Shadow Neural Radiance Fields for 3D Earth Observation

Dawa Derksen - Postdoctoral Research Fellow

Advanced Concepts Team – ESTEC Netherlands

dawa.derksen@esa.int

- Supervision : Dario Izzo

Presentation outline

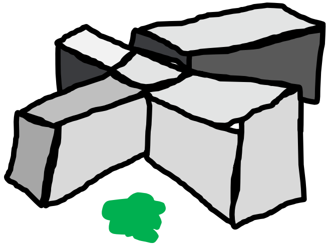
1. Introduction to Photogrammetry for Earth Observation
2. Basics of Neural Volume Rendering
3. Shadow Neural Radiance Field
4. Preliminary results
5. Conclusion and further applications

Photogrammetry for Earth Observation

Multi-view satellite images

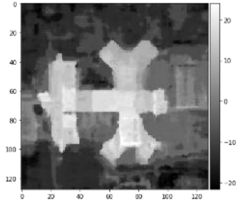
Source : WorldView-3 (0.3m RGB)

Volumetric representation



New views

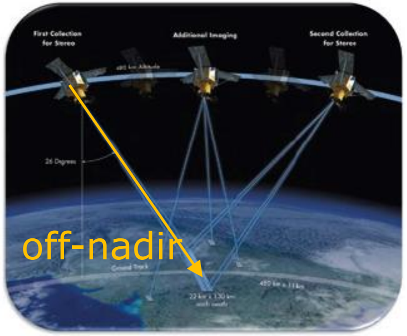
Depth (topography)



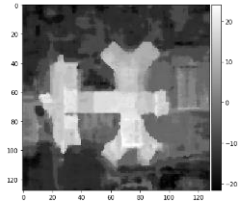
Polygonal 3d surface model

Photogrammetry for EO – Scope and applications

- Use of optical images only
  - **No active sensor** e.g. radar, lidar
    - Low energy consumption
  - Off-nadir images exhibit vertical structure



- Tomography
  - Vegetation height estimation (biomass)
  - Land cover classification
  - Change detection
    - Building destruction, deforestation



- Other domains
  - Autonomous landing, deep space exploration

Prior work (EO): Semi-Global Matching

- Mutual information based **pixel matching** [Hirschmüller 2008]
  - Requires similar capture conditions [Krauß 2018]
  - Efforts to include tolerance to non-correlation [Rupnik 2018, 2019]

- Sources of non-correlation : **time-dependent** phenomena
  - Shadows
  - Specularity – matte/glossy

  Light source (sun)

  - Transient objects – cars, airplanes
  - Seasonal variation – leaf growth/loss
  - Weather effects – snow
  - Land cover change – deforestation, construction

  Scene itself

Prior work (EO): Semi-Global Matching

- Limitations of SGM
    1. Structures with overhang (trees, monuments)
    2. Semi-opaque/transparent surfaces (glass)
    3. Pixel matching sensitive to non-correlation
- Can we benefit from non-correlation ?
    - Position/shape of shadows varies according to scene geometry
    - Slow temporal variations can indicate the presence of vegetation
    - Cars are mainly in parking lots or on roads (flat surfaces)
- "New" paradigm in 3D graphics - *Representation learning*
    - Exploit the capacity of ML algorithms to memorize data
    - Create a learning scheme with physics-inspired rules

Presentation outline

1. Introduction to photogrammetry for Earth Observation

2. Basics of Neural Volume Rendering
    1. Neural Radiance Field (NeRF)
    2. Sensor model and view synthesis

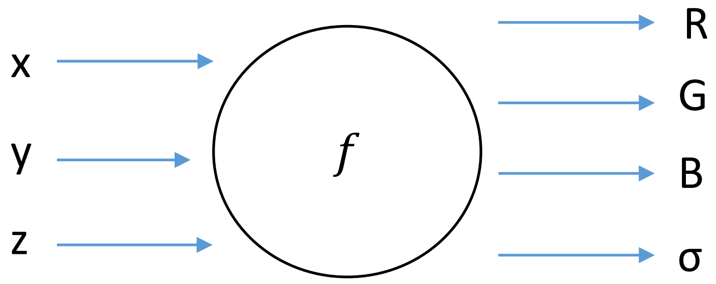
3. Shadow Neural Radiance Field

4. Preliminary results

5. Conclusion and further applications

Volumetric representations

- Model volume properties as a function of 3D space : $p = f(x, y, z)$
- p can represent different aspects of the volume
    - **Color** (R, G, B) and **opacity** (σ) for optical rendering [Max 1995]
    - Semantic class labels [Kohli 2020]
    - Other local properties : density, pressure, temperature

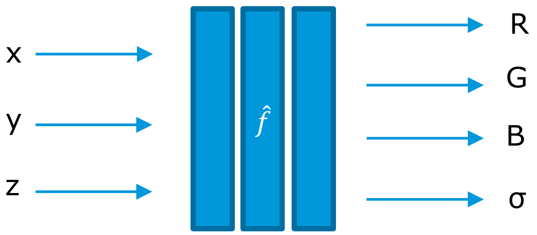


Volumetric representation for optical rendering

Neural Radiance Field (NeRF)

- **Continuous** representation using a neural network [Mildenhall 2020]

$$f(x, y, z, \dots) = (\boldsymbol{c}, \boldsymbol{\sigma}, \dots)$$

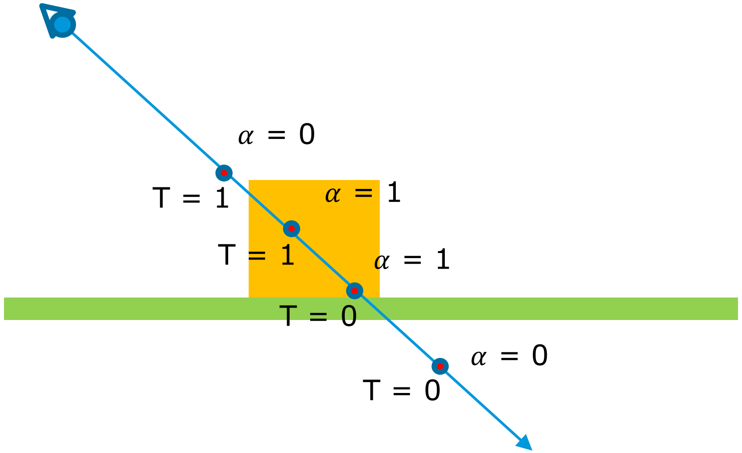


- Compact compared to voxel grid [Lombardi 2019]
  - Resolution depends on network size
- Trivial to add new dimensions (inputs)
  - Ex : Addition of viewing angles for specular effects
- Trivial to add new volume properties (outputs)
  - Ex : IR bands, albedo, class label, etc.

