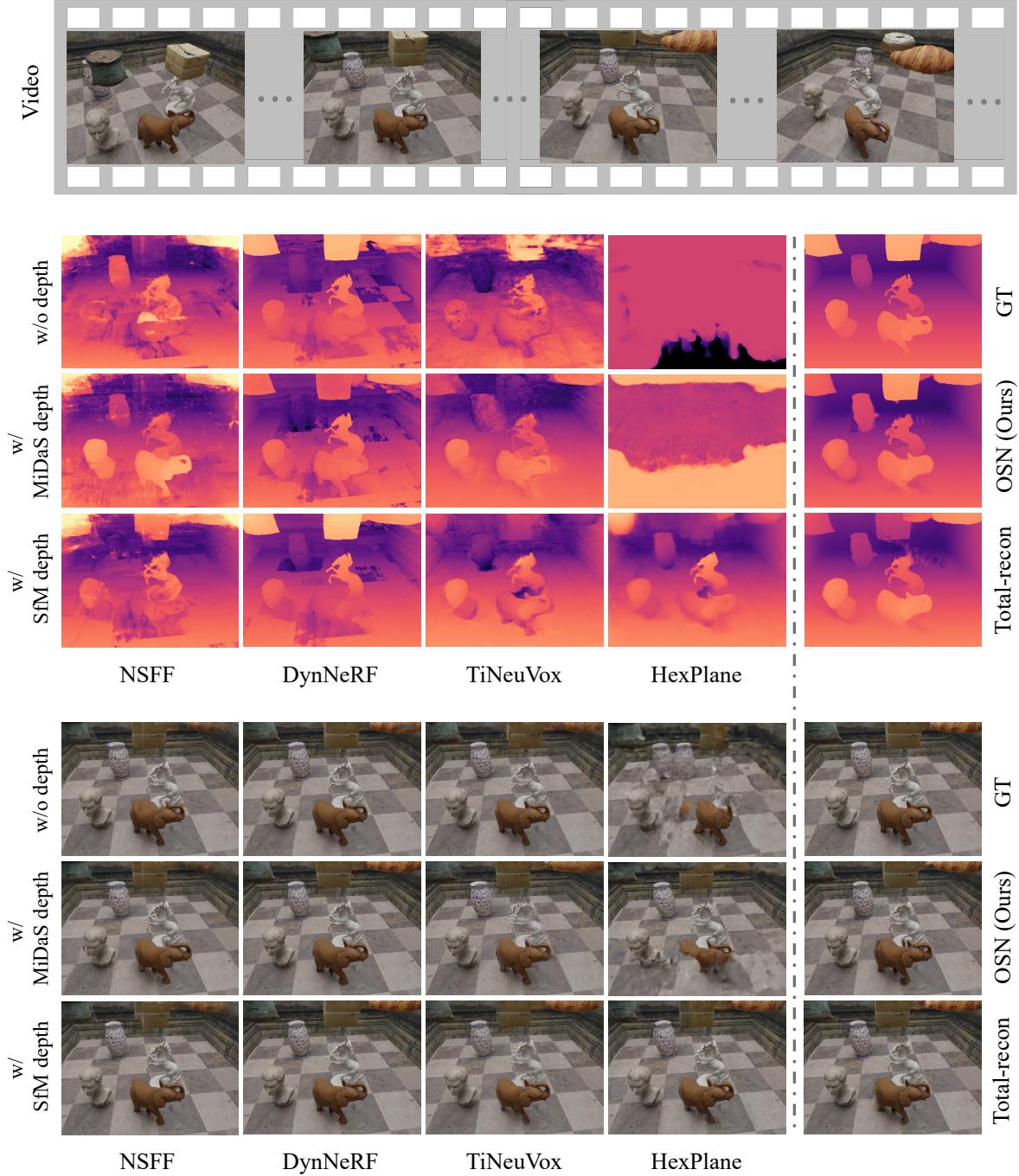


Figure 4: Qualitative results on new synthetic "Chessboard++" scene with appearing and disappearing objects.

