GeCoNeRF: Few-shot Neural Radiance Fields via Geometric Consistency

Min-Seop Kwak*, Jiuhn Song*, Seungryong Kim†

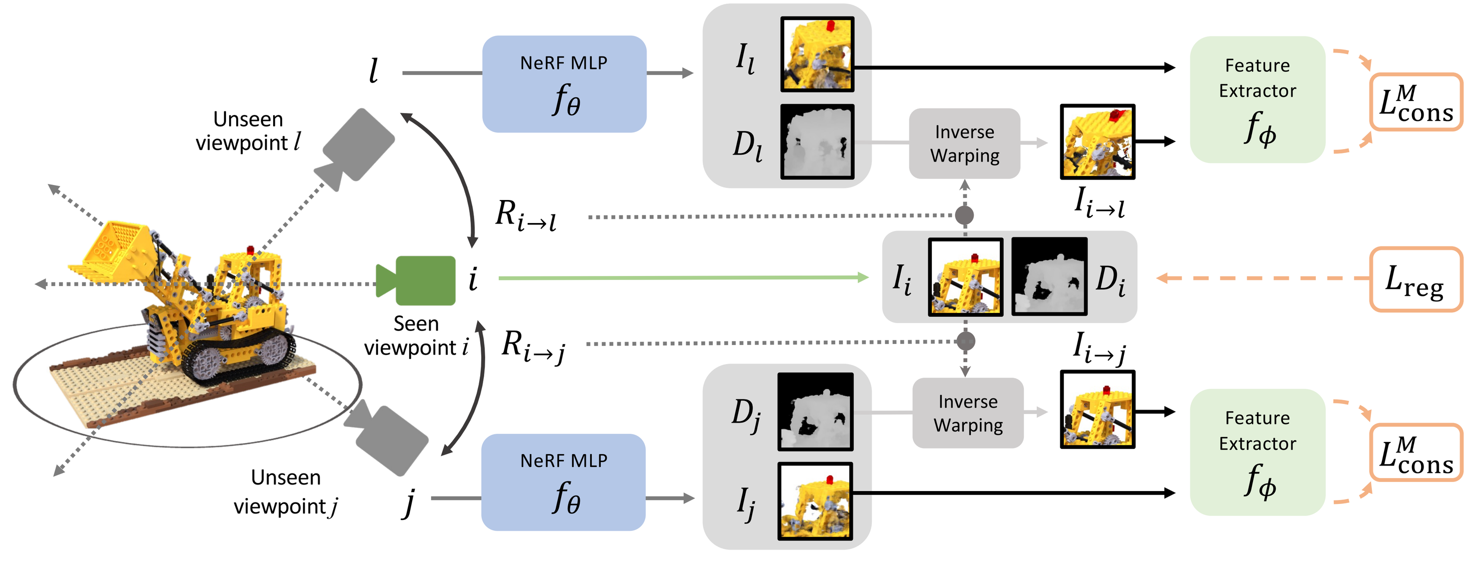
Computer Vision Laboratory, Korea University

