école normale supérieure paris–saclay
université PARIS-SACLAY
CENTRE BORELLI

centreborelli.github.io/satnerf

Sat-NeRF: Learning Multi-View Satellite Photogrammetry With Transient Objects and Shadow Modeling Using RPC Cameras

Roger Marí, Gabriele Facciolo, Thibaud Ehret

[1]Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, 91190, Gif-sur-Yvette, France

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA
IEEE COMPUTER SOCIETY
EARTHVISION

Contributions

- A NeRF variant robust to the radiometric inconsistencies of multi-date satellite images, like shadows caused by the sun and small transient objects.
- Study the advantages of a clever use of satellite RPC camera models to increase the performance of NeRF architectures for satellite images.

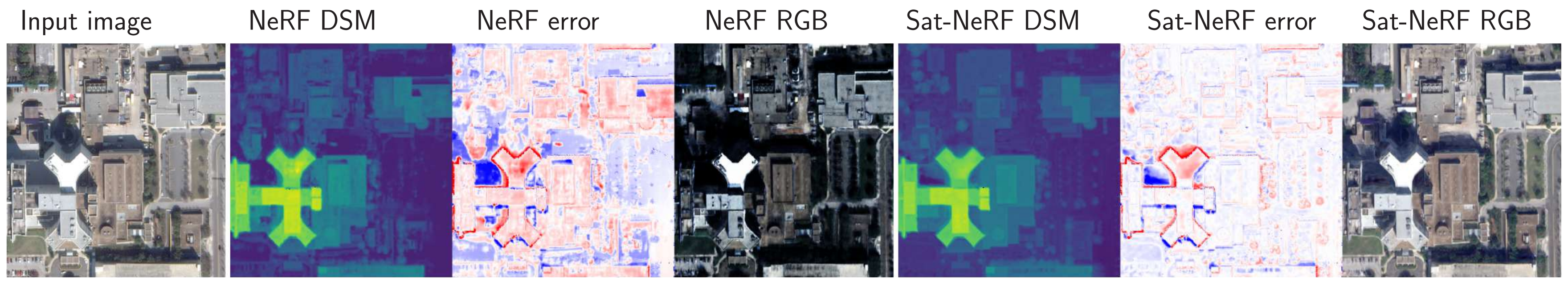
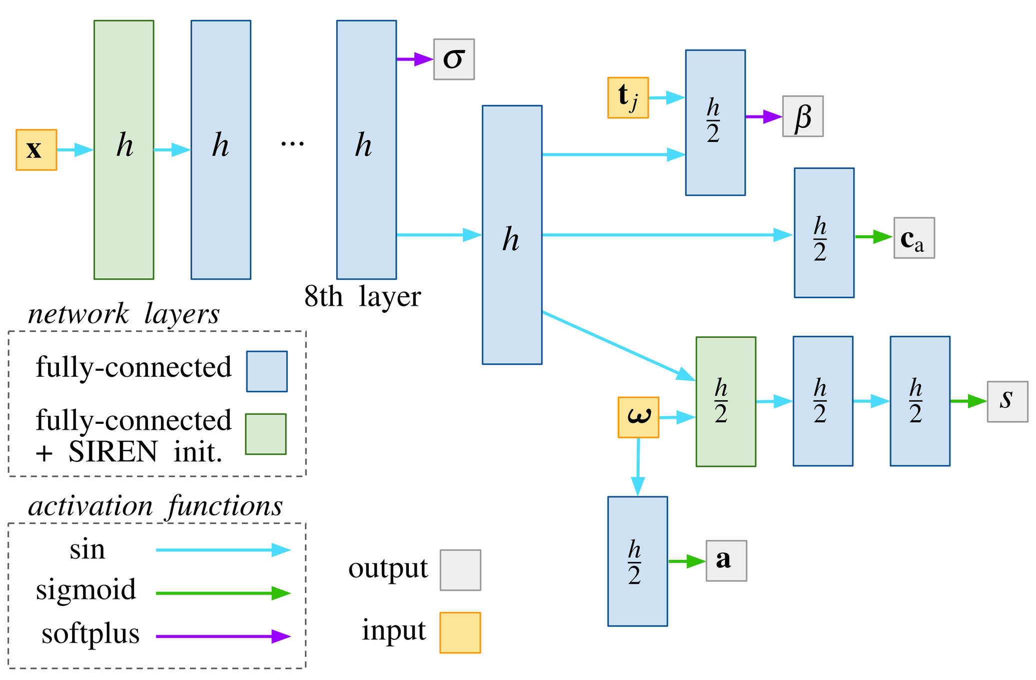


