PIGEON

Lukas Haas, Michal Skreta, Silas Alberti, Chelsea Finn
CVPR 2024

Bastien DOIGNIES - GDL. LIRIS - 15/01/2025

Geoguessr

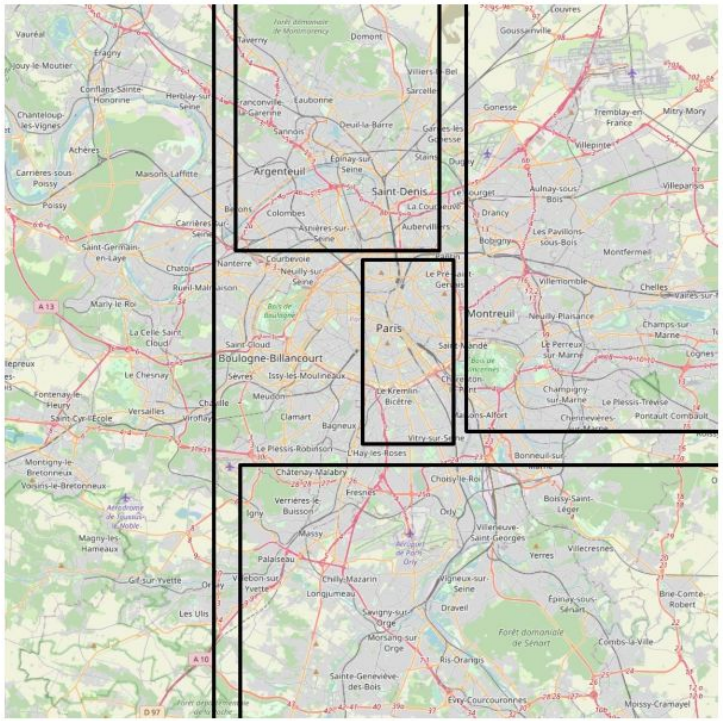

GEOGUESSR
EXPLORE THE WORLD!
Get dropped anywhere from the busy streets of New York to the beautiful beaches of Bali.
Join 85 million players today!

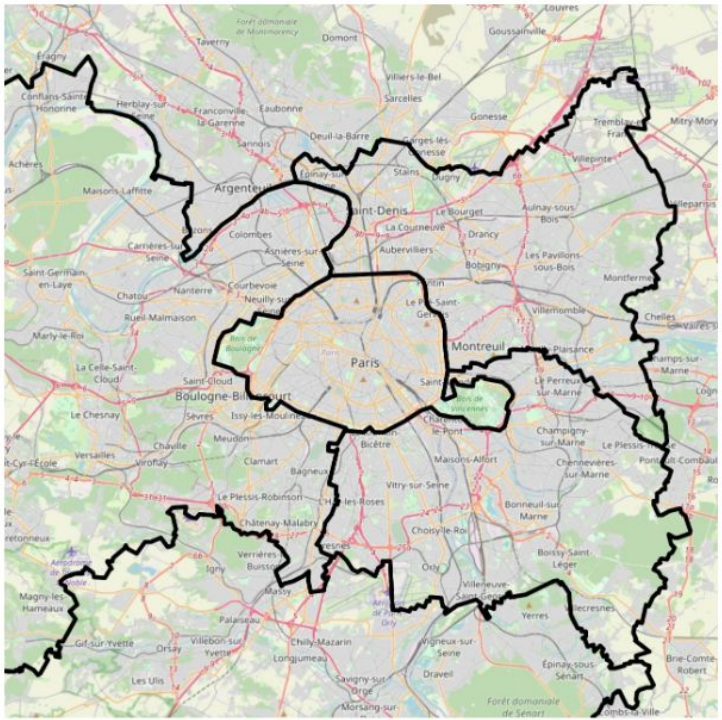
Gameplay ?

https://www.geoguessr.com/

Un pro :

Comment trouver une géolocalisation ?





Problème de classification : Geocells

Comment trouver une géolocalisation ?

