

FlowMap: High-Quality Camera Poses, Intrinsics, and Depth via Gradient Descent

—Supplemental Material—

Anonymous 3DV submission

Paper ID 265

1. Additional Discussions

1.1. Limitations.

FlowMap has several limitations that suggest exciting directions for future work. First, FlowMap requires off-the-shelf correspondences from optical flow and point tracking methods. An exciting direction is to remove the dependence on correspondence altogether or to jointly learn correspondence extraction. Second, we mainly analyze FlowMap in the setting of per-scene optimization, where our results demonstrate that the gradients provided by FlowMap’s formulation are robustly lead to high-quality depth and camera parameters. It is natural to attempt to use these gradients to train a feed-forward structure-from-motion method. Lastly, through its dependence on optical flow or point tracks, FlowMap can currently only process continuous video, in contrast to conventional SfM methods which can operate on unstructured image collections. This is a natural assumption for applications in embodied intelligence, navigation, and robotics, but limits applications in computer graphics. Leveraging unstructured correspondences, e.g. via [3], may be used to overcome this limitation.

1.2. Relationship to Conventional SfM.

Across many applications of conventional SfM, such as the reconstruction of large, non-continuous image collections, FlowMap cannot serve as a drop-in replacement, and we note that this is not our objective. Rather, we demonstrate that a self-supervised, end-to-end differentiable, feed-forward formulation that can naturally be integrated into neural network vision models surprisingly approaches COLMAP’s performance on the downstream task of novel view synthesis *in the context of video data*. Here, FlowMap has the potential to make camera pose and depth supervision unnecessary for 3D deep learning, paving the way for training on unannotated, internet-scale video data.

1.3. Memory and Time Requirements.

FlowMap’s complexity in time and memory is linear with the number of input video frames. During each optimization step, FlowMap recomputes depth for each frame, then derives poses and intrinsics from these depths to generate gradients. In practice, FlowMap optimization for a 150-frame video takes about 20 minutes, with a peak memory usage of about 36 GB. Precomputing point tracks and optical flow takes approximately 2 minutes. Note that FlowMap’s runtime could be reduced by early stopping, and its memory usage could be reduced by performing backpropagation on video subsets during each step, but we leave these optimizations to future work.

1.4. Sequence Length and Drift.

Since adjacent frames in typical 30 FPS videos usually contain redundant information, we run FlowMap on subsampled videos. We perform subsampling by computing optical flow on the whole video, then selecting frames so as to distribute the overall optical flow between them as evenly as possible. With this strategy, we find that an object-centric, full 360° trajectory as is common in novel view synthesis papers is covered by about 90 frames. We note that FlowMap does not have a loop closure mechanism. Rather, point tracks provide long-range correspondences that prevent the accumulation of drift in long sequences.

2. Additional Ablation Studies

In Tab. 1, we include three additional ablations. The “Random Init.” ablation uses a randomly initialized CNN for FlowMap training with 2,000 steps of optimization. The “Random Init. (20k)” ablation is identical, but runs for 20,000 optimization steps. The “No Corresp. Weights” ablation removes the correspondence weights used in FlowMap’s Procrustes-solving step. We note that the “Random Init. (20k)” ablation’s performance almost matches FlowMap’s, indicating that although pre-training