[*NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*, Mildenhall et al. 2020]

Emissive rendering equation

- How to extract an image from the volumetric representation ?
- $I_n = \int_{t=n}^{f} e^{-\int_n^t \sigma(x)dx} \alpha(t)c(t)dt + I_f$ [Max 1995]
- Numerically estimated by sampling (c, σ) along the ray



- $I = \sum_{i=1}^{N} T_i \alpha_i c_i$ ← colors

Fully differentiable!

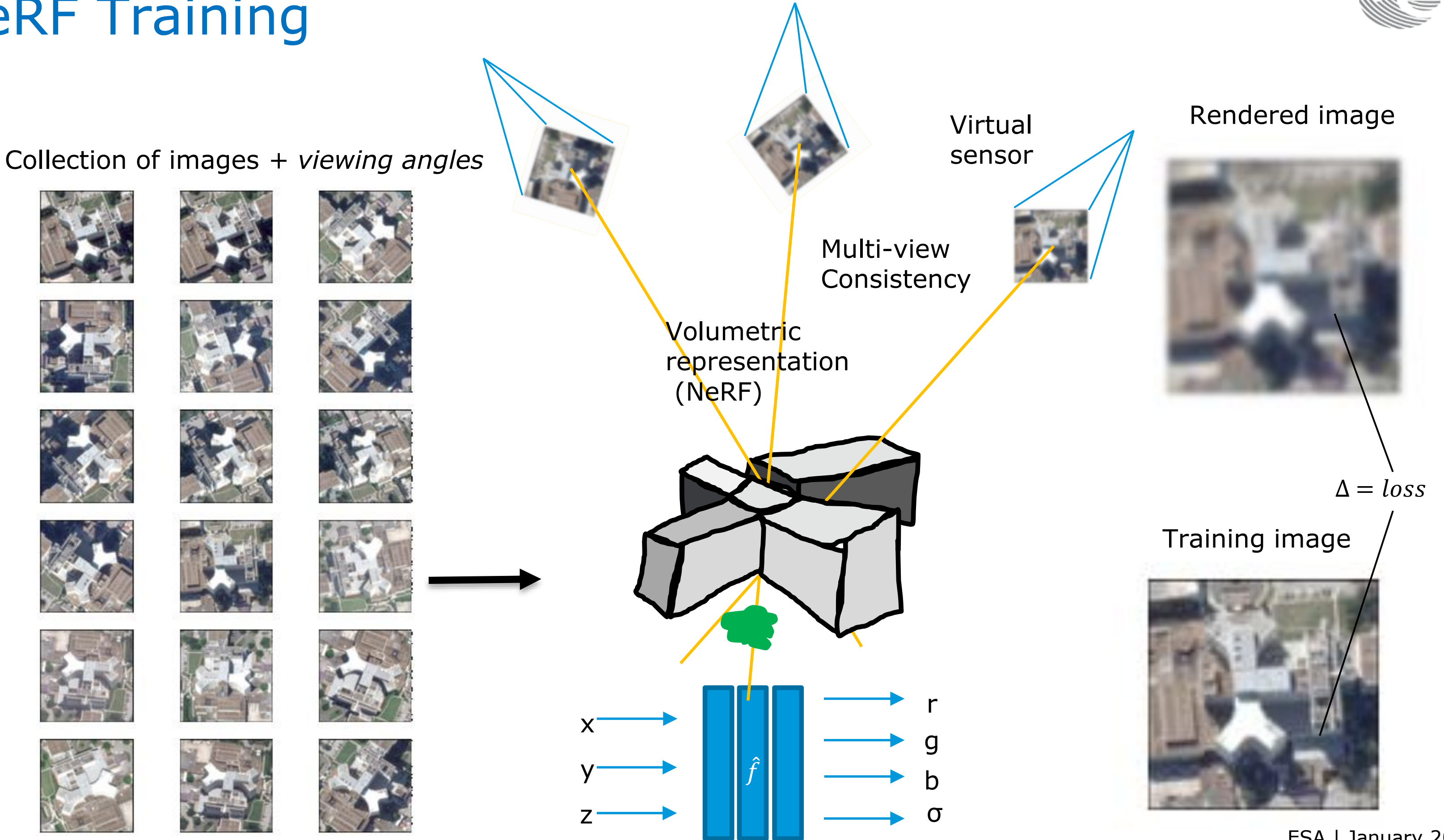
- α : How much light is contributed by ray segment i

$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$

- T : How much light is blocked earlier along the ray

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

NeRF Training

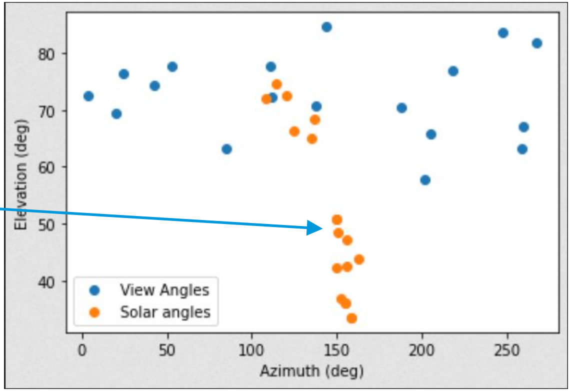


Neural Radiance Field (NeRF) – view synthesis

- About 20 training images – specular effects successfully modeled

Limitations of “off the shelf” NeRF

- Designed for images with constant lighting conditions
- Unfit for off-nadir satellite imagery
    - Solar direction varies across data set
    - Other kinds of space-based imagery
        - Comets, moons, asteroids, other satellites, etc.