Country

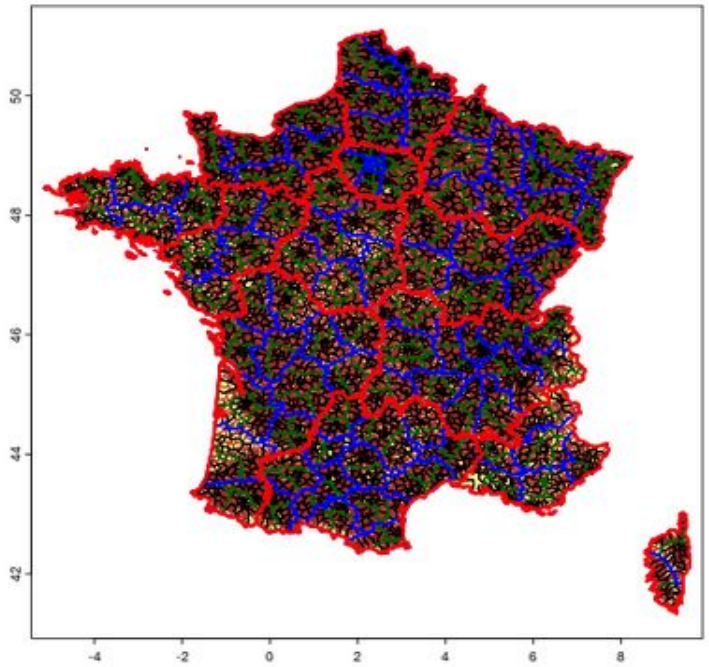
France

Geopackage

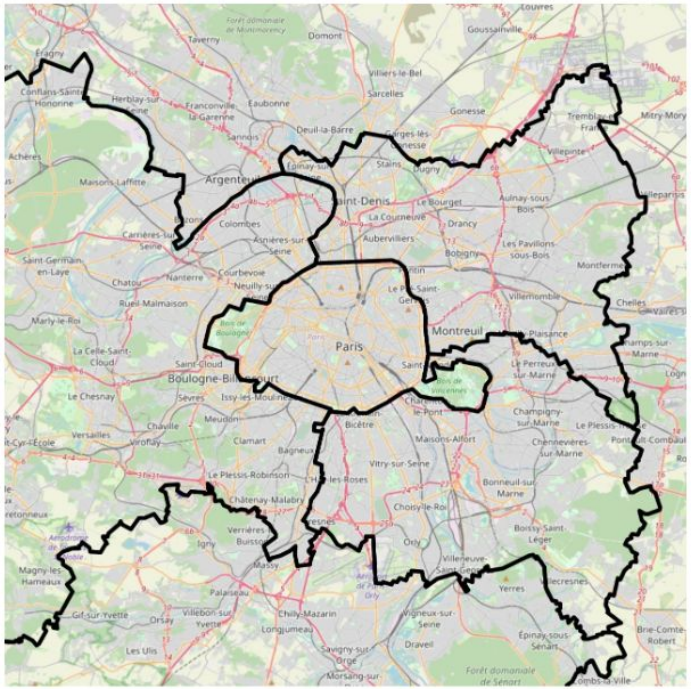
Shapefile

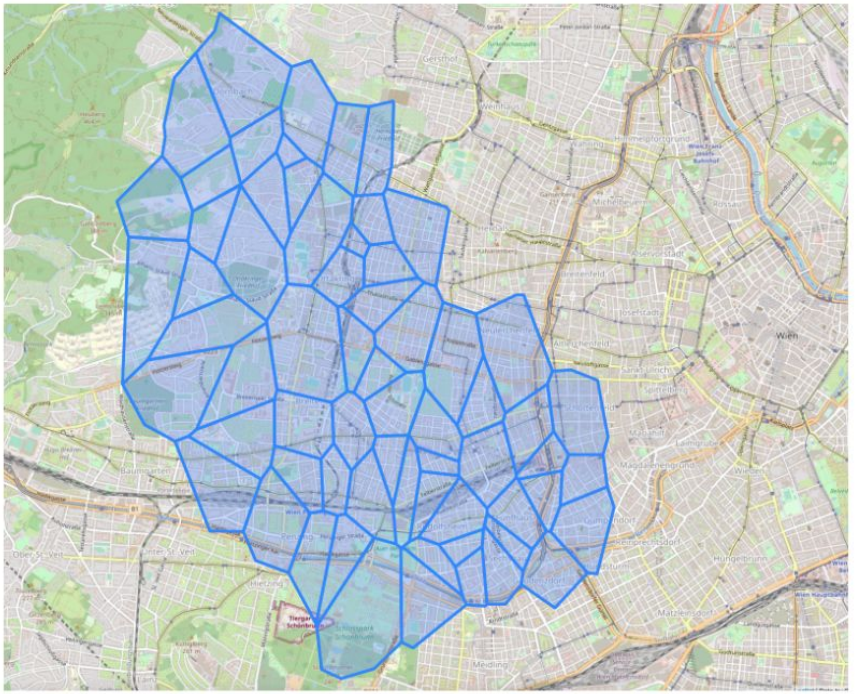
GeoJSON: level-0, level1, level2, level3, level4