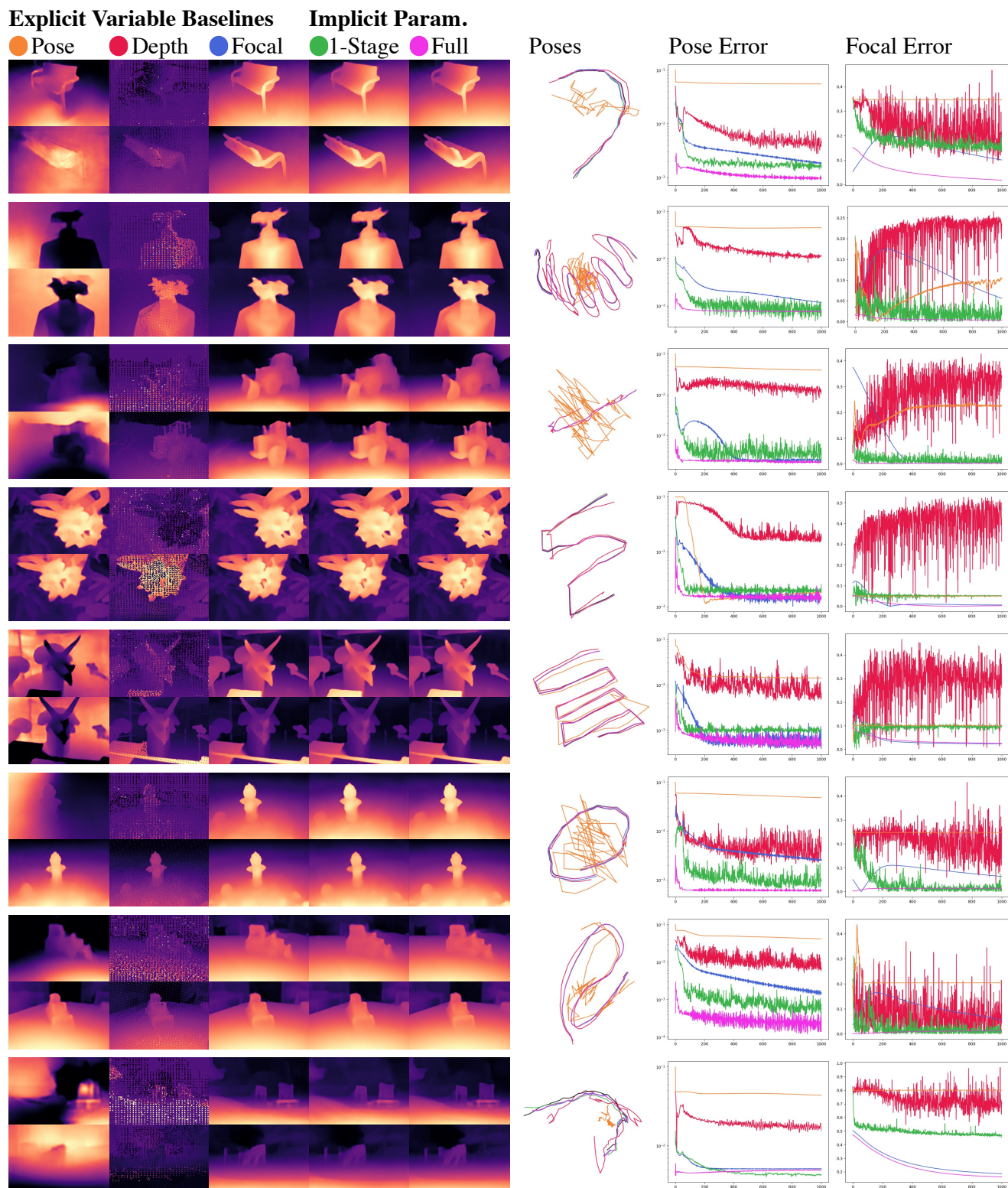


Figure 1. **Pose and Geometry Convergence for Free-Variable vs. Proposed Parameterizations.** We plot poses, depths, focal lengths, and pose error (ATE) obtained with our proposed parameterizations (“Full”) vs. those obtained with free-variable parameterizations at various optimization steps. With our proposed reparameterizations (“Full”) as a baseline, we ablate either depth, focal length, or poses as free-variable optimizations and plot the resulting optimizations’ pose and depth estimates. For instance, “Depth” corresponds to making the depth an explicit free-variable in the optimization. Using pose-as-variable and depth-as-variable often lead to “hollow-face” geometry, where the geometry is effectively inverted but still mostly satisfies the optical flow constraints. We also show results from a single-stage FlowMap pipeline, which only uses the implicit parameterization of intrinsics rather than switching to regressed intrinsics halfway through optimization. Note that the plotted lines for “Full” are initialized with the results of “1-Stage” and represent the second stage (explicit focal length) of FlowMap optimization.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
FlowMap	27.70	0.863	0.089
Single Stage	26.66	0.842	0.112
Expl. Focal Length	25.15	0.788	0.141
Expl. Depth	8.84	0.168	0.684
Expl. Pose	16.00	0.533	0.495
No Tracks	25.83	0.822	0.122
Random Init.	25.54	0.808	0.129
Random Init. (20k)	27.25	0.850	0.101
No Corresp. Weights	24.18	0.765	0.168

Table 1. **Additional Ablations.** We report additional ablation results averaged across all scenes alongside the ablations found in the main paper.

helps FlowMap converge much more quickly, it is not necessary for accuracy. In Tab. 2, we report per-scene ablation results. Finally, Fig. 1 compares convergence between FlowMap and the free-variable parameterization variants on more scenes.

3. Additional Results

3.1. Pre-Trained Depth vs. Fine-Tuned Depth vs. High-Resolution Fine-Tuned Depth.

In Fig. 4, we compare the depths produced by FlowMap’s initialization to the depths produced after FlowMap optimization. We additionally compare these results to a MiDaS CNN fine-tuned at a significantly higher resolution. We find that per-scene fine-tuning leads to high-quality depth predictions. This is illustrated by Fig. 3, which demonstrates FlowMap’s ability to generate high-quality, consistent depths. However, it is worth noting that FlowMap’s off-the-shelf depths are slightly blurry. To investigate whether this is a limitation of our loss or the architecture of the depth-predicting CNN, we also perform optimization at a higher resolution. We find that this leads to crisp depth maps, demonstrating that blurry depth maps are a result of insufficient capacity of the MiDaS backbone and not a limitation of our camera-induced flow loss. Notably, the poses barely change in this fine-tuning stage. It is likely that replacing the MiDaS depth predictor with a more powerful depth backbone would lead to sharper depth without high-resolution fine-tuning.

3.2. Effects of Pretraining Expanded Figure

We include an expanded figure with an additional ATE plot for the pretrained vs from-scratch optimization study in Fig. 6.

3.3. Large-Scale Robustness Study

We study FlowMap’s robustness by using it to estimate camera poses for 420 CO3D scenes from 10 categories. We compare these trajectories to CO3D’s pose annotations, which were computed using COLMAP. Since the quality of CO3D’s ground-truth trajectories varies between categories, we focus on categories that have been used to train novel view synthesis models [1, 5, 6], where pose accuracy is expected to be higher. We find that FlowMap’s mean ATE (0.0056) is lower than DROID-SLAM’s (0.0082) and similar to the mean ATE obtained by re-running COLMAP and comparing the results to the provided poses (0.0038). This demonstrates that FlowMap consistently estimates poses which are close to COLMAP’s. We note that COLMAP failed to estimate poses for 36 scenes, possibly because we ran it at a sparser frame rate to be consistent with our method or because the original annotations were generated using different COLMAP settings; we exclude COLMAP’s failures from the above mean ATE. See Fig. 7 for distributions of ATE values with respect to CO3D’s provided camera poses.

3.4. Additional Point Clouds and Qualitative Pose Reconstructions

In Fig. 3, we display 12 additional point clouds plus estimated camera poses across popular datasets and scenes across the LLFF, Tanks and Temples, MipNeRF 360, and CO3D datasets. FlowMap robustly recovers camera poses and scene geometry across these diverse, challenging, and real-world sequences.