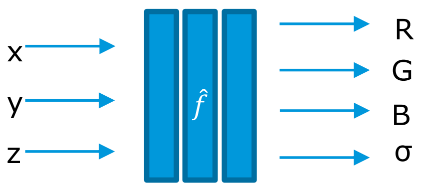
- Direct application leads to inaccuracies in shaded areas
    1. Floating particles
    2. Lenticular painting effects


Art by Randy Noborikawa

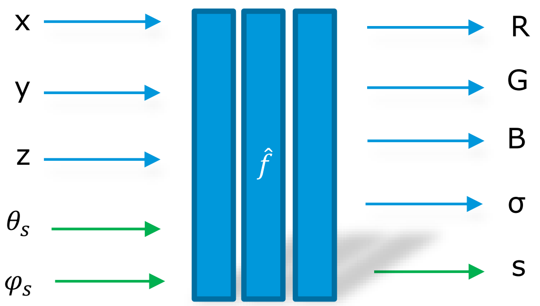
First research objectives

Propose and evaluate a **shadow-aware** NeRF model

1. Proper modeling of shaded areas (limited depth inaccuracies)
2. Detect shadows
3. Synthesize accurate views at unseen solar angles (interpolation, extrapolation)
4. Shadow removal



Standard NeRF



Shadow NeRF (SNeRF)

Presentation outline

1. Introduction to photogrammetry for Earth Observation

2. Basics of Neural Volume Rendering

3. Shadow Neural Radiance Field
    1. Directional/diffuse light source modeling
    2. Solar Correction rays

4. Preliminary results

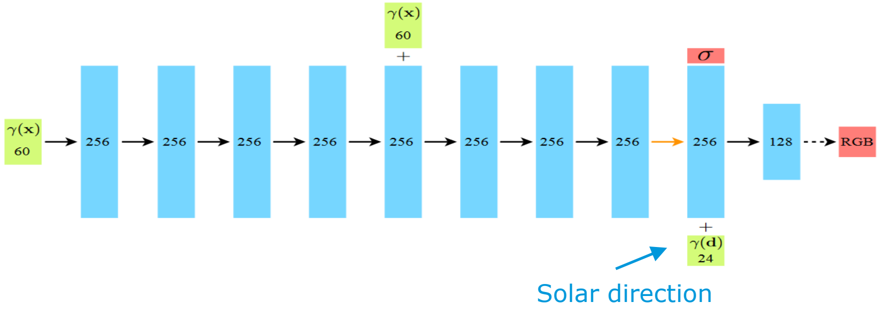
5. Conclusion and further applications

Naïve shadow model

- Make use of **solar direction** $\boldsymbol{\omega_s} = (\phi_s, \theta_s)$
  - Available as metadata for most EO imagery
- As input to the representation network
  - Similar to NeRF viewing direction for specular effects

- Result : Poor interpolation due to under-regularized color values at unseen solar directions
- *No explicit disentanglement of illumination vs. color*

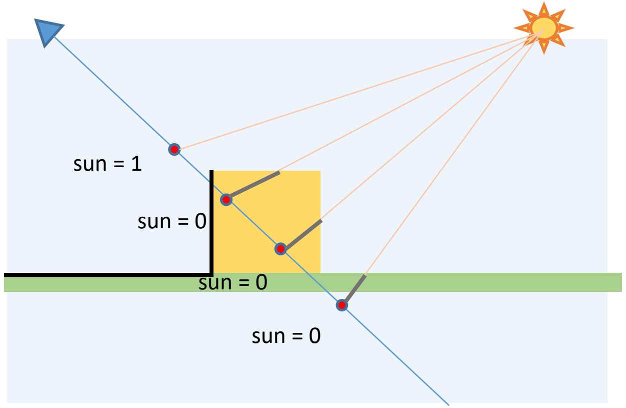
- $f(x, y, z, \boldsymbol{\omega_s}) = [c(x, y, z, \boldsymbol{\omega_s}), \sigma(x, y, z)]$



Shadow Neural Radiance Field

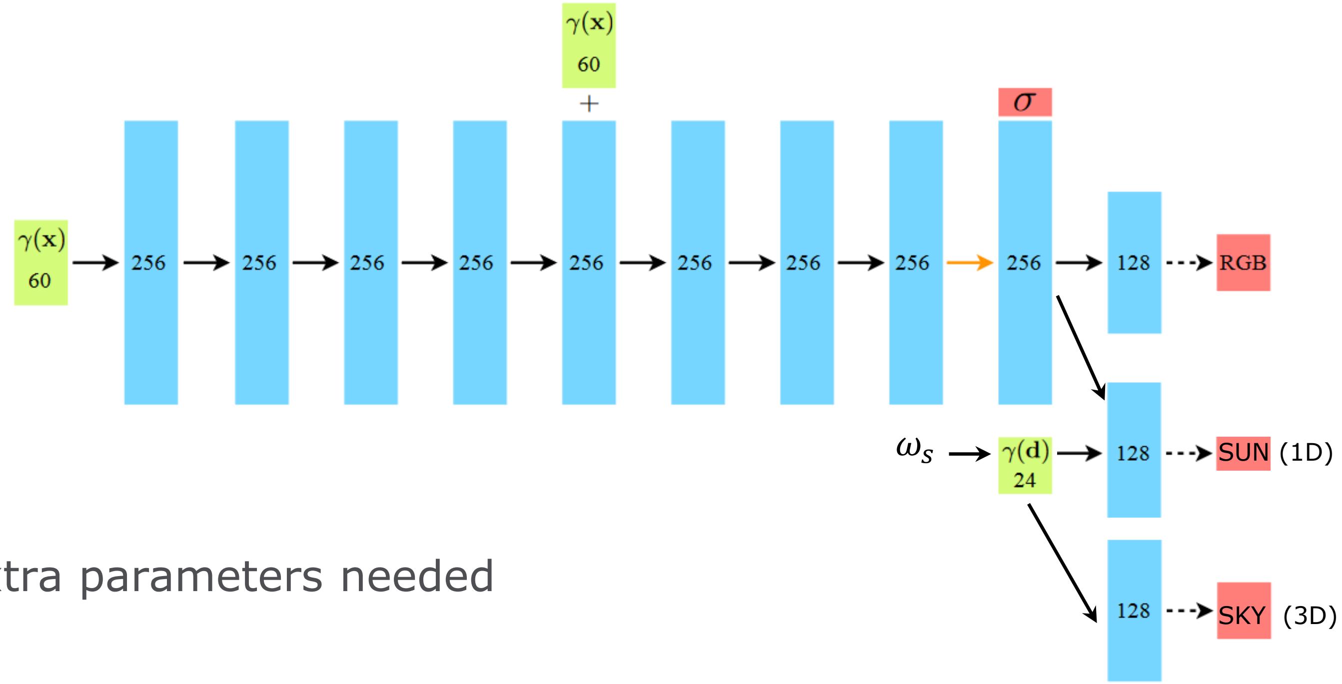
- Represent incoming illumination
- Introduce **two new physical properties** (model outputs)
  1. Ratio of incoming solar light $\mathbf{sun}(\mathrm{x}, \mathrm{y}, \mathrm{z}, \omega_s) \rightarrow [0..1]$
  2. Sky color (diffuse source) $\boldsymbol{sky}(\omega_s) \rightarrow [0..1]^3$
- New rendering function