KMZ: level-0, level1, level2, level3, level4



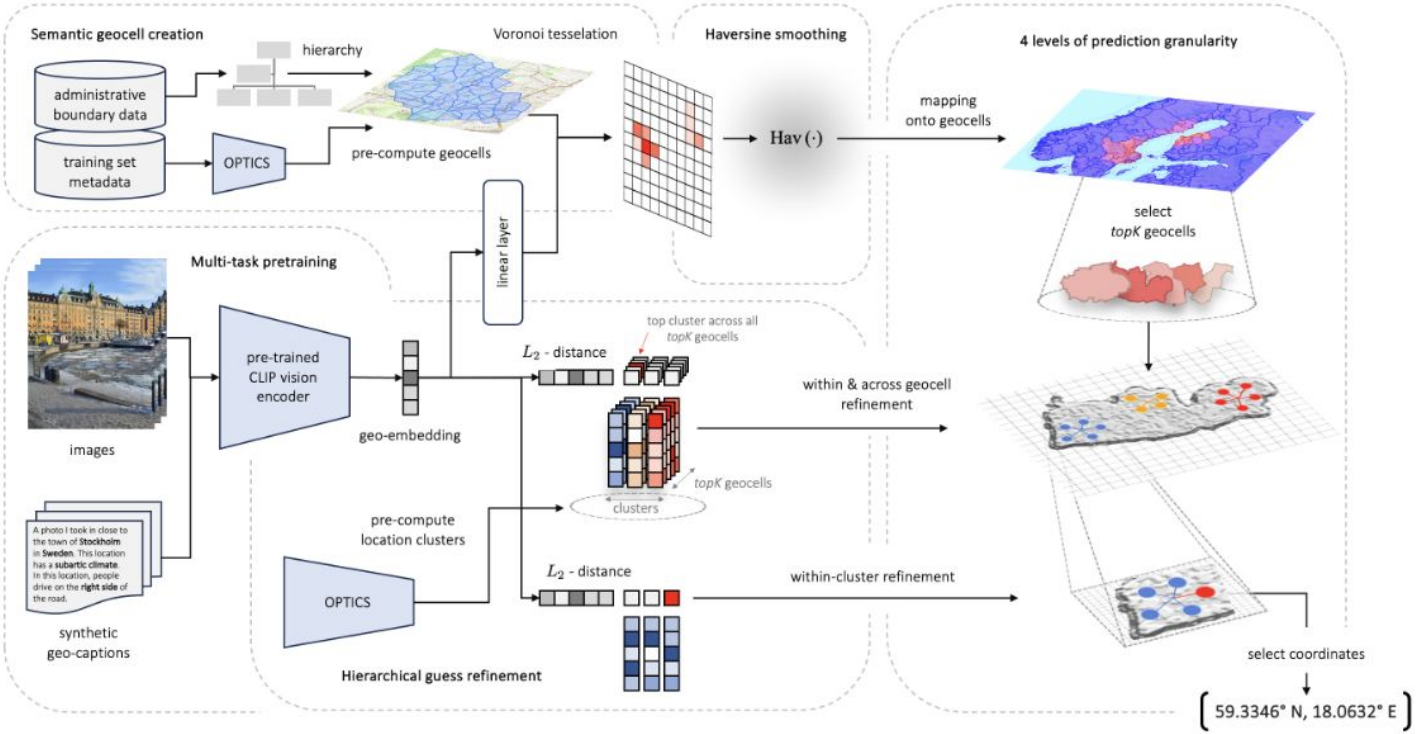
Comment trouver une géolocalisation ?