Figure 1: RGB renderings and surface models learned with NeRF and Sat-NeRF.

Architecture and fundamentals



$\mathcal{F} : (\mathbf{x}, \boldsymbol{\omega}, \mathbf{t}_j) \mapsto (\sigma, \mathbf{c}_\text{a}, s, \mathbf{a}, \beta)$

inputs:
$\mathbf{x}$: spatial coordinates
$\boldsymbol{\omega}$: direction of solar rays
$\mathbf{t}_j$: transient embedding of image j

outputs:
σ: volume density
$\mathbf{c}_\text{a}$: albedo RGB color
s: shadow-aware shading scalar
$\mathbf{a}$: ambient RGB color
β: uncertainty coefficient

Figure 2: Sat-NeRF MLP architecture.

Sat-NeRF is trained using the volume rendering strategy of NeRF [1]. Each ray $\mathbf{r}$ of a batch $\mathcal{R}$ is discretized into a set of points $\mathbf{x}_i = \mathbf{o} + t_i\mathbf{d}$, where $t_i \in [t_\text{near}, t_\text{far}]$, and $\mathbf{o}$ and $\mathbf{d}$ are the origin and direction vectors. The rendered color $\mathbf{c}(\mathbf{r})$ and depth $d(\mathbf{r})$ of $\mathbf{r}$ are

$$\mathbf{c}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i \qquad d(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i t_i. \tag{1}$$

where $\mathbf{c}_i$ is the color predicted at $\mathbf{x}_i$. Opacity α_i and transmittance T_i are obtained from the volume density σ_i and $\sigma_j : j < i$ (to handle occlusions).

The Sat-NeRF loss uses a color term L_RGB and two auxiliary solar correction L_RGB and depth supervision L_DS terms, weighted with $\lambda_\text{SC}, \lambda_\text{DS} \in \mathbb{R}$:

$$L_\text{RGB}(\mathcal{R}) + \lambda_\text{SC} L_\text{SC}(\mathcal{R}_\text{SC}) + \lambda_\text{DS} L_\text{DS}(\mathcal{R}_\text{DS}) \tag{2}$$

Uncertainty weighting for transient objects

Similarly to [2] we let the MLP learn an uncertainty coefficient β that weights the contribution of each point to the color term of the loss.

$$L_\text{RGB}(\mathcal{R}) = \sum_{\mathbf{r} \in \mathcal{R}} \frac{\|\mathbf{c}(\mathbf{r}) - \mathbf{c}_\text{GT}(\mathbf{r})\|_2^2}{2\beta'(\mathbf{r})^2} + \left(\frac{\log \beta'(\mathbf{r}) + \eta}{2} \right) \tag{3}$$

where $\beta'(\mathbf{r}) = \beta(\mathbf{r}) + 0.05$ and $\eta = 3$ to prevent negative values.

$$\beta(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \beta_i \tag{4}$$

Shadow-aware irradiance model

Color is learned as in [3], using a linear combination of an albedo color $\mathbf{c}_\text{a}$ with a shading scalar s and ambient hue $\mathbf{a}$ related to the solar direction $\boldsymbol{\omega}$.