$$I = \sum_{i=1}^{N} T_i \alpha_i c_i{}'$$



- Different "color" in rendering equation : $c_i' = c_i l_i$
- Incoming light color: $l_i = (sun\, \mathbb{1}_3 + (1 - sun) sky)$
  - Weighted sum between sky and $\mathbb{1}_3$

SNeRF network architecture



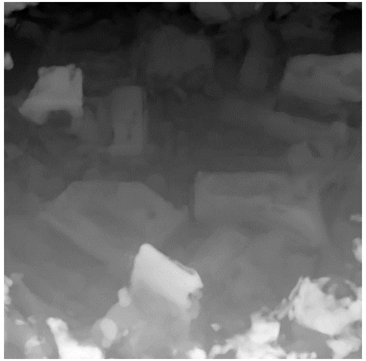
- Few extra parameters needed

First result and limitation

- Sun function entirely learned from *non-correlation* in the data set
  - Areas that are always in the shadow have a dark RGB color and are modeled as always in the sun
  - World-view 3 data only has morning pictures : certain building facades are always in the shadow

Training images

Synthetic images



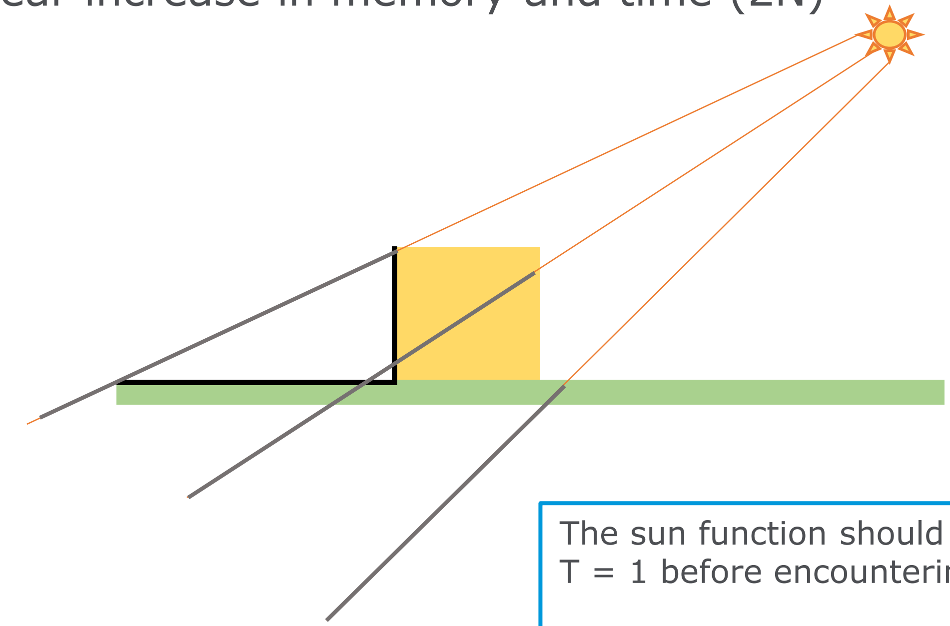
3d surface

Sun function at surface

- Could be improved using the learned scene geometry

Solar correction rays

- Make the network learn the shadows based on the scene geometry
- Linear increase in memory and time (2N)

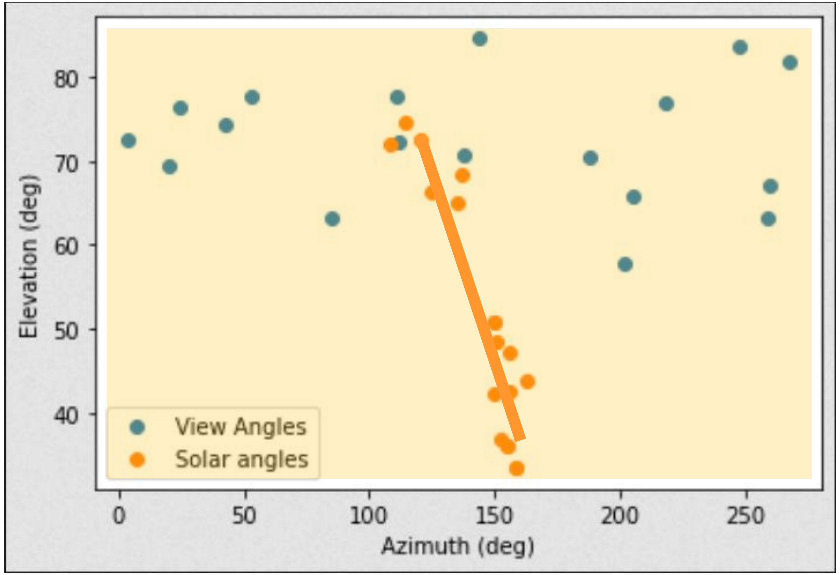


The sun function should be equal to T along a solar correction ray.
T = 1 before encountering a surface, 0 after.

$$L_s = \frac{1}{N}\sum_{i=1}^{N}(T_i - sun_i)^2$$

Solar correction rays

- Which solar angles to consider for ray generation ?
  1. Training image angles
     - Achieve high fidelity on training images
  2. Any other angles !
     1. Sampled to fill in the gaps between training angles
     2. Sampled regularly on the entire upper hemisphere



Overview of changes to NeRF volume rendering

1. 2 new dimensions : *solar direction* as network input
    - Requires knowledge of incoming light conditions

2. 2 new physical properties : *sun* and *sky* as network outputs
    - Small extra cost in memory (larger network)

3. Shadow-aware rendering
    - Outgoing light depends on color of incoming light (*sun* and *sky*)
    - No extra cost at inference time
4. Solar correction rays
    - Learn *sun* based on scene geometry
    - Linear increase in memory and time (2N) during training