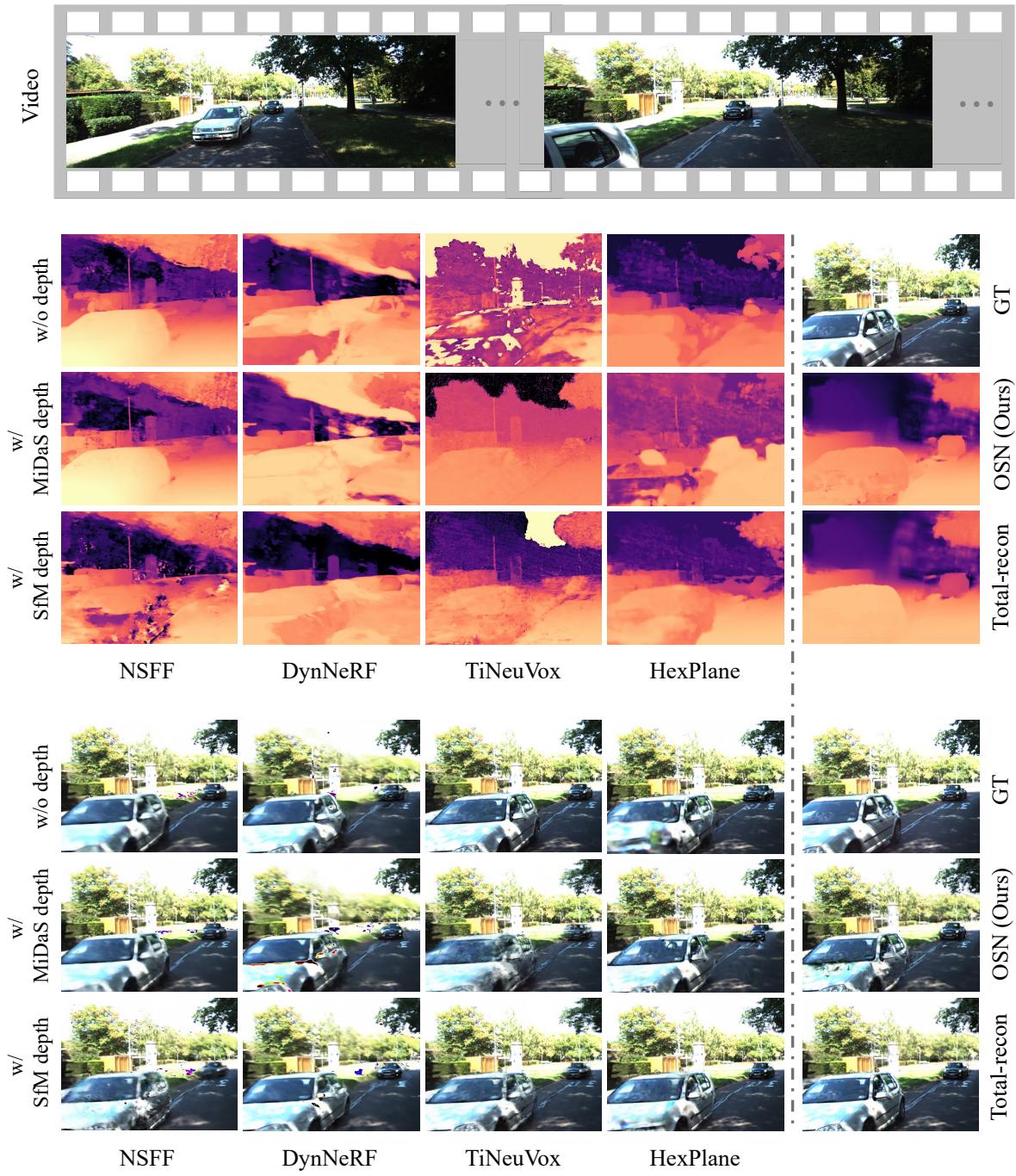


Figure 5: Qualitative results on selected scene from Sequence 0007 of KITTI dataset.

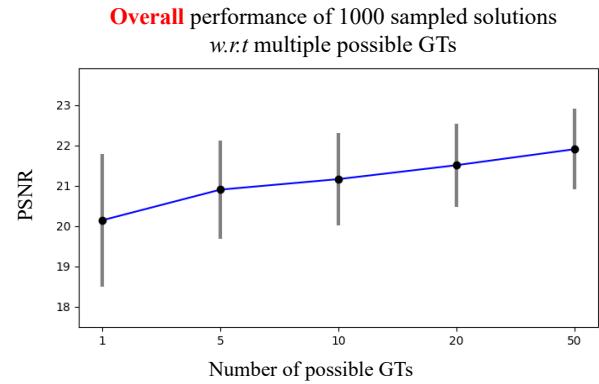


Figure 6: Overall performance of 1000 sampled solutions *w.r.t* different number of possible GTs.

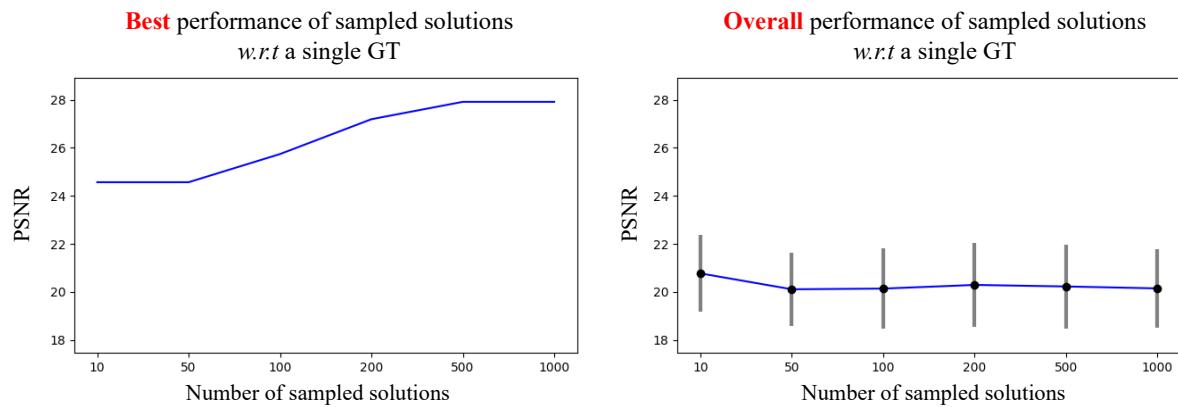


Figure 7: Best and overall performance of different number of sampled solutions *w.r.t* a single GT.