Project Page
QR Code

Motivation

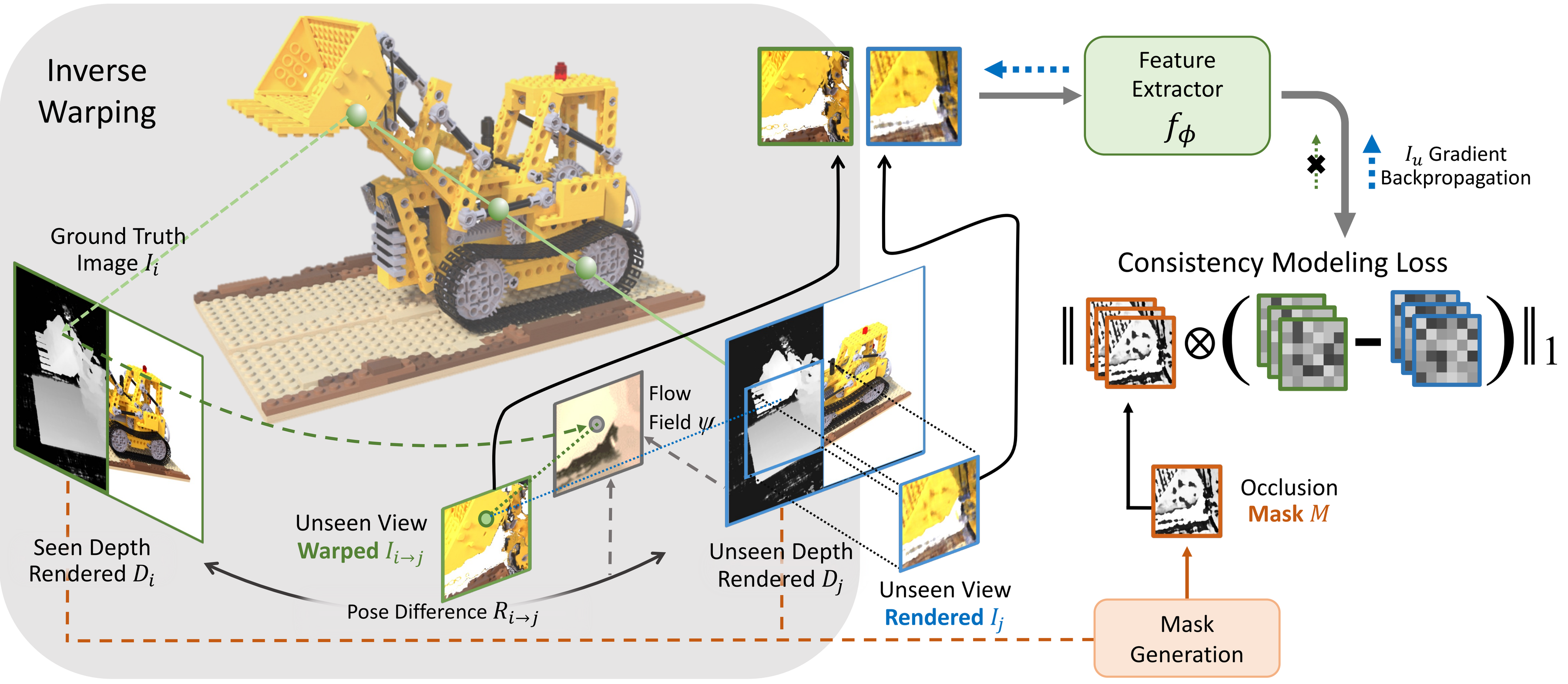
- NeRF requires numerous (100+) densely, well distributed calibrated images for optimization, which limits applicability
- The task of few-shot NeRF aims to optimize high-fidelity neural radiance field in such sparse scenario
- Previous works' reliance on handcrafted methods or inability to extract local and fine structures limit their performance
- We propose a novel methodology, GeCoNeRF, short for Geometric Consistency NeRF, a regularization technique that enforces a geometric consistency across different views in a self-supervised manner with a depth-guided warping and a geometry-aware consistency modeling.

Architecture



- Warping occurs between observed and unobserved viewpoints using depth geometry recovered & rendered by NeRF
- Consistency loss function $\mathcal{L}^M_{cons}$ is applied between the unobserved viewpoint image and *warped* observed viewpoint image for regularization
- Disparity regularization loss $\mathcal{L}_{reg}$ regularizes depth at seen viewpoints.

Overview and Contributions



Rendered Depth-Guided Warping

- To generate pseudo ground truth patches for unknown viewpoints, ground truth image I_i is warped to unseen viewpoint j using unseen rendered depth D_j in a following manner:

$$D(\mathbf{r}_p) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}_p(t))dt \qquad p_{j\to i} \sim K R_{j\to i} D_j(p_j) K^{-1} p_j \qquad I_{i\to j}(p_j) = \text{sampler}(I_i; p_{j\to i})$$

