SENSOR-ROBUST DYNAMIC NEURAL RADIANCE FIELDS:
SYNTHETIZING NEW VIEWS FROM 100 YEARS OF ARCHIVAL AERIAL IMAGES

Laboratory: LASTIG, Univ Gustave Eiffel, IGN-ENSG (STRUDEL and GeoVIS teams)

Localisation: IGN, Saint Mandé, France

Supervision: Loic Landrieu, PhD; Mathieu Brédif, PhD

Remuneration: 513 euros / month

Starting Date: April 2022, up to 6 month duration

Key Words: Deep Learning, Neural Radiance Fields, Inverse Problems, New View Synthesis, Digital Heritage

Development Environment: Linux, Python, PyTorch.

Context

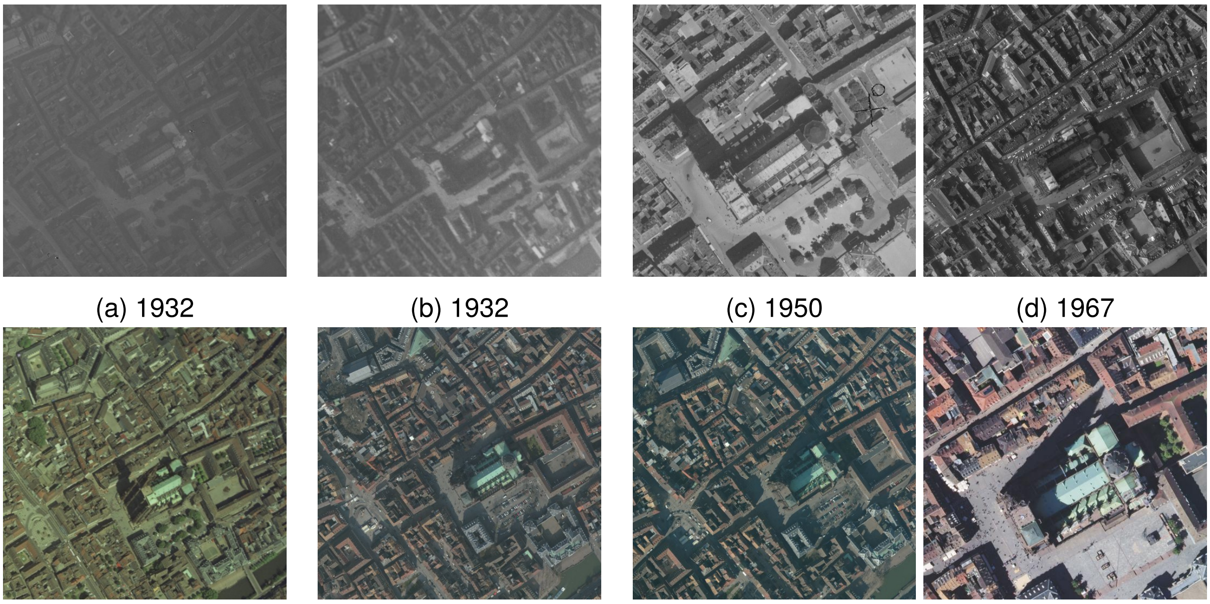
IGN is the public institution in charge of the production and distribution of geographical information in France. LASTIG, the research lab associated with IGN, has privileged access to over a hundred years of archival aerial data. These images have been taken across the French territory, with various sensors as photography technology evolved: analog, digital, black and white, RGB and infrared, etc. The images have already been geographically aligned and are accessible through the open-data platform: See examples of aligned images at `https://remonterletemps.ign.fr/telecharger`. Some regions have been photographed nearly every year for the last century, and landmarks are visible at different angles and with varying qualities.

Neural Radiance Fields (NeRFs) are a recent development in view synthesis [2], and are capable of generating novel views of objects from a set of images at various poses, without any 3D supervision. This work has been adapted to reconstruct views of city scenes from off-nadir satellite images by explicitly taking into account the sun illumination [1]. However, all these works make several hypotheses that are not true in our case: (i) the images come from identical sensors (ii) the scenes are static, except for transient objects, such as cars.