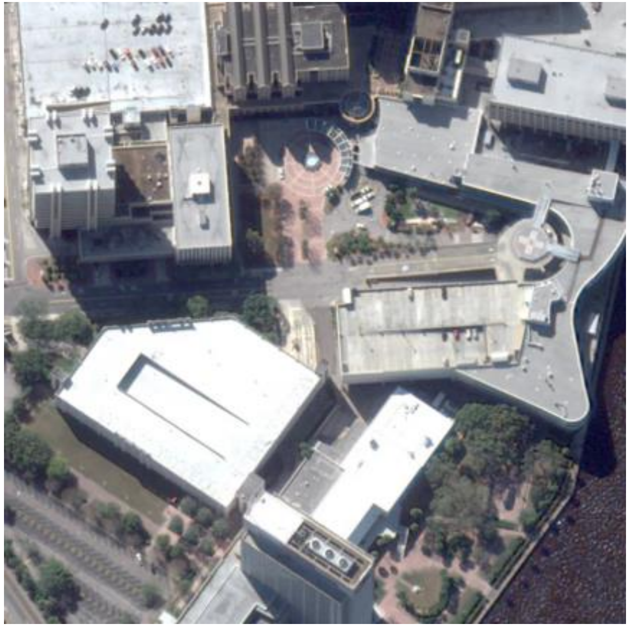
Presentation outline

1. Introduction to photogrammetry for Earth Observation
2. Basics of Neural Volume Rendering
3. Shadow Neural Radiance Field
4. Preliminary results
    1. Solar correction rays
    2. Different areas
5. Conclusion and further applications

Experimental setup

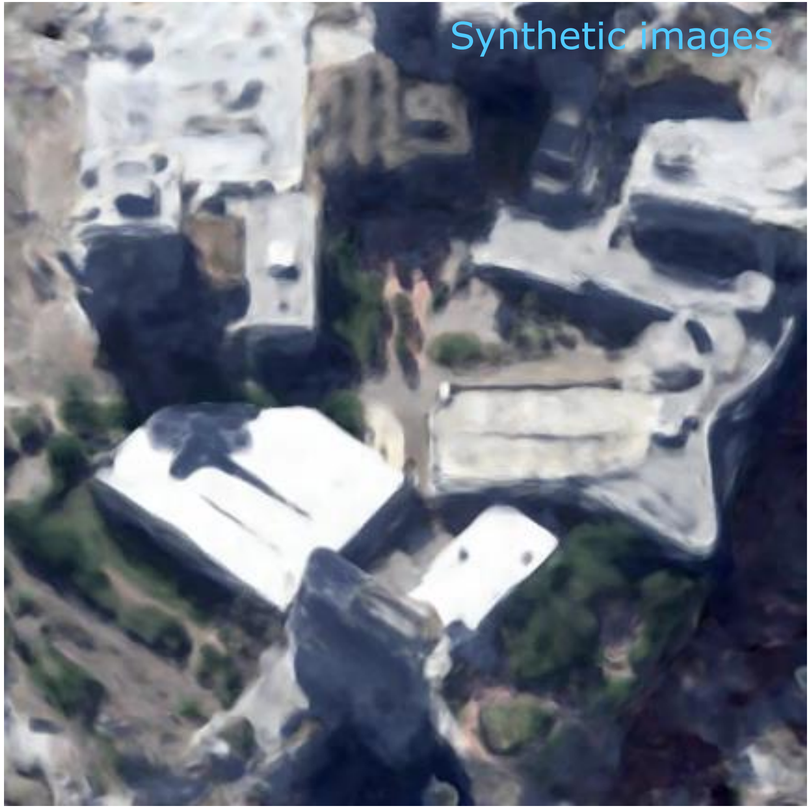
- 4 areas of 300m x 300m
  - Worldview-3 [IEEE GRSS Data Fusion contest 2019]
- 512x512 pixel images at 0.6m
  - 10-20 images / area
- Network size = 800k parameters
  - One network / area
- Hardware : 1 GPU with 12 GB ram
  - ≈12h training time / area