3.5. Failure Cases

While running FlowMap, we observed failures on several scenes. These include the Tanks-and-Temples Auditorium scene (our model struggles with rotation-dominant trajectories), the LLFF Leaves scene (our model falls into

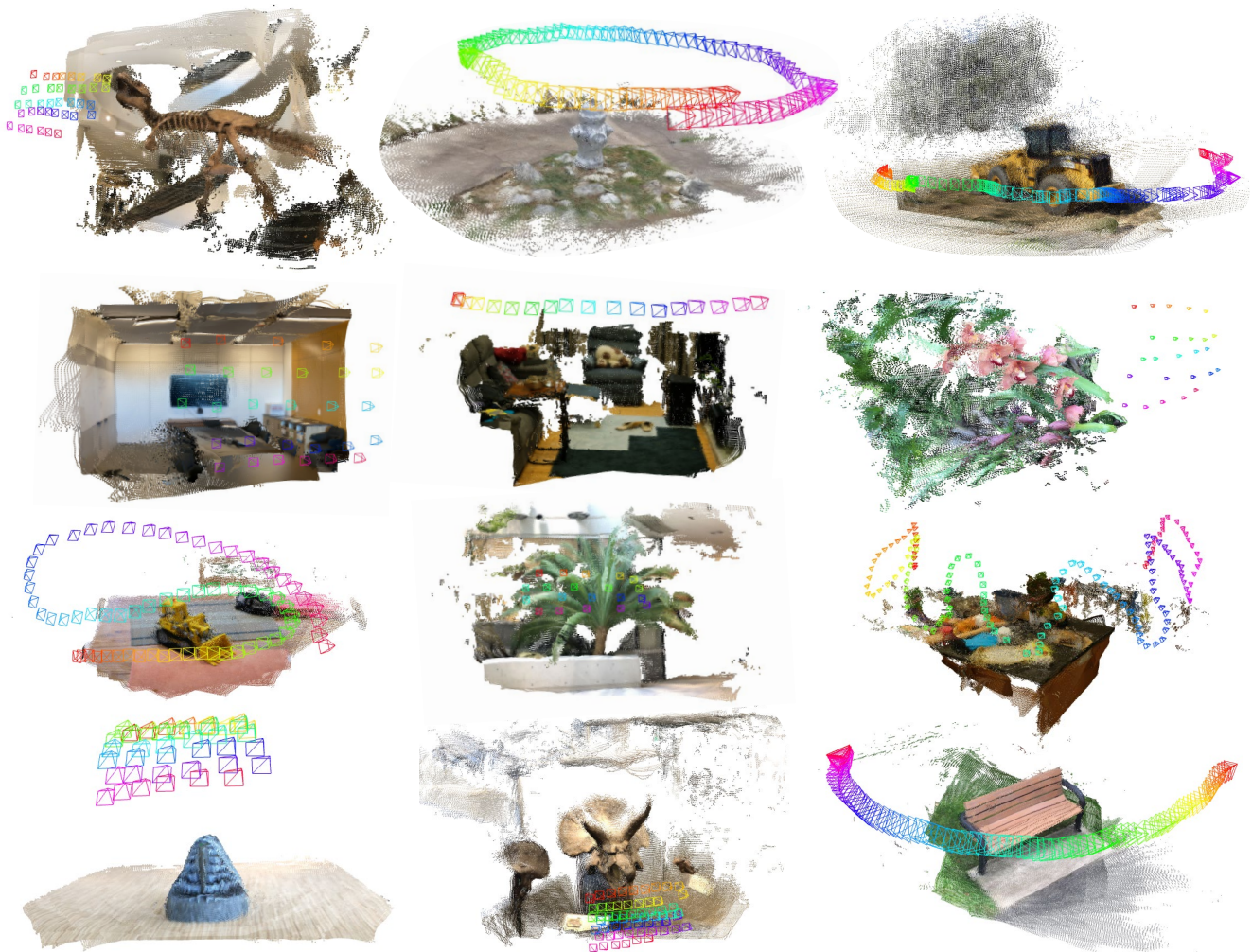


Figure 3. **Additional Point Clouds** Here we plot additional point clouds across the Tanks and Temples, LLFF, Mip-NeRF360, and CO3D datasets.

a “hollow-face minimum”), and the Tanks-and-Temples Lighthouse scene (this video features a large lens flare which degrades the optical flow). Future extensions to FlowMap could use an occlusion-aware formulation to avoid hollow-face minima.

4. Implementation Details

4.1. Procrustes Solver Details

Our pose solver is the one introduced in FlowCam [4]; see [4] for details. The only difference is that instead of selecting 1000 random points for the Procrustes estimation, we fix the points (uniformly spaced throughout the image) when performing per-scene overfitting. We find that fixing the points used for the pose solver allows the network to better overfit confidence weights and subsequently yields better poses.

4.2. Intrinsic Solver Details

For the intrinsic solver, we assume a pinhole camera estimate and discretize a set of 60 candidate focal lengths between .5 and 2 (in resolution-independent units). We use a softmin on the flow error maps, as discussed in the main paper. We scale the error maps by a temperature factor of 10 and weight the error maps by the flow confidence weights. See Fig. 8 for illustration.

4.3. Depth NN (MiDaS) details

For our depth network, we use the lightweight CNN version of MiDaS [2], pretrained with the publicly available weights trained on relative-depth estimation. We optimize the entire network weights during training.