Recent work has proposed adding a temporal dimension to NeRFs [4, 3], allowing NeRFs to encode dynamic scenes. However, their formulation is ill-suited to capture the multiple temporal dynamics in century-long image time series (daylight, seasonal, transient objects, urban change). Adapting NeRFs to a dynamic, multi-sensor setting would allow us to recreate the changing landscape of the French territory across the last century as free-viewpoint videos.

Objective

This internship aims to adapt the current NeRF-based methods to temporal sequences of aligned aerial images taken with various sensors and across the last century. This poses several challenges:



(a) 1932 (b) 1932 (c) 1950 (d) 1967

(e) 1980 (f) 2007 (g) 2007 (h) 2021

Figure 1: **Century-Spanning Image Time Series.** Sample image crops from 7 time steps out of more than 60 available time steps. (a) and (b) are different time steps of the same year (various sensors, image qualities, and acquisition dates). Each time step exhibits multi-view images, as illustrated by (f) and (g): same date, same sensor, 4 to 9 views in general for each ground point.

Sensor Robustness as an Inverse Problem: The last century saw dramatic technological advances in photography. As a consequence, the radiometric quality varies a lot across images. We propose to select a given acquisition as pivot modality (ideally RGB, high resolution, low distortion, consistent illumination) and to learn transformations from this pivot modality to the different acquisition dates. This can entail lowering the resolution, converting to monochrome, adding radial distortion, and so on. The pixelwise radiometric supervision of the NeRF will then take place through this transformation (*i.e.*, an inverse problem setting). In practice, we denote by I a ground truth image acquired at a time T, potentially of low quality; J the render of the NeRF with I's extent and pose in the high-quality pivot modality; and ϕ_T an operator transforming the pivot modality to the radiometric/sensor setting of acquisition time T, the loss writes:

$$\mathcal{L}(J) = \|I - \phi_T(J)\|^2 \,. \tag{1}$$

Multi-Scale Dynamics: Sequences of images spanning a century are subject to several temporal dynamics: (i) the time of day/ day of the year are hugely influential through the illumination conditions, but also seasonal changes (vegetation, snow). The proposed approach must be able to model this influence. (ii) Transient objects such as cars or pedestrians can appear in some images. These objects should not be rendered. (iii) The urban landscape evolves slowly: vegetalisation, densification, and so on. A carefully designed dynamic NeRF would be able to capture these changes.

Results: Provided that all challenges can be solved, we would be able to train dynamic NeRFs from sequences of historical images, allowing us to generate free view-point high-quality RGB videos retracing the evolution of urban landscapes across the century. If the internship is successful, we will write an article on the subject and release both code and datasets in open access.

The tentative planning of the internship is as follows:

- **Month 1-2.** Bibliography on NeRFs; curation of a set of stable scenes; tuning of transformation operators; training sensor-robust NeRFs with the inverse problem formulation.
- **Month 3-4.** Curation of a set of dynamic scenes; designing dynamic NeRFs able to capture urban dynamics.
- **Month 5-6.** Rendering of complex scenes; curation of a larger dataset; extensive comparison with baseline methods (*e.g.*, comparison with ground truth height from LiDAR); writing the article.

Opportunity. The intern will have a privileged opportunity to postulate to LASTIG's Ph.D. offers.

Profile

- Master 2 student in computer science, applied mathematics, or remote sensing.
- Familiarity with computer vision, machine learning, and deep learning.
- Mastery of Python, familiarity with PyTorch;
- Curiosity, rigor, motivation;
- (Optional) Familiarity with (differential) renderings;
- (Optional) Experienced with aerial/satellite images or image time series.

Contact

Send a CV and a short letter of purpose (∼20 lines max) stating your interest in this internship and the relevance of your experience to `loic.landrieu@ign.fr` and `mathieu.bredif@ign.fr`.

References

[1] Dawa Derksen and Dario Izzo. Shadow neural radiance fields for multi-view satellite photogrammetry. In *CVPR Workshop*, 2021.

[2] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[3] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *CVPR*, 2021.

[4] Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. Space-time neural irradiance fields for free-viewpoint video. In *CVPR*, 2021.

-