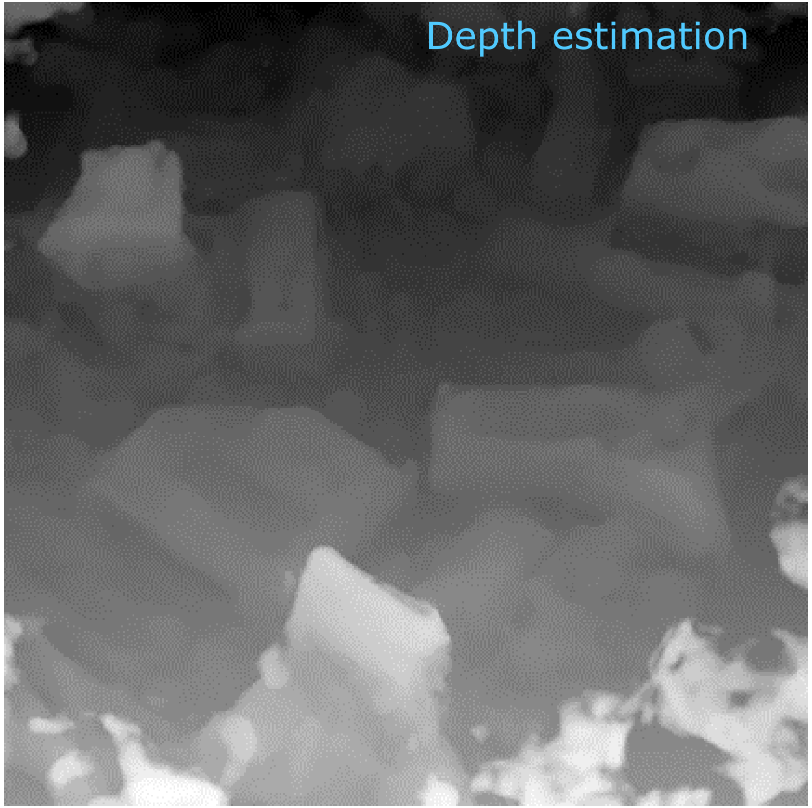
Training images - area 1



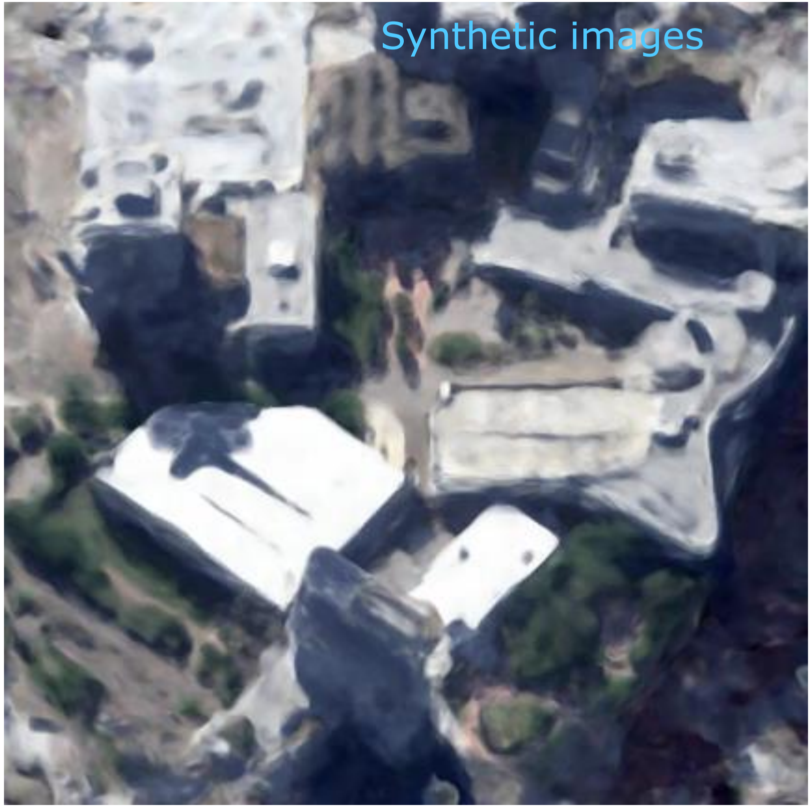
http://www.grss-ieee.org/community/technical-committees/data-fusion/2019-ieee-grss-data-fusion-contest/