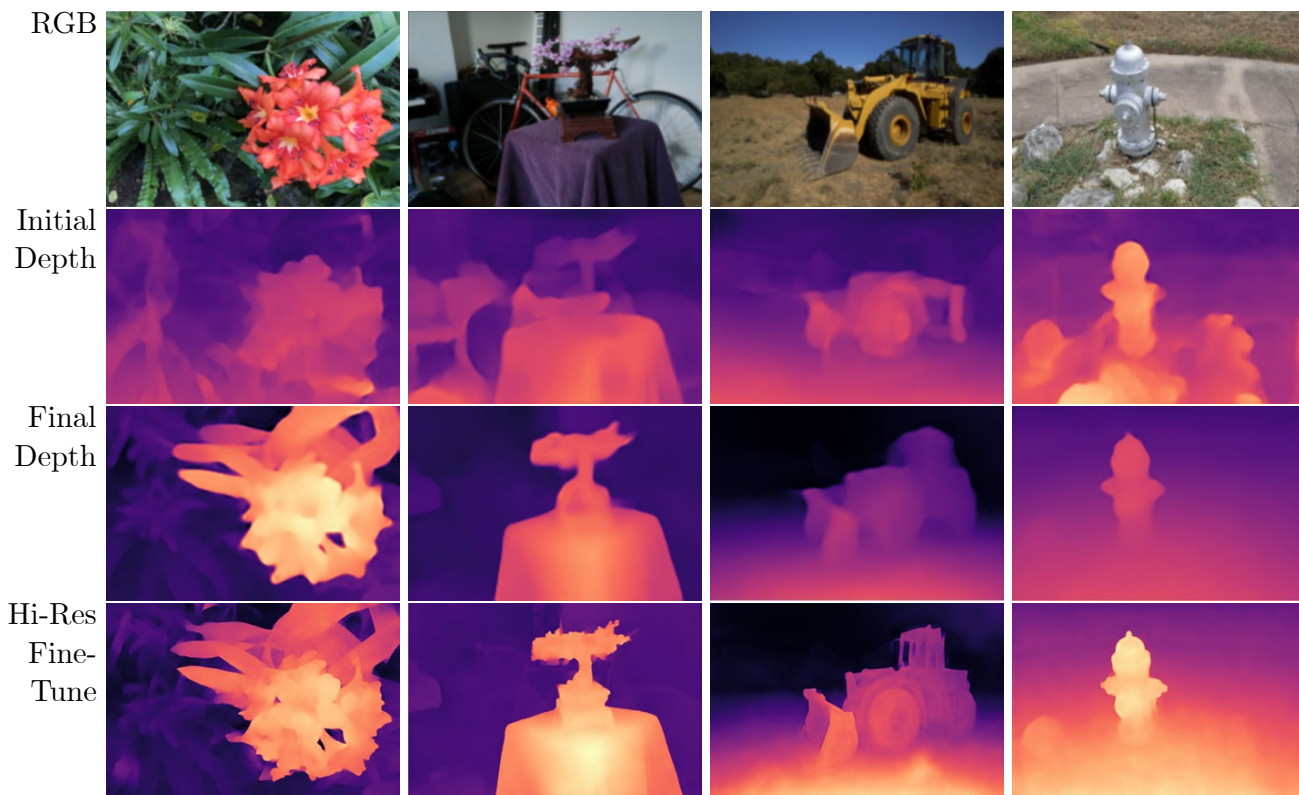


Figure 4. **Depth Estimates Before and After Optimization.** The depth prediction neural network can either be randomly initialized or pre-trained, though pre-trained depth networks lead to much faster convergence. In the second row, we show the output of the depth prediction neural network after pre-training it on a dataset consisting of CO3D, KITTI, and RealEstate10k. These estimates converge to high-quality depth within only a few hundred FlowMap optimization steps. We see that the quality of the initial, pre-trained depth predictions is not critical to achieve accurate reconstructions. Although we estimate geometry at a lower resolution during optimization to manage memory constraints, we can quickly fine-tune at high-resolution for more detailed depth maps if necessary (bottom row).

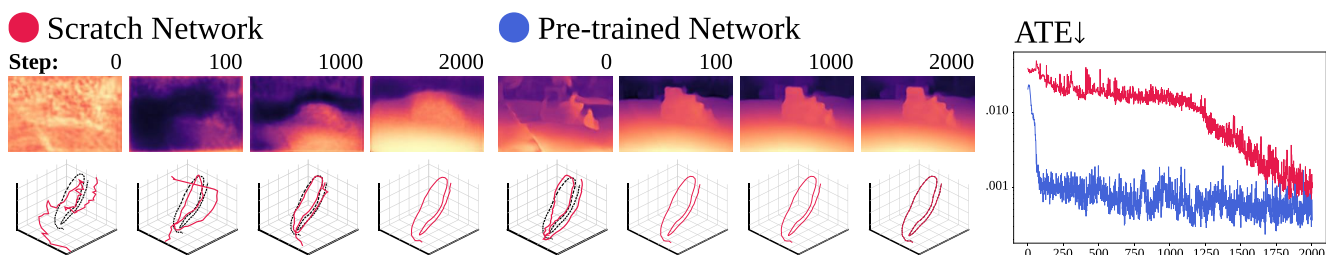


Figure 6. **Effects of pretraining.** While a randomly initialized FlowMap network often provides accurate poses after optimization, pre-training leads to faster convergence and slightly improved poses. Here we plot depth estimates at specific optimization steps (left) as well as pose accuracy with respect to COLMAP during optimization (right). Randomly initialized FlowMap networks often require more than 20,000 steps to match the accuracy of a pre-trained initialization at 2,000 steps.

163 4.4. Correspondence Weight MLP

164 The correspondence weight MLP is a three-layer MLP with
165 ReLU activations and 128 hidden units per layer. It takes as
166 input two corresponding image features and outputs a per-
167 correspondence weight between 0 and 1 via a sigmoid ac-
168 tivation. Here we use intermediate feature maps from the
169 depth network as the image features. These weights are

used in the weighted Procrustes pose solver.

5. Experiment Details

5.1. Image Resolution

To manage computational cost (our current implementation
loads the entire video into memory), we compute optical

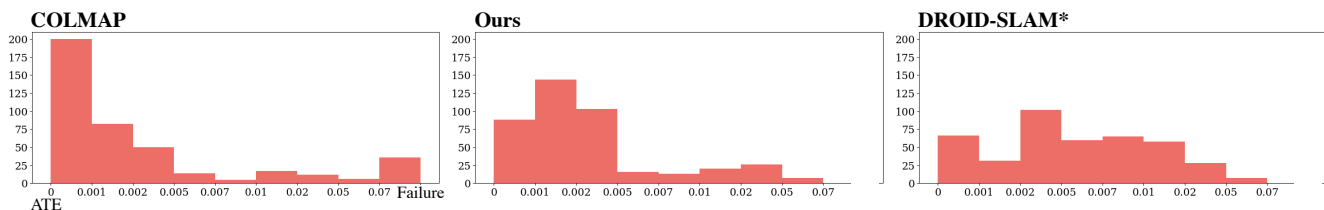
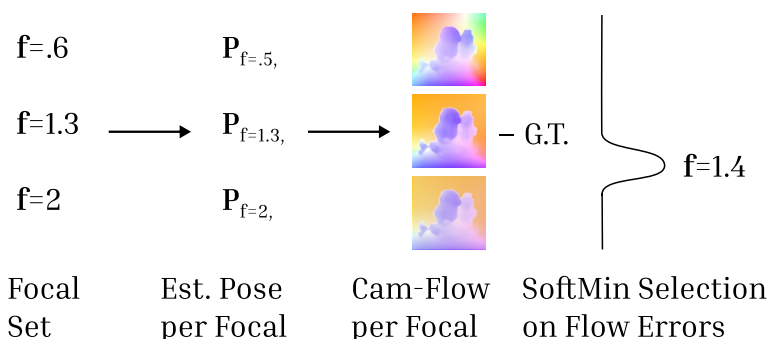


