IntrinsicNeRF: Learning Intrinsic Neural Radiance Fields for Editable Novel View Synthesis

[Weicai Ye, Shuo Chen]Co-Authors, Chong Bao, Hujun Bao, Marc Pollefeys, Zhaopeng Cui, Guofeng Zhang*

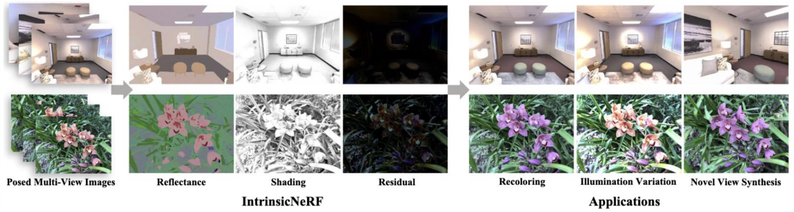
code zju3dv.github.io/intrinsic_nerf

ICCV23 PARIS

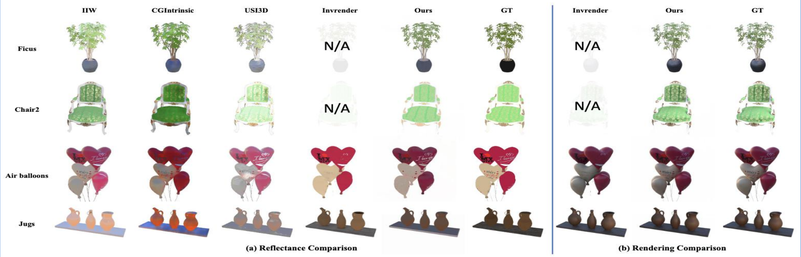
1. Motivation

Our Problem:

- Input: multi-view posed images of static scenes
- Output: factorize them into multi-view **consistent intrinsic components**: reflectance, shading, and residual layers;
- Support online applications: recoloring, **editable view synthesis**, etc.

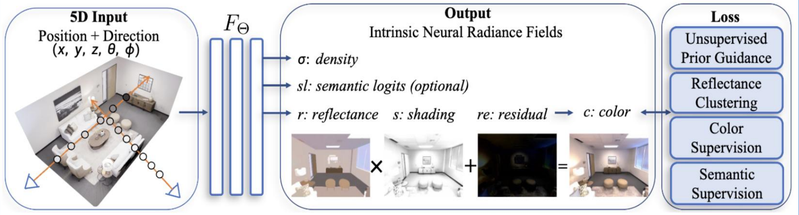


- Existing neural rendering with inverse rendering method (e.g. PhySG and InvRender): **rely on accurate surface**, may fail if bad geometry; can only perform editable view synthesis on **object-specific** scenes.



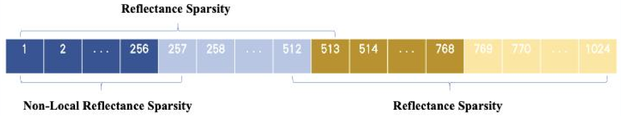
Our Solution:

- **Intrinsic Neural Radiance Fields** which **introduce intrinsic decomposition into NeRF**
- **Unsupervised intrinsic prior** with **distance-aware point sampling**
- **Adaptive reflectance iterative clustering** optimization
- **Hierarchical clustering and indexing** method with **semantic constraints**



2. Unsupervised Intrinsic Prior

- Distance-Aware Point Sampling: first randomly sample 512 points, and then randomly sample the remaining 512 points in the eight neighborhoods of each sampled point.

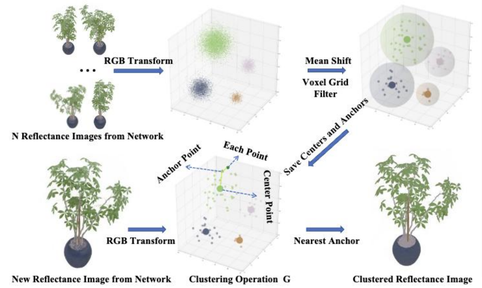


- Chromaticity Prior: $L_{chrom(\mathbf{x})} = \|c_r(\mathbf{x}) - c(\mathbf{x})\|_2^2,$
- Reflectance Sparsity: $L_{reflect(\mathbf{x})} = \sum_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} \omega_{cs(\mathbf{x},\mathbf{y})} \|r(\mathbf{x}) - r(\mathbf{y})\|_2^2,$
- Non-Local Reflectance Sparsity: $L_{non-local(\mathbf{x})} = \sum_{\mathbf{y}\in\mathcal{F}(\mathbf{x})} \omega_{cs(\mathbf{x},\mathbf{y})} \|r(\mathbf{x}) - r(\mathbf{y})\|_2^2,$
- Shading Smoothness: $L_{shade(\mathbf{x})} = \sum_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} \|c(\mathbf{x}) - c(\mathbf{y})\|_2^2 \|s(\mathbf{x}) - s(\mathbf{y})\|_2^2.$
- Intrinsic Residual Constraints: $L_{residual(\mathbf{x})} = \|re(\mathbf{x})\|_2^2$
- Intensity Prior: $L_{intensity(\mathbf{x})} = \|i_r(\mathbf{x}) - i(\mathbf{x})\|_2^2,$