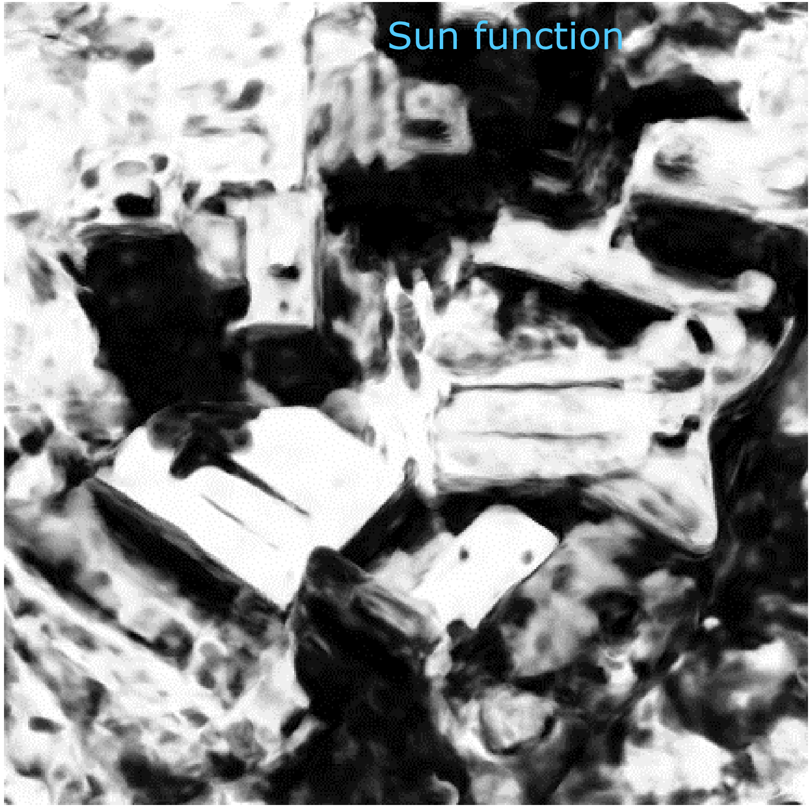
Depth estimation



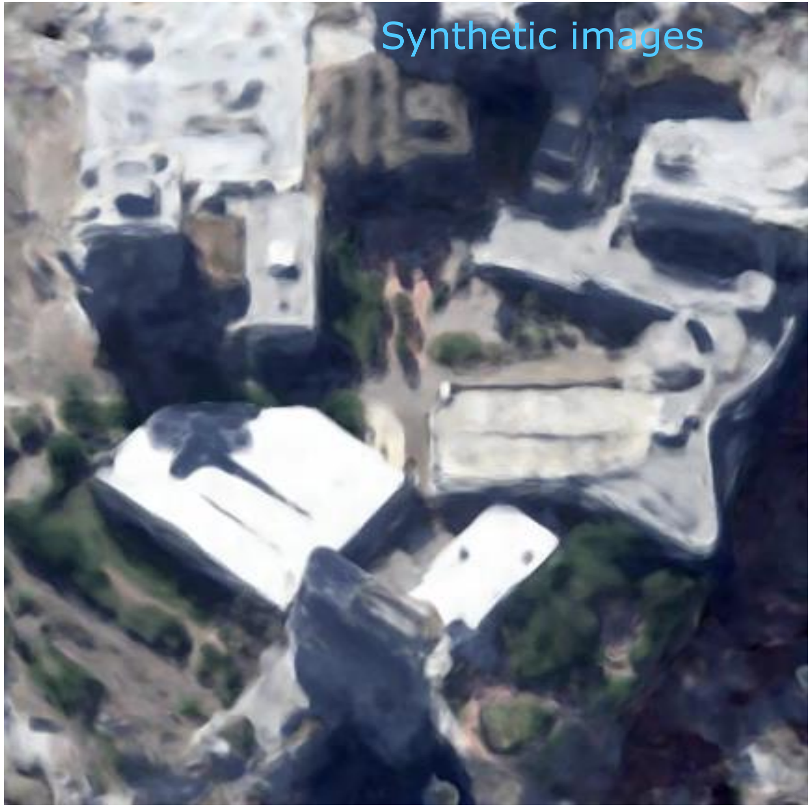


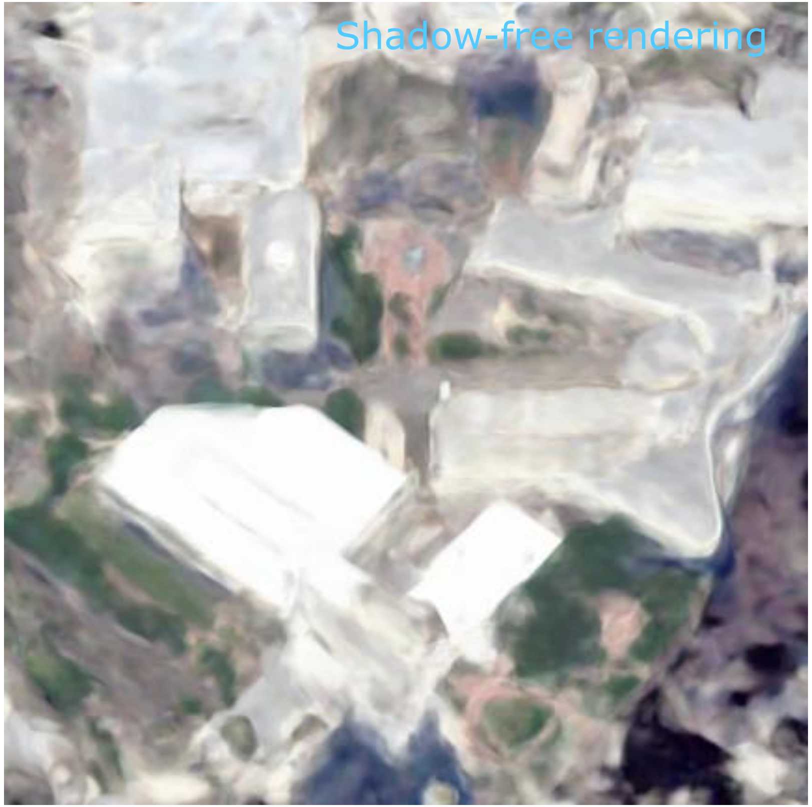
Rendered sun function - shadow detection





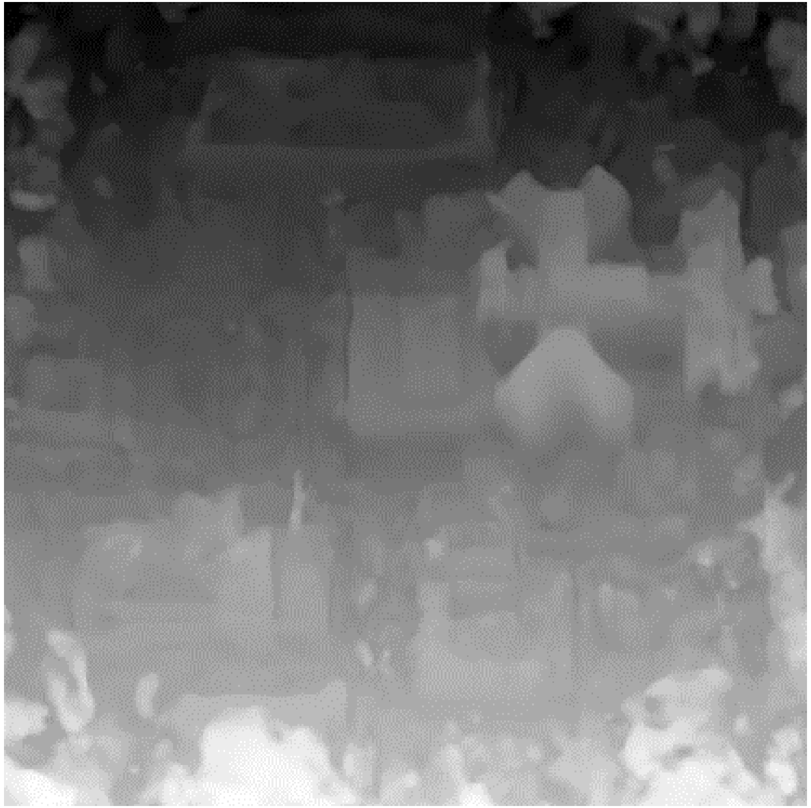
Shadow-free rendering



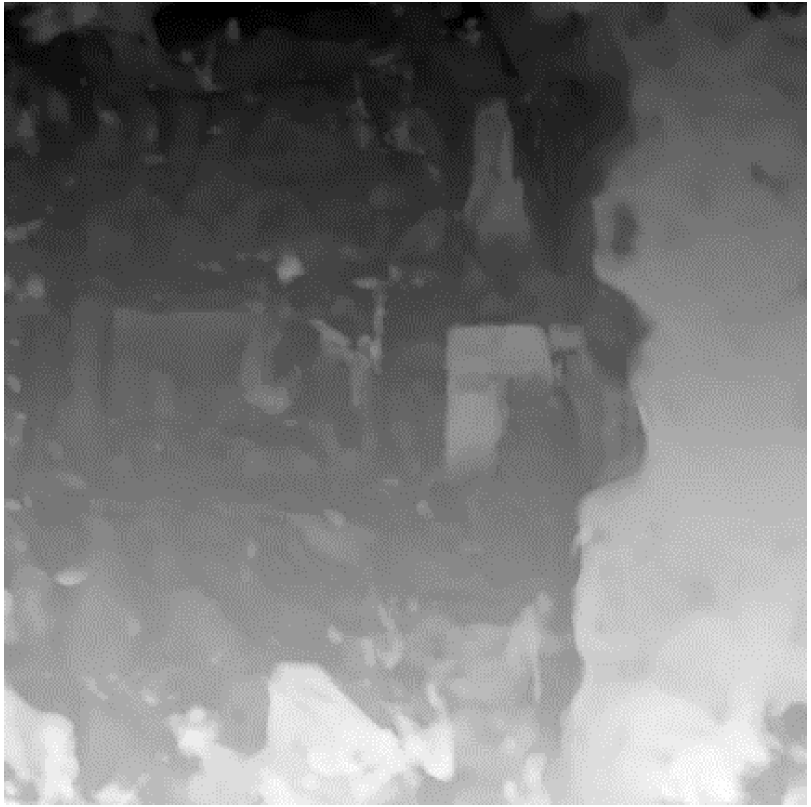


Other areas – diffuse urban cover

Other areas – Industrial / commercial area



Other areas – Coastal diffuse urban



Quantitative evaluation

- Average altitude estimation error (RMSE) = 3-5 m (w.r.t airborne LiDAR)
- Average train image PSNR = 25-30
- Average test image PSNR = 20-25