Figure 7. **Large-scale Robustness Study.** We run FlowMap and DROID-SLAM on 420 CO3D scenes across 10 categories and plot mean ATEs with respect to CO3D’s COLMAP-generated pose metadata. We also re-run COLMAP on the same data. Compared to DROID-SLAM, which requires ground-truth intrinsics, FlowMap produces notably lower ATEs. FlowMap’s ATE distribution is similar to one obtained by re-running COLMAP, with most ATEs falling under 0.005 in both cases.

(a) Intrinsics Estimation via Best-Explaining Focal

Cam-Induced Flow per Candidate Focal *Choose Focal with Flow Closest to GT*



(b) Depth Estimation via CNN

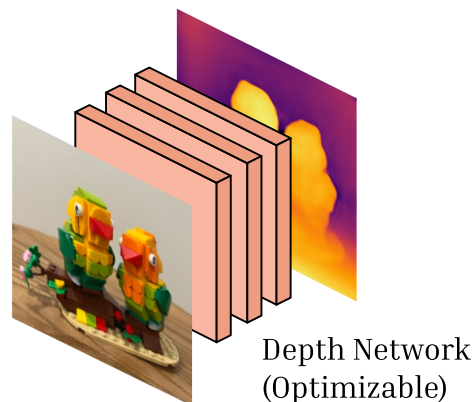


Figure 8. In (a) we illustrate our implicit focal length formulation, which considers a set of candidate focal lengths, assigns each one an error score, and softly selects the focal length with the lowest error. To calculate the error score for a focal length, we use that focal length to estimate a pose, and then compare the resulting pose-induced optical flow to the ground truth optical flow. In (b) we illustrate that we parameterize depth via the output of a monocular depth prediction CNN.

flow and point tracks at a resolution of around 700,000 pixels, then perform FlowMap optimization at 1/16th the resolution.

5.2. Hyperparameters

We train for 2000 steps using Adam and use a learning rate of $3e-5$. For the pose-as-variable experiments, we choose Euler angles as the parameterization of the rotation matrix.

5.3. Pre-Training Details

Before performing per-scene fine-tuning, we found it useful to learn a large-scale prior for better initialization. We use the same FlowMap loss formulation but train it on datasets of videos (instead of optimizing on a single scene). We use videos from CO3D, Real Estate 10K, and KITTI for pre-training. Note that we only use the raw videos from these datasets (no intrinsics, poses, or sparse geometry).

6. Limitations

While our method is much faster than MVS COLMAP, it is about 30 percent slower than COLMAP at its highest quality setting (on long sequences, about 20 minutes for our method vs. 14 minutes for COLMAP). It additionally requires significantly more GPU memory than COLMAP does. Our method’s pose and intrinsics predictions are less accurate and robust than COLMAP’s, as measured by ATE, though after Gaussian Splatting with fine-tuning of camera parameters, we often perform on par with COLMAP.

Our method further depends on correspondences estimated by point tracks and optical flow. While existing methods for computing point tracks and optical flow are robust, failures sometimes occur, and these failures can affect FlowMap’s accuracy if they are significant. On the other hand, FlowMap will directly improve alongside advancements in these domains.

Finally, our method is constrained to work on frame sequences with significant overlap (i.e., videos) and fails when input sequences contain significant scene motion. The

latter limitation is shared with COLMAP, though we hope that our method may serve as a step towards novel methods that address this shortcoming.

References

- [1] Eric R Chan, Koki Nagano, Matthew A Chan, Alexander W Bergman, Jeong Joon Park, Axel Levy, Miika Aittala, Shalini De Mello, Tero Karras, and Gordon Wetzstein. Generative novel view synthesis with 3d-aware diffusion models. *Proceedings of the International Conference on 3D Vision (3DV)*, 2023.
- [2] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [3] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4938–4947, 2020.
- [4] Cameron Smith, Yilun Du, Ayush Tewari, and Vincent Sitzmann. Flowcam: Training generalizable 3d radiance fields without camera poses via pixel-aligned scene flow. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [5] Ayush Tewari, Tianwei Yin, George Cazenavette, Semon Rezhikov, Joshua B Tenenbaum, Frédo Durand, William T Freeman, and Vincent Sitzmann. Diffusion with forward models: Solving stochastic inverse problems without direct supervision. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [6] Christopher Wewer, Kevin Raj, Eddy Ilg, Bernt Schiele, and Jan Eric Lenssen. latentsplat: Autoencoding variational gaussians for fast generalizable 3d reconstruction. In *arXiv*, 2024.