$$\mathbf{c}_i = \mathbf{c}_\text{a}(\mathbf{x}) \cdot \big(s(\mathbf{x}, \boldsymbol{\omega}) + (1 - s(\mathbf{x}, \boldsymbol{\omega})) \cdot \mathbf{a}(\boldsymbol{\omega})\big), \tag{5}$$

The solar correction term is used to prevent instabilities in s:

$$L_\text{SC}(\mathcal{R}_\text{SC}) = \sum_{\mathbf{r} \in \mathcal{R}_\text{SC}} \left(\sum_{i=1}^{N_\text{SC}} (T_i - s_i)^2 + 1 - \sum_{i=1}^{N_\text{SC}} T_i \alpha_i s_i \right), \tag{6}$$

where the rays in $\mathcal{R}_\text{SC}$ follow the solar direction $\boldsymbol{\omega}$.

Point sampling using RPC camera models

Given the predefined altitude boundaries $[h_\text{max}, h_\text{min}]$, the ray $\mathbf{r}$ intersecting the pixel $\mathbf{p}$ in image j corresponds to the line between the boundary points:

$$\mathbf{x}_\text{start} = \mathcal{L}_j(\mathbf{p}, h_\text{max})_\text{ECEF} \qquad \mathbf{x}_\text{end} = \mathcal{L}_j(\mathbf{p}, h_\text{min})_\text{ECEF}, \tag{7}$$

where $\mathcal{L}_j$ is the localization function of image j. All point coordinates are expressed in the interval $[-1, 1]$ using an offset and scaling normalization.

RPC refinement and depth supervision

Bundle adjustment is used to refine the RPCs of all training images [4]. Without this preprocessing step, we observe a loss of accuracy in all metrics.

The sparse point cloud used in the BA can be reused for depth supervision:

$$L_\text{DS}(\mathcal{R}_\text{DS}) = \sum_{\mathbf{r} \in \mathcal{R}_\text{DS}} w(\mathbf{r}) \left(d(\mathbf{r}) - \|\mathbf{X}(\mathbf{r}) - \mathbf{o}(\mathbf{r})\|_2 \right)^2, \tag{8}$$

where $\|\mathbf{X}(\mathbf{r}) - \mathbf{o}(\mathbf{r})\|_2$ is the target depth of the BA point $\mathbf{X}(\mathbf{r})$. $\mathcal{R}_\text{DS}$ is the set of rays intersecting BA points. $w(\mathbf{r})$ is a reprojection error aware weight.

Results

- Trained/tested using the IEEE GRSS Data Fusion Contest 2019 benchmark.

	PSNR ↑				SSIM ↑				Altitude MAE [m] ↓			
Area index	004	068	214	260	004	068	214	260	004	068	214	260
NeRF	17.93	10.26	15.26	14.95	0.559	0.536	0.736	0.443	3.327	2.591	2.691	3.257
S-NeRF + SC	26.14	24.07	24.93	21.24	0.871	0.891	0.943	0.825	1.472	1.374	2.406	2.299
Sat-NeRF	26.16	24.80	25.54	21.88	0.876	0.903	0.951	0.840	1.416	1.275 ●	2.125	2.428
Sat-NeRF + SC	26.67	25.07	25.50	21.78	0.884	0.908	0.950	0.842	**1.288** ●	**1.249** ●	2.009	1.864
Sat-NeRF + SC (no BA)	21.55	22.87	24.53	20.96	0.571	0.874	0.942	0.816	1.577	1.392	2.176	1.875
Sat-NeRF + DS	26.43	25.27	25.69	21.94	0.879	0.913	0.952	0.842	1.420	1.298	1.714 ●	**1.624** ●
Sat-NeRF + DS + SC	26.62	25.00	25.66	21.66	0.881	0.909	0.952	0.839	1.366 ●	1.277	**1.676** ●	1.638 ●
S2P (10 pairs) [5]	—	—	—	—	—	—	—	—	1.370 ●	1.174 ●	1.811 ●	1.640 ●

Table 1: Numerical results using the test images (unseen at training time). The best altitude MAE values are given gold ● , silver ● and bronze ● medals.