	NeRF	**SNeRF + SC**
Test PSNR	20.27	**23.56**
Test SSIM	0.7589	**0.897**
Av. Alt. err (m)	6.971	**4.361**

- More complete evaluation (60 areas) in upcoming publication

Presentation outline

1. Basics of Neural Volume Rendering
2. Shadow Neural Radiance Field
3. Preliminary results
4. Conclusion and further applications
    1. Overview
    2. Current limitations
    3. Outlook

Overview of advantages

- Advantages of SNeRF compared to SGM
  1. No need for pairs/triplets in stereo configuration
     - Any number of images
     - Potentially different resolutions and spectral bands
  2. Shadows are a part of the model
     - Act as a source of information rather than a perturbation
     - Can be detected and even removed

- If successful, could alleviate constraints for off-nadir EO satellites
  - No more need for stereo configuration with similar capture conditions
  - More diverse viewing angles and shadow positions are better

Current limitations

1. Need to train one model for each area
    - Not yet scalable to very large areas
    - Possible generalization using *hypernetworks* [Stitzmann 2020] or *network conditioning* [Trevithick 2020], demonstrated on computer vision tasks
    - Learn from commonalities between scenes
2. Poor performance on flat, texture-free surfaces
    - Water, buildings
    - Unable to learn if there are no visual cues
3. Perturbed by other time-dependent effects
    - NeRF in the Wild [Martin-Brualla 2020]
    - *Conditional representation* can distinguish *static* from *time-dependent* elements

Possible extensions

1. Include viewing angles to account for specular effects (full NeRF)
    - Make color $\boldsymbol{c}$ depend on viewing angles
    - Will this interfere with the shadow model ?

2. Multiple light sources: small extra cost
    - Model one *sun function* for each directional light source
    - Model one *sky function* for each diffuse light source
    - Render a weighted sum with more than 2 elements

3. Multispectral images
    - Infrared bands bring more information for vegetation and water
    - Lower resolution than RGB images: fusion of sources

Outlook

- Implicit representation learning is a form of self-supervised learning
    - Paradigm where the images themselves are the training data
    - Physics-inspired rendering and modeling: more than just memorizing images
    - Use of ANN as core model allows for back-propagation through the integration

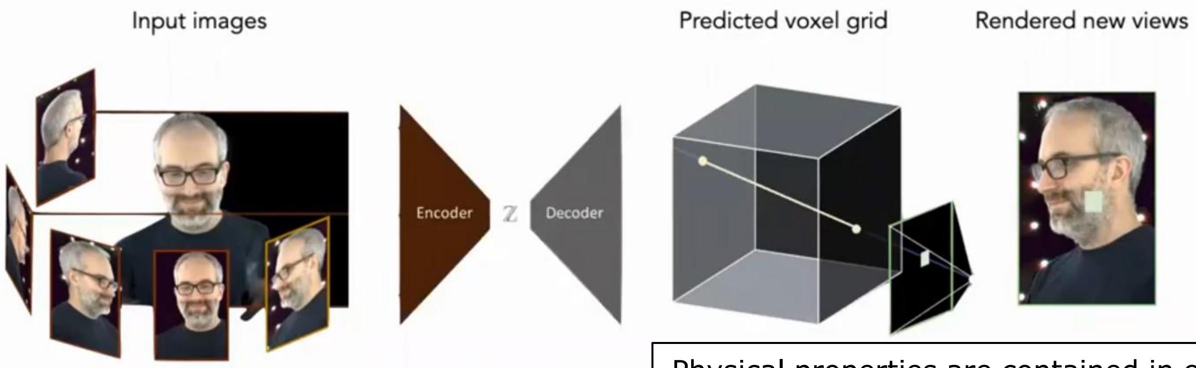
- More compact than image representation
    - Image set = $18 \times 512 \times 512 \times 3 \approx 14M$ parameters (floats)
    - SNeRF $\approx 1M$ parameters

- Other research in implicit representation learning for space ongoing at ESA
    - NeRF/SNeRF on images of satellites for debris identification/capture
    - GravANN – estimating the inner density of celestial bodies

Bibliography

- [Hirschmüller 2008] Stereo Processing by Semi-global Matching and Mutual Information - *International Symposium on Visual Computing, 2008*
- [Rupnik 2018] 3D reconstruction from multi-view VHR satellite images in MicMac - *ISPRS Journal of Photogrammetry and Remote Sensing*, *139*, 201-211.
- [Krauß 2018] Cross-track satellite stereo for 3D modelling of urban areas - *European Journal of Remote Sensing (2018)*
- [Rupnik 2019] More Surface Detail with One-Two Pixel Matching - *Conférence Française de Photogrammétrie et de Télédétection 2018.*
- [Max 1995] Optical models for direct volume rendering - *IEEE Transactions on Visualization and Computer Graphics* 1.2 (1995): 99-108.
- [Kohli 2020] Inferring Semantic Information with 3D Neural Scene Representations. *arXiv preprint arXiv:2003.12673* (2020).
- [Lombardi 2019] Neural Volumes: Learning Dynamic Renderable Volumes from Images. *arXiv preprint arXiv:1906.07751* (2019).
- [Mildenhall 2020] NeRF : Representing scenes as Neural Radiance Fields for view synthesis. *arXiv preprint arXiv:2003.08934* (2020).
- [Stitzmann 2020] Implicit Neural Representations with Periodic Activation Functions. *Advances in Neural Information Processing Systems* 33 (2020).
- [Trevithick 2020] GRF: Learning a General Radiance Field for 3D Scene Representation and Rendering. *arXiv preprint arXiv:2010.04595, (2020)*
- [Martin-Brualla 2020] NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. *arXiv preprint arXiv:2008.02268* (2020).

Prior work (vision) : discrete representations



Physical properties are contained in each voxel
Ex : Color, opacity

- Neural Volumes [Lombardi 2019]
- A CNN is trained to generate a voxel representation given a set of images
  - Outputs (voxel grids) require high storage capacity
  - High redundancy in discrete representation