3DV 2025 Submission #265. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

	Fern (LLFF)					Flower (LLFF)					Fortress (LLFF)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	23.70	0.801	0.096	4.8	0.00233	29.07	0.877	0.084	6.6	0.00079	31.13	0.906	0.060	7.8	0.00049
Single Stage	23.64	0.797	0.098	4.9	0.00294	29.06	0.877	0.086	6.7	0.00065	31.05	0.908	0.058	7.9	0.00054
Expl. Focal Length	23.07	0.787	0.119	4.6	0.00296	29.13	0.874	0.079	6.4	0.00293	29.82	0.891	0.062	7.6	0.00223
Expl. Depth	4.71	0.001	0.785	2.9	0.00785	8.04	0.007	0.839	3.6	0.00666	2.60	0.001	0.774	4.1	0.00664
Expl. Pose	4.71	0.001	0.785	4.5	0.01118	15.51	0.569	0.428	6.3	0.00192	16.49	0.577	0.594	7.4	0.01302
No Tracks	23.58	0.796	0.099	4.3	0.00316	29.29	0.879	0.084	5.5	0.00337	30.92	0.906	0.059	6.3	0.00143
Random Init.	22.68	0.756	0.113	4.8	0.00371	28.47	0.864	0.084	6.6	0.00303	30.96	0.904	0.059	7.8	0.00068
Random Init. (20k)	23.50	0.791	0.098	44.2	0.00312	29.33	0.880	0.083	59.4	0.00054	31.04	0.911	0.057	69.7	0.00047
No Corresp. Weights	23.27	0.784	0.104	4.4	0.00311	27.61	0.844	0.090	6.0	0.00554	24.05	0.709	0.138	7.1	0.01363
	Horns (LLFF)					Orchids (LLFF)					Room (LLFF)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	28.35	0.903	0.071	10.6	0.00049	19.16	0.615	0.132	5.5	0.00127	32.93	0.958	0.037	7.8	0.00274
Single Stage	28.19	0.899	0.071	10.7	0.00051	19.33	0.623	0.129	5.6	0.00120	32.75	0.959	0.037	7.9	0.00265
Expl. Focal Length	24.57	0.823	0.153	10.5	0.00100	18.96	0.606	0.151	5.3	0.00184	32.13	0.953	0.040	7.6	0.00274
Expl. Depth	5.94	0.002	0.788	5.3	0.00603	6.25	0.003	0.886	3.2	0.00647	5.78	0.004	0.616	4.1	0.00710
Expl. Pose	14.66	0.506	0.685	10.0	0.01230	12.80	0.279	0.429	5.2	0.01496	16.92	0.767	0.466	7.3	0.00596
No Tracks	28.32	0.900	0.071	8.2	0.00173	19.21	0.616	0.133	4.8	0.00195	28.98	0.922	0.068	6.3	0.00938
Random Init.	23.93	0.729	0.172	10.6	0.00486	18.85	0.594	0.146	5.4	0.00188	29.19	0.920	0.067	7.8	0.00422
Random Init. (20k)	28.33	0.900	0.068	94.2	0.00054	19.40	0.629	0.126	49.5	0.00112	31.92	0.949	0.043	69.4	0.00288
No Corresp. Weights	28.17	0.893	0.072	9.6	0.00169	18.76	0.597	0.148	5.1	0.00307	31.81	0.952	0.041	7.1	0.00331
	Trex (LLFF)					Bonsai (MipNeRF 360)					Kitchen (MipNeRF 360)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	26.27	0.880	0.075	9.7	0.00655	32.24	0.950	0.047	24.2	0.00048	30.47	0.936	0.049	10.9	0.00041
Single Stage	26.65	0.886	0.073	9.8	0.00655	32.21	0.951	0.048	24.3	0.00046	30.26	0.925	0.051	11.1	0.00072
Expl. Focal Length	24.57	0.852	0.107	9.5	0.00766	21.36	0.689	0.231	24.1	0.00184	21.29	0.645	0.174	10.7	0.00449
Expl. Depth	5.56	0.002	0.759	4.8	0.04406	10.46	0.045	0.633	11.3	0.02830	5.08	0.016	0.753	5.3	0.00827
Expl. Pose	15.09	0.540	0.544	9.2	0.02277	13.12	0.425	0.577	22.8	0.01407	14.18	0.387	0.587	10.3	0.01669
No Tracks	24.36	0.831	0.110	7.8	0.02011	25.53	0.863	0.115	18.0	0.00291	25.48	0.794	0.112	8.5	0.00302
Random Init.	24.84	0.835	0.099	9.7	0.00598	18.38	0.585	0.342	24.2	0.01380	24.94	0.764	0.113	10.9	0.00345
Random Init. (20k)	26.45	0.882	0.076	86.5	0.00991	18.75	0.600	0.342	214.8	0.01433	31.69	0.945	0.044	96.8	0.00023
No Corresp. Weights	25.33	0.859	0.094	8.8	0.01083	25.59	0.841	0.118	21.8	0.00141	24.62	0.742	0.118	9.9	0.00422
	Counter (MipNeRF 360)					Barn (Tanks & Temples)					Caterpillar (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	26.80	0.862	0.121	24.2	0.00076	27.10	0.872	0.090	22.3	0.00048	28.25	0.830	0.113	22.3	0.00030