Dataset description

Area index	004	068	214	260
# train/test	9/2	17/2	21/3	15/2
max alt. [m]	1	30	73	13
min alt. [m]	-24	-27	-29	-30

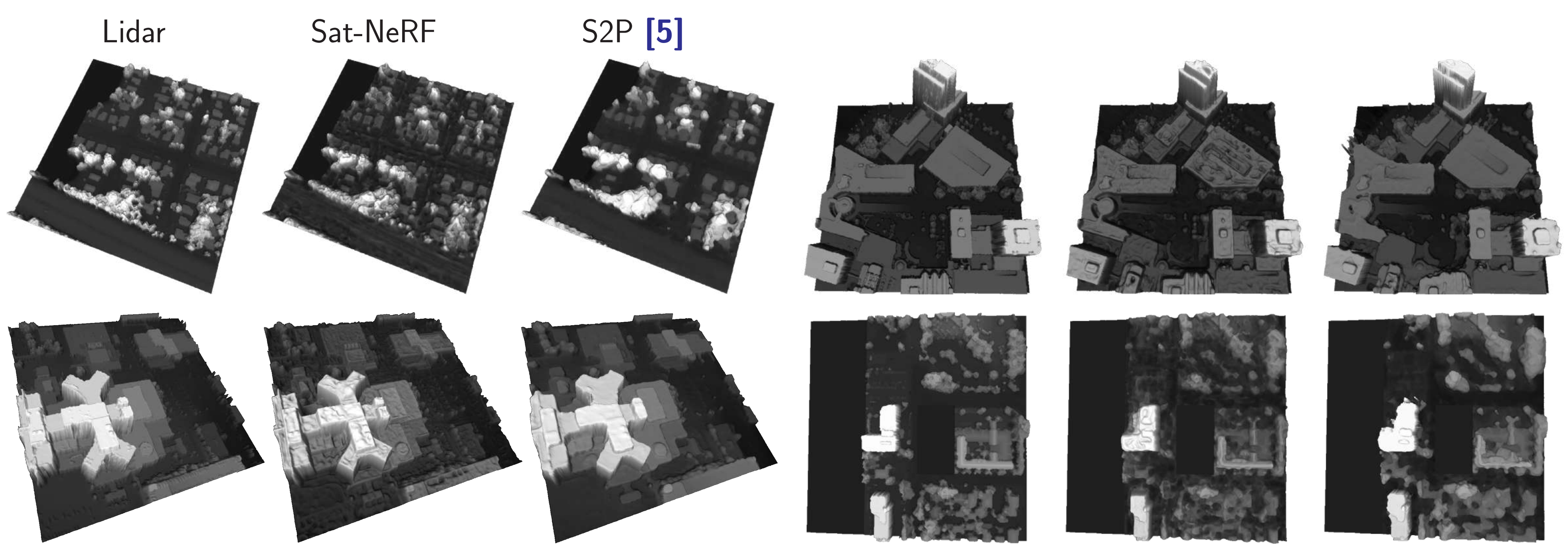


Figure 3: Compared to classic satellite MVS [5], Sat-NeRF surface models provide finer details and sharper structures but exhibit strong local irregularities.