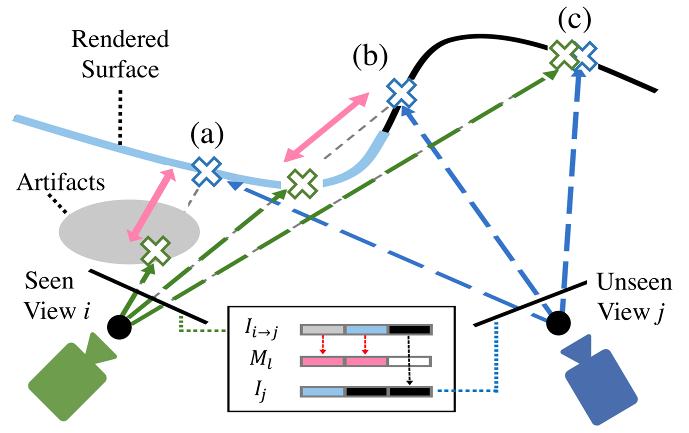
Feature-level Consistency Modeling

- We define the consistency between the unknown view rendered patch I_j and warped patch $I_{i\to j}$ to encourage additional regularization for local geometric consistency.
- To overcome failures in modeling non-Lambertian surfaces and fine-grained occlusions, we propose masked feature-level regularization loss that focuses upon structural consistency.

$$\mathcal{L}^M_{\text{cons}} = \sum_{l=1}^{L} \frac{1}{C_l m_l} \| M_l \odot (f^l_\phi(I_{i\to j}) - f^l_\phi(I_j)) \|$$

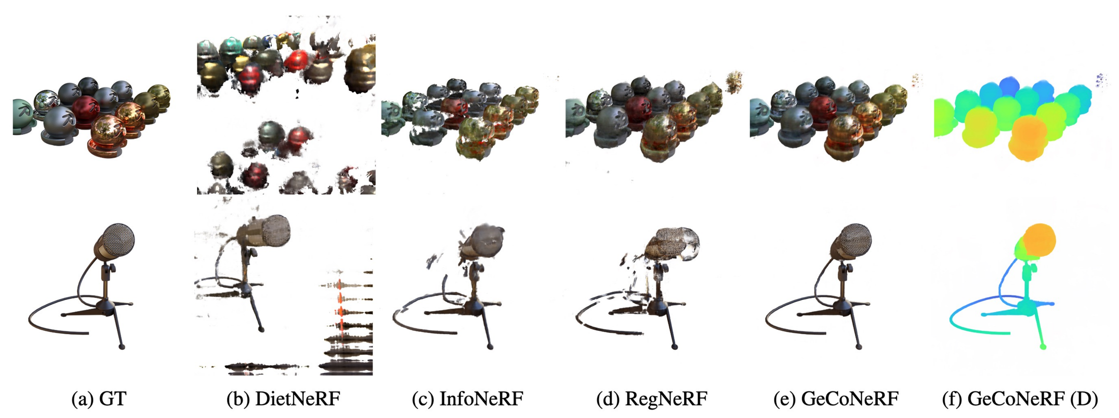
Occlusion Handling, Disparity Regularization, Progressive Annealing

- Consistency mask M_l is constructed to let NeRF ignore regions with geometric inconsistencies, along with seen view disparity regularization.
- We recognize that progressive freq. encoding annealing has powerful regularization effect and apply it at the beginning stages of training.



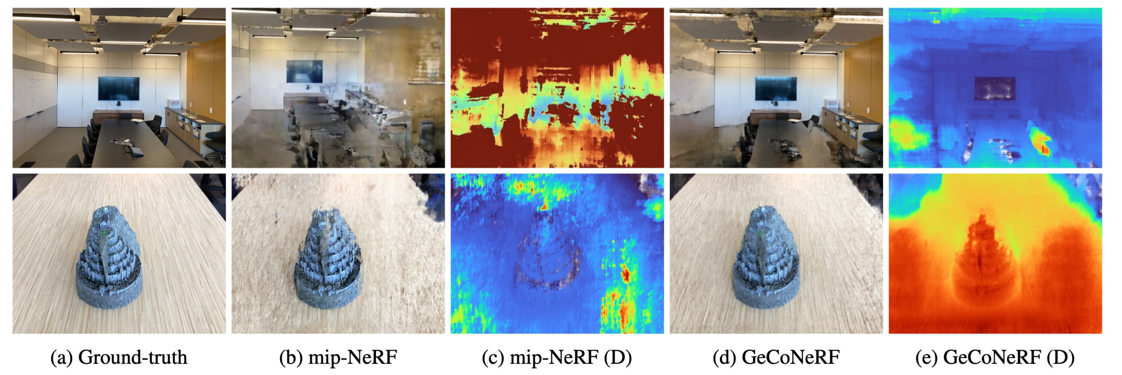
$$M(p_j) = [\| D_j(p_j) - D_i(p_{j\to i}) \| < \tau]$$