3. Clustering Optimization

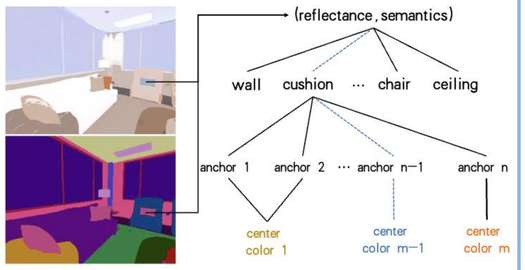
Adaptive Reflectance Clustering

- RGB Transform: $f([\mathrm{r},\mathrm{g},\mathrm{b}]) = [\beta\frac{\mathrm{r}+\mathrm{g}+\mathrm{b}}{3}, \frac{\mathrm{r}}{\mathrm{r}+\mathrm{g}+\mathrm{b}}, \frac{\mathrm{g}}{\mathrm{r}+\mathrm{g}+\mathrm{b}}],$
- Mean Shift
- Clustering Operation G
- Voxel Grid Filter
- Optimization: $L_{cluster(\mathbf{x})} = \|r_{cluster}(\mathbf{x}) - r(\mathbf{x})\|_2^2.$



Hierarchical Clustering and Indexing Method

- Different adjacent instances of similar reflectance in a scene are incorrectly clustered together.
- Introduce hierarchical clustering with semantic constraints



4. Experiments

Quantitative Results

- Blender Object: for reflectance estimation, achieved the best results on our dataset and ranked 2nd on Invrender dataset; for view synthesis, achieved the best results on both dataset.

Method	Reflectance (Invrender dataset) PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	LMSE ↓	View Synthesis (Invrender dataset) PSNR ↑	SSIM ↑	LPIPS ↓	Reflectance (our dataset) PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	LMSE ↓	View Synthesis (our dataset) PSNR ↑	SSIM ↑	LPIPS ↓
IIW[4]	22.0284	0.9307	0.0847	0.0099	**0.0120**	-	-	-	20.5299	0.9079	0.1131	0.0102	0.0727	-	-	-
CGIntrinsic[31]	20.1583	0.9209	0.0996	0.0129	0.0141	-	-	-	18.3542	0.8999	0.1229	0.0156	0.0659	-	-	-
USI3D [36]	20.7571	0.9267	0.0887	0.0079	0.0149	-	-	-	19.1489	0.9115	0.1070	0.0135	0.0524	-	-	-
NeRFactor[76]	19.9167	0.9156	0.1354	0.0059	0.0210	23.0133	0.9277	0.0822	21.4440	0.9170	0.1055	**0.0063**	**0.0444**	20.6880	0.8733	0.1185
PhySG[74]	23.3748	0.9231	0.1092	0.0034	0.0396	25.4225	0.9388	0.0804	-	-	-	-	-	-	-	-
Invrender [72]	**26.3078**	**0.9380**	**0.0572**	0.0022	0.0226	29.3870	0.9522	0.0505	-	-	-	-	-	-	-	-
Baseline	16.3209	0.8637	0.1301	0.0254	0.1955	**34.0036**	**0.9670**	**0.0252**	14.8572	0.8397	0.1738	0.0451	0.1849	**28.2604**	**0.9383**	**0.0339**
Baseline + w/ prior.	21.7370	0.9278	0.1086	0.0055	0.0186	33.4909	0.9638	0.0304	20.9646	0.9140	0.1216	0.0095	0.0538	28.0633	0.9370	0.0369
Ours	24.2642	0.9371	0.0880	**0.0021**	0.0173	33.4967	0.9630	0.0306	**22.5677**	**0.9267**	**0.0975**	0.0066	0.0474	27.9494	0.9357	0.0372

- Ablation Studies of Each Loss Constraints.

Metric \ Method	w/o L_{chrom}	w/o $L_{reflect}$	w/o $L_{non-local}$	w/o L_{shade}	w/o $L_{cluster}$	w/o $L_{residual}$	w/o $L_{intensity}$	w/o all prior	Ours
PSNR ↑	22.0243	22.4955	23.3032	22.9874	21.3508	21.1288	18.7466	15.5891	**23.4160**
MSE ↓	0.0067	0.0060	0.0044	0.0048	0.0075	0.0074	0.0172	0.0352	**0.0043**
LMSE ↓	0.0392	0.0378	0.0323	0.0338	0.0362	0.0387	0.0339	0.1902	**0.0323**

- Comparable Results for View Synthesis on Blender Object.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
NeRF [46]	31.0838	0.9525	0.0302
Ours	30.7230	0.9494	0.0339

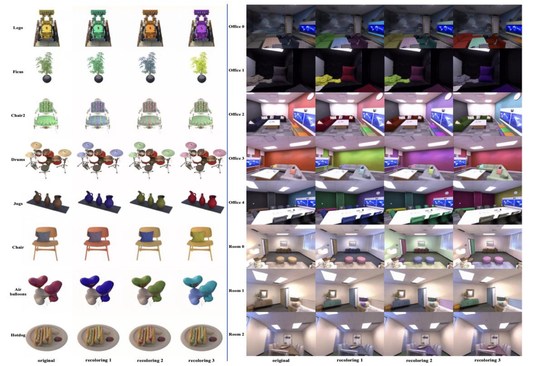
- Comparable Results for View Synthesis and Semantic Segmentation on Replica Scene.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	mIoU ↑
[79]	30.9770	0.8955	0.1066	0.9725
Ours	30.7044	0.8908	0.1140	0.9702

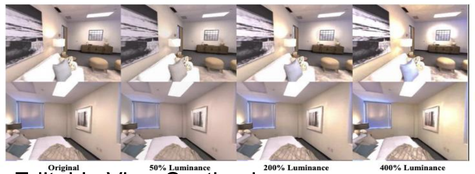
Qualitative Results

- Reflectance Comparison on Replica Scene

- Scene Recoloring



- Illumination Variation



- Editable View Synthesis