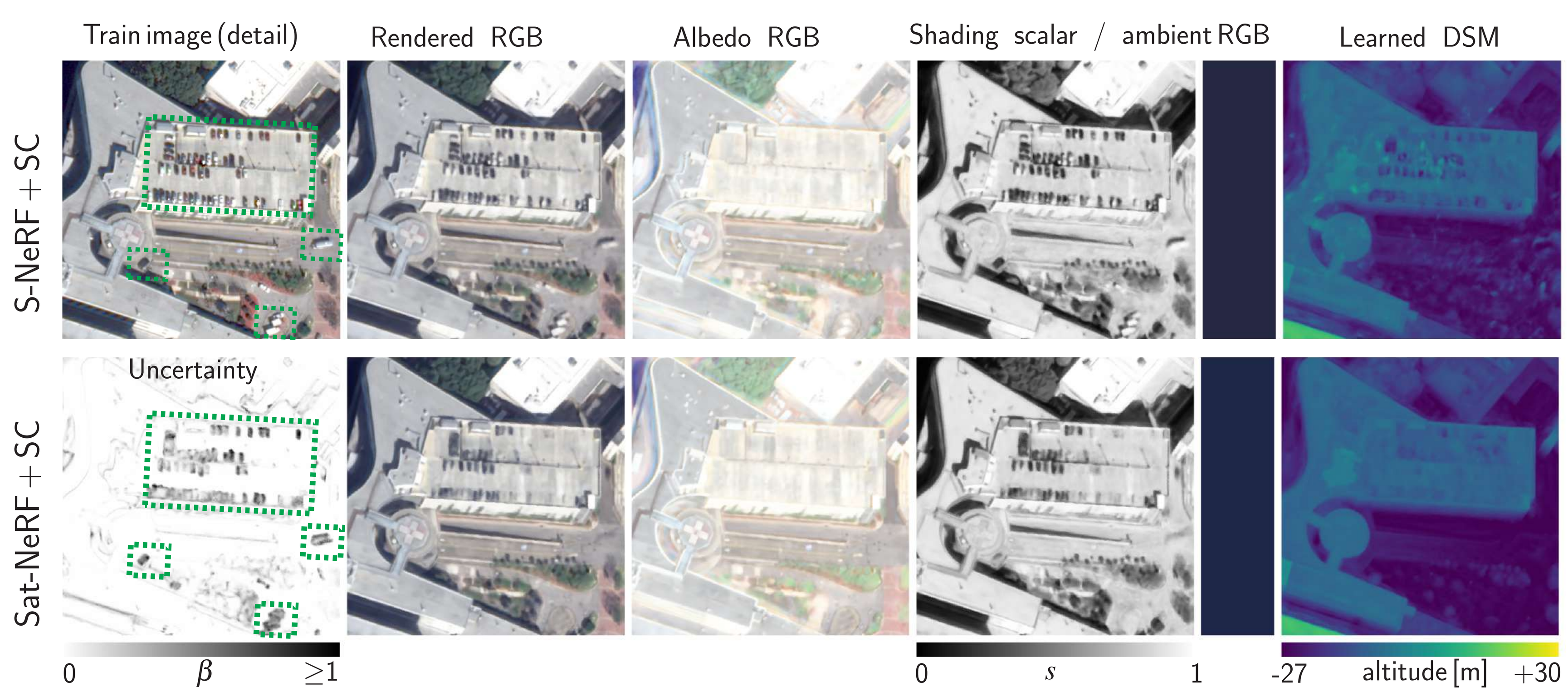


Figure 4: The uncertainty coefficient β learned by Sat-NeRF helps improve geometry learning with respect to previous NeRF variants for satellite images [3].

References

[1] Mildenhall, Ben et al. NeRF: Representing scenes as neural radiance fields for view synthesis. *ECCV*, 2020. doi.org/10.1007/978-3-030-58452-8_24

[2] Martin-Brualla, Ricardo et al. NeRF in the wild: Neural radiance fields for unconstrained photo collections. *CVPR*, 2021. doi.org/10.1109/CVPR46437.2021.00713

[3] Derksen, Dawa et al. Shadow neural radiance fields for multi-view satellite photogrammetry. *CVPR Workshops*, 2021. doi.org/10.1109/CVPRW53098.2021.00126

[4] Marí, Roger et al. A generic bundle adjustment methodology for indirect RPC model refinement of satellite imagery. *IPOL*, 2021. doi.org/10.5201/ipol.2021.352

[5] Facciolo, Gabriele et al. Automatic 3D reconstruction from multi-date satellite images. *CVPR Workshops*, 2017. doi.org/10.1109/CVPRW.2017.198