Qualitative Comparisons



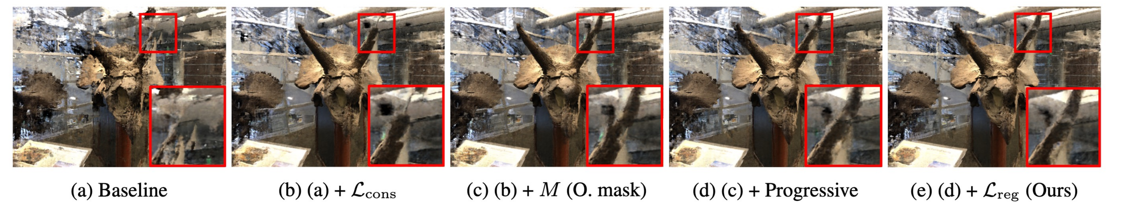
Comparison w/ Previous Methods

In 3-view setting, our method captures fine details more robustly (such as the wire in the *mic* scene) and produces less artifacts (background in the *materials* scene) compared to previous methods.



Comparison w/ Baseline

Comparison with baseline mip-NeRF on LLFF dataset shows that our model learns of coherent depth and geometry in sparse three-view scenario.



Qualitative Ablation Study

Our qualitative ablation results on Horns scene shows the contribution of each module in the performance of our model.

Quantitative Comparisons

Table 1. **Quantitative comparison on NeRF-Synthetic (Mildenhall et al., 2020) and LLFF (Mildenhall et al., 2019) datasets.**

Methods	NeRF-Synthetic (Mildenhall et al., 2020)				LLFF (Mildenhall et al., 2019)			
	PSNR ↑	SSIM ↑	LPIPS ↓	Avg. ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Avg. ↓
NeRF (Mildenhall et al., 2020)	14.73	0.734	0.451	0.199	13.34	0.373	0.451	0.255
mip-NeRF (Barron et al., 2021)	17.71	0.798	0.745	0.178	14.62	0.351	0.495	0.246
DietNeRF (Jain et al., 2021)	16.06	0.793	0.306	0.151	14.94	0.370	0.496	0.232
InfoNeRF (Kim et al., 2022)	18.65	0.811	0.230	0.111	14.37	0.349	0.457	0.238
RegNeRF (Niemeyer et al., 2022)	18.01	0.842	0.352	0.132	**19.08**	0.587	**0.336**	0.146
GeCoNeRF (Ours)	**19.23**	**0.866**	**0.201**	**0.096**	18.77	**0.596**	0.338	**0.145**

Quantitative comparisons show our model's competitive results in NeRF-Synthetic dataset and LLFF dataset, whose PSNR results show large improvement in comparison to mip-NeRF baseline and competitive compared to previous SOTA, RegNeRF.

Table 2. **Ablation study.**

Components	PSNR↑	SSIM↑	LPIPS↓	Avg.↓
(a) Baseline	14.62	0.351	0.495	0.246
(b) (a) + $\mathcal{L}_{\text{cons}}$	18.10	0.529	0.408	0.164
(c) (b) + M (O. mask)	18.24	0.535	0.379	0.159
(d) (c) + Progressive	18.46	0.552	0.349	0.151
(e) (d) + $\mathcal{L}_{\text{reg}}$ (Ours)	**18.55**	**0.592**	**0.340**	**0.150**

Table 3. **Progressive training ablation.**

Components	PSNR↑	SSIM↑	LPIPS↓	Avg. ↓
w/o prog. anneal	18.50	0.852	0.781	0.161
w/o prog. pose	16.96	0.799	0.811	0.194
w/o both	17.04	0.788	0.823	0.197
GeCoNeRF (Ours)	**19.23**	**0.866**	**0.723**	**0.148**