The World Map Hidden in Latent Space: Emergent Geographical Reasoning in SSL

Laboratory: IMAGINE/LIGM, ENPC

Location: Marne-la-Vallée, Champs-sur-Marne

Advisors: Loic Landrieu (ENPC, PhD. HDR) and Nicolas Dufour (Kyutai, PostDoc).

Remuneration: 1500 euros gross

Starting Date: 1st semester of 2026, 5 or 6 months duration

Keywords: Geolocation, Self-supervised Learning, Emerging Properties, Platonic Representation Hypothesis

Development Environment: Linux, Python, PyTorch.

Global Image Geolocation.

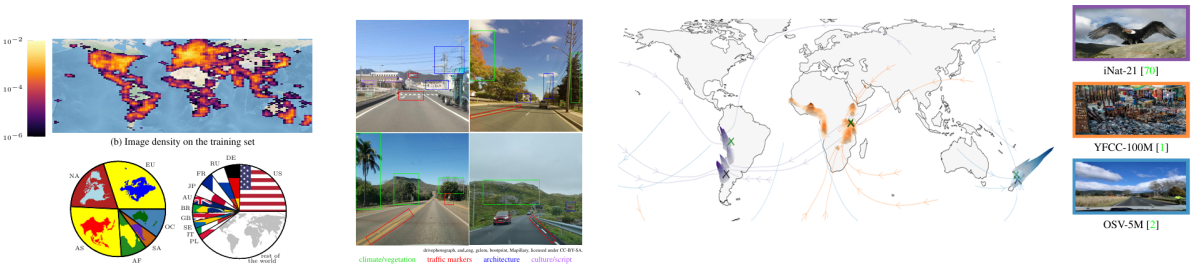
Image geolocation has recently emerged as a challenging benchmark for evaluating the reasoning abilities of Large Language Models (LLMs) and Vision–Language Models (VLMs) [4, 3, 1]. Successfully geolocating an image requires extracting a variety of subtle visual cues (vegetation, road-sign typography, architectural styles, climate indicators) and integrating them with broad knowledge of the world. Strikingly, these models often perform remarkably well on this task despite not being explicitly trained with geolocation supervision.

Our preliminary experiments suggest that their latent visual representations correlate with geographic location. In other words, latent spaces appear to be organized in a geographically coherent and surprisingly consistent way. This observation raises several fundamental questions about the nature and structure of learned representations.

Does geographical reasoning emerge naturally from large-scale visual self-supervision? Is this a concrete instance of convergence toward a shared statistical model of the world [5]?

Is the emergence of geographically meaningful latent structure simply an efficient outcome of large-scale self-supervized learning? What does this reveal about the inductive biases and internal mechanisms of modern foundation models, and can these insights be exploited to improve the accuracy and interpretability of geolocation systems? This internship aims to explore these questions both empirically and theoretically.

To learn more about the geolocation task and our previous work on the topic you can check this Underscore_ episode: https://www.youtube.com/watch?v=s5oHvfFUsbE&t=24s.



(a) OpenStreetView5M (b) Visual Cues (c) Geolocation Ambiguity

Figure 1: **Image Geolocation Models.** a: The OpenStreetView5M dataset [3] provides large-scale, globally distributed imagery for training and benchmarking geolocation systems. b: Geolocation relies on combining diverse and often subtle visual cues present in the environment. c: Recent diffusion-based approaches model geolocation ambiguity and output probability maps rather than single-point predictions [4].

Objective

This internship is part of a broader research effort on the emergence of geographical understanding in large-scale visual models. We outline four main research directions:

- **Characterizing the geographical organization of latent spaces in large models.** For instance, using methods such as sparse autoencoders [2] to identify and disentangle geographically meaningful factors.
- **Studying whether textual or visual chains of thought can reveal or explain these latent structures.** Conversely, can such analyses improve the interpretability and transparency of geographical reasoning?
- **Investigating whether geographical reasoning can be encouraged or amplified.** For example, by granting models access to explicit geographic resources (e.g., querying maps or spatial databases).
- **Exploring in-context learning for image and video geolocation.** Given that geolocation capabilities seem to emerge naturally, can this property be leveraged within diffusion models [6, 7] to address geolocation tasks in a zero-shot or few-shot setting?

While the internship will primarily focus on the first question, it will contribute to a broader, long-term research agenda. **A successful internship may lead to the opportunity to continue this work as a PhD thesis.**

Requested Profile

- Student in Master 2 in computer science, applied mathematics, or other relevant courses;
- Familiarity with machine learning and computer vision concepts;
- Experienced with Python and familiar with PyTorch;
- Curiosity, rigor, scientific reasoning, critical thinking;
- (Optional) Experienced with LLMs, VLMs, Image/Video diffusion models;

- (Optional) Good level of written English:
- (Optional) Prior experience running experiments on a GPU cluster on a distributed settings (DDP, FSDP).
- **(Bonus) Plays GeoGuessr!**

Contact

Send a CV, a link to the project you are most proud of (PDF of a report and, if possible, a link to a Github), your latest grade reports, and a short statement of purpose (∼10 lines) explaining your interest for this internship to the following adresses: loic.landrieu@enpc.fr and nicolas.dufour@enpc.fr.

References

[1] GeoArena: An open platform for bench-marking large vision-language models on worldwide image geolocalization.

[2] Anthropic. Towards monosemanticity: Decomposing language models with dictionary learning.

[3] Guillaume Astruc, Nicolas Dufour, Ioannis Siglidis, Constantin Aronssohn, Nacim Bouia, Stephanie Fu, Romain Loiseau, Van Nguyen Nguyen, Charles Raude, Elliot Vincent, and Loic Landrieu. OpenStreetView-5M: The many roads to global visual geolocation. In *CVPR*, 2024.

[4] Nicolas Dufour, Vicky Kalogeiton, David Picard, and Loic Landrieu. Around the world in 80 timesteps: A generative approach to global visual geolocation. In *CVPR*, 2025.

[5] Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. The Platonic representation hypothesis. *ICML position paper*, 2024.

[6] Black Forest Labs et al. FLUX. 1 Kontext: Flow matching for in-context image generation and editing in latent space. *arxiv*, 2025.

[7] Team Wan et al. WAN: open and advanced large-scale video generative models. *arxiv*, 2025.