Single Stage	26.65	0.857	0.124	24.3	0.00083	26.28	0.857	0.104	22.4	0.00102	28.32	0.834	0.110	22.4	0.00026
Expl. Focal Length	21.82	0.697	0.237	24.1	0.00355	27.04	0.873	0.086	21.9	0.00085	27.12	0.789	0.133	22.1	0.00046
Expl. Depth	8.80	0.029	0.719	11.4	0.01003	17.01	0.591	0.489	10.7	0.00923	9.15	0.016	0.732	10.7	0.00841
Expl. Pose	14.40	0.506	0.554	22.9	0.00611	18.09	0.625	0.455	20.9	0.02160	17.57	0.491	0.554	21.1	0.00817
No Tracks	23.91	0.788	0.183	18.1	0.00240	25.41	0.837	0.122	16.1	0.00363	27.33	0.807	0.133	16.1	0.00095
Random Init.	26.05	0.847	0.131	24.2	0.00088	26.24	0.864	0.100	22.2	0.00079	26.27	0.750	0.169	22.2	0.00147
Random Init. (20k)	26.88	0.867	0.115	214.3	0.00064	26.80	0.871	0.091	197.3	0.00049	28.01	0.823	0.122	197.3	0.00031
No Corresp. Weights	17.93	0.575	0.391	21.8	0.01099	24.53	0.820	0.133	20.3	0.00244	25.93	0.734	0.174	20.1	0.00106
	Church (Tanks & Temples)					Courthouse (Tanks & Temples)					Family (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	28.29	0.883	0.074	22.4	0.00061	27.51	0.911	0.055	22.2	0.00129	27.96	0.889	0.067	22.1	0.00039
Single Stage	27.60	0.875	0.079	22.3	0.00061	27.67	0.914	0.054	22.3	0.00150	27.65	0.880	0.075	22.4	0.00033
Expl. Focal Length	27.29	0.856	0.088	22.0	0.00093	26.86	0.897	0.069	22.1	0.00234	27.10	0.873	0.082	22.0	0.00082
Expl. Depth	16.21	0.518	0.474	10.7	0.02314	3.52	0.001	0.745	10.7	0.00718	4.09	0.001	0.773	10.8	0.01403
Expl. Pose	17.66	0.582	0.457	21.0	0.00807	19.68	0.726	0.251	21.0	0.00511	15.79	0.562	0.507	21.1	0.03074
No Tracks	26.93	0.851	0.100	16.1	0.00259	25.27	0.858	0.108	16.1	0.00442	27.00	0.869	0.088	16.1	0.00172
Random Init.	27.45	0.858	0.089	22.2	0.00112	26.55	0.894	0.071	22.2	0.00314	26.36	0.858	0.093	22.2	0.00148
Random Init. (20k)	28.67	0.886	0.074	197.4	0.00030	27.62	0.911	0.054	197.8	0.00101	28.07	0.892	0.066	196.9	0.00019
No Corresp. Weights	27.86	0.875	0.081	20.3	0.00086	25.62	0.868	0.086	20.1	0.00264	19.01	0.629	0.313	20.2	0.00672
	Francis (Tanks & Temples)					Horse (Tanks & Temples)					Ignatius (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	31.90	0.903	0.080	22.4	0.00058	28.35	0.917	0.064	22.4	0.00054	24.54	0.773	0.131	22.4	0.00037
Single Stage	32.20	0.905	0.078	22.4	0.00054	11.42	0.635	0.496	22.3	0.03959	24.48	0.764	0.131	22.4	0.00024
Expl. Focal Length	30.89	0.884	0.108	22.0	0.00040	27.82	0.905	0.074	22.0	0.00102	23.12	0.723	0.157	22.0	0.00071
Expl. Depth	7.42	0.006	0.631	10.8	0.01956	2.67	0.000	0.691	10.7	0.02555	5.68	0.006	0.867	10.7	0.02181
Expl. Pose	18.19	0.639	0.464	20.9	0.03102	14.60	0.661	0.468	21.0	0.03918	12.48	0.314	0.640	20.9	0.02886
No Tracks	30.72	0.887	0.100	16.1	0.00113	25.50	0.882	0.101	16.1	0.00241	23.54	0.727	0.163	16.1	0.00144
Random Init.	29.44	0.862	0.122	22.3	0.00289	25.07	0.871	0.119	22.2	0.00380	23.50	0.737	0.159	22.1	0.00084
Random Init. (20k)	31.56	0.899	0.085	197.7	0.00138	28.16	0.915	0.067	197.0	0.00066	24.47	0.771	0.133	197.5	0.00034
No Corresp. Weights	28.92	0.850	0.130	20.1	0.00397	25.82	0.871	0.100	20.2	0.00275	21.89	0.655	0.197	20.2	0.00108
	M60 (Tanks & Temples)					Museum (Tanks & Temples)					Panther (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	23.23	0.805	0.190	22.4	0.00838	28.48	0.862	0.078	22.2	0.00070	27.50	0.882	0.105	22.3	0.00112
Single Stage	23.30	0.803	0.187	22.3	0.00832	28.62	0.866	0.076	22.4	0.00058	27.31	0.881	0.104	22.4	0.00118
Expl. Focal Length	19.65	0.696	0.278	22.1	0.01400	28.15	0.850	0.092	22.0	0.00124	2				

	Fern (LLFF)					Flower (LLFF)					Fortress (LLFF)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	23.70	0.801	0.096	4.8	0.00233	29.07	0.877	0.084	6.6	0.00079	31.13	0.906	0.060	7.8	0.00049
COLMAP	24.04	0.818	0.133	0.4	N/A	29.60	0.884	0.090	2.5	N/A	25.69	0.892	0.087	1.4	N/A
COLMAP (MVS)	24.33	0.826	0.094	6.7	N/A	29.82	0.888	0.085	11.3	N/A	30.97	0.909	0.059	13.9	N/A
DROID-SLAM*	23.13	0.752	0.125	0.1	0.00089	28.48	0.860	0.079	0.2	0.00162	30.05	0.856	0.065	0.3	0.00038
NoPE-NeRF*	19.33	0.520	0.580	1227.0	0.01470	19.63	0.540	0.470	1777.7	0.02581	21.00	0.530	0.510	533.4	0.02068
	Horns (LLFF)					Orchids (LLFF)					Room (LLFF)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	28.35	0.903	0.071	10.6	0.00049	19.16	0.615	0.132	5.5	0.00127	32.93	0.958	0.037	7.8	0.00274
COLMAP	27.82	0.888	0.095	1.5	N/A	19.33	0.636	0.126	0.7	N/A	25.69	0.927	0.096	0.3	N/A
COLMAP (MVS)	28.68	0.902	0.067	20.5	N/A	19.79	0.657	0.117	8.7	N/A	33.43	0.963	0.035	14.6	N/A
DROID-SLAM*	28.37	0.881	0.064	0.5	0.00045	18.44	0.555	0.179	0.2	0.00072	27.63	0.924	0.078	0.3	0.00051
NoPE-NeRF*	11.88	0.370	0.820	2597.7	0.07315	13.11	0.270	0.620	1377.9	0.05492	17.79	0.650	0.590	2500.5	0.03714
	Trex (LLFF)					Bonsai (MipNeRF 360)					Kitchen (MipNeRF 360)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	26.27	0.880	0.075	9.7	0.00655	32.24	0.950	0.047	24.2	0.00048	30.47	0.936	0.049	10.9	0.00041
COLMAP	27.95	0.912	0.062	1.1	N/A	32.64	0.949	0.058	6.9	N/A	28.82	0.936	0.056	3.4	N/A
COLMAP (MVS)	28.92	0.922	0.049	18.4	N/A	33.14	0.957	0.045	52.2	N/A	31.33	0.948	0.045	22.4	N/A
DROID-SLAM*	27.36	0.898	0.067	0.3	0.00062	31.96	0.947	0.045	0.9	0.00016	29.75	0.903	0.054	0.4	0.00015
NoPE-NeRF*	18.71	0.550	0.550	2614.1	0.04796	13.49	0.370	0.770	2615.2	0.04475	14.86	0.370	0.710	516.3	0.05471
	Counter (MipNeRF 360)					Barn (Tanks & Temples)					Caterpillar (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	26.80	0.862	0.121	24.2	0.00076	27.10	0.872	0.090	22.3	0.00048	28.25	0.830	0.113	22.3	0.00030
COLMAP	28.39	0.899	0.107	4.1	N/A	27.18	0.874	0.108	3.5	N/A	28.05	0.825	0.134	6.6	N/A
COLMAP (MVS)	28.61	0.909	0.089	52.9	N/A	27.91	0.889	0.075	51.5	N/A	28.52	0.839	0.103	51.1	N/A
DROID-SLAM*	27.78	0.890	0.099	0.7	0.00019	27.03	0.877	0.082	0.8	0.00029	28.13	0.829	0.108	0.9	0.00020
NoPE-NeRF*	12.44	0.390	0.770	2607.8	0.03342	13.06	0.460	0.710	2608.4	0.03761	16.42	0.390	0.680	2469.9	0.03112
	Church (Tanks & Temples)					Courthouse (Tanks & Temples)					Family (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	28.29	0.883	0.074	22.4	0.00061	27.51	0.911	0.055	22.2	0.00129	27.96	0.889	0.067	22.1	0.00039
COLMAP	27.93	0.866	0.107	6.3	N/A	27.79	0.916	0.056	5.9	N/A	27.13	0.878	0.092	5.0	N/A
COLMAP (MVS)	28.71	0.890	0.068	50.9	N/A	28.56	0.926	0.044	51.9	N/A	28.40	0.897	0.062	50.9	N/A
DROID-SLAM*	27.79	0.869	0.084	0.8	0.00065	27.94	0.916	0.051	0.9	0.00034	27.78	0.873	0.081	0.8	0.00040
NoPE-NeRF*	12.91	0.400	0.700	2575.8	0.02752	14.92	0.510	0.590	2599.3	0.03462	12.87	0.470	0.700	2597.4	0.03232
	Francis (Tanks & Temples)					Horse (Tanks & Temples)					Ignatius (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	31.90	0.903	0.080	22.4	0.00058	28.35	0.917	0.064	22.4	0.00054	24.54	0.773	0.131	22.4	0.00037
COLMAP	31.85	0.896	0.124	3.6	N/A	27.34	0.903	0.097	3.4	N/A	24.95	0.781	0.153	5.6	N/A
COLMAP (MVS)	32.73	0.913	0.069	51.1	N/A	28.82	0.926	0.062	53.2	N/A	24.93	0.795	0.113	51.2	N/A
DROID-SLAM*	22.23	0.753	0.275	0.9	0.00041	27.61	0.909	0.069	0.8	0.00051	24.28	0.750	0.142	0.8	0.00025
NoPE-NeRF*	17.27	0.570	0.640	524.9	0.02569	9.87	0.590	0.700	2587.4	0.04710	10.90	0.260	0.780	2583.2	0.04241
	M60 (Tanks & Temples)					Museum (Tanks & Temples)					Panther (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	23.23	0.805	0.190	22.4	0.00838	28.48	0.862	0.078	22.2	0.00070	27.50	0.882	0.105	22.3	0.00112
COLMAP	22.04	0.803	0.219	6.2	N/A	28.94	0.863	0.100	5.3	N/A	27.32	0.882	0.129	5.0	N/A
COLMAP (MVS)	21.75	0.791	0.221	51.9	N/A	29.05	0.874	0.070	50.6	N/A	27.96	0.891	0.101	52.2	N/A
DROID-SLAM*	22.66	0.792	0.195	0.7	0.00667	27.74	0.833	0.096	0.8	0.00088	27.48	0.878	0.106	0.8	0.00150
NoPE-NeRF*	12.67	0.490	0.720	2485.1	0.04258	14.26	0.430	0.800	2606.9	0.03224	13.71	0.500	0.690	2591.0	0.03854
	Playground (Tanks & Temples)					Train (Tanks & Temples)					Truck (Tanks & Temples)				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE
FlowMap	24.29	0.727	0.192	22.2	0.00096	26.22	0.870	0.077	22.2	0.00082	24.34	0.828	0.098	22.3	0.00078
COLMAP	22.24	0.684	0.292	7.4	N/A	26.09	0.857	0.104	8.4	N/A	25.57	0.848	0.104	4.9	N/A
COLMAP (MVS)	22.92	0.693	0.230	51.6	N/A	27.43	0.888	0.063	51.4	N/A	26.39	0.864	0.080	50.4	N/A
DROID-SLAM*	21.11	0.642	0.301	0.7	0.00284	26.51	0.872	0.069	0.8	0.00088	21.48	0.739	0.208	0.8	0.00127
NoPE-NeRF*	13.53	0.360	0.770	2613.1	0.04120	13.18	0.440	0.670	2614.8	0.04052	11.71	0.410	0.740	2603.3	0.04583
	Bench (CO3D)					Hydrant (CO3D)									
Method	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE	PSNR ↑	SSIM ↑	LPIPS ↓	Time (min.) ↓	ATE					
FlowMap	33.17	0.927	0.045	22.0	0.03094	29.05	0.865	0.083	22.1	0.00083					
COLMAP	19.87	0.600	0.309	17.2	N/A	30.46	0.900	0.070	8.0	N/A					
COLMAP (MVS)	20.00	0.616	0.292	53.2	N/A	30.70	0.908	0.057	50.8	N/A					
DROID-SLAM*	22.48	0.699	0.206	0.9	0.03433	29.46	0.880	0.073	0.7	0.00024					
NoPE-NeRF*	13.20	0.500	0.750	2604.0	0.03432	16.74	0.300	0.790	2605.8	0.03864					

Table 3. Results for all individual scenes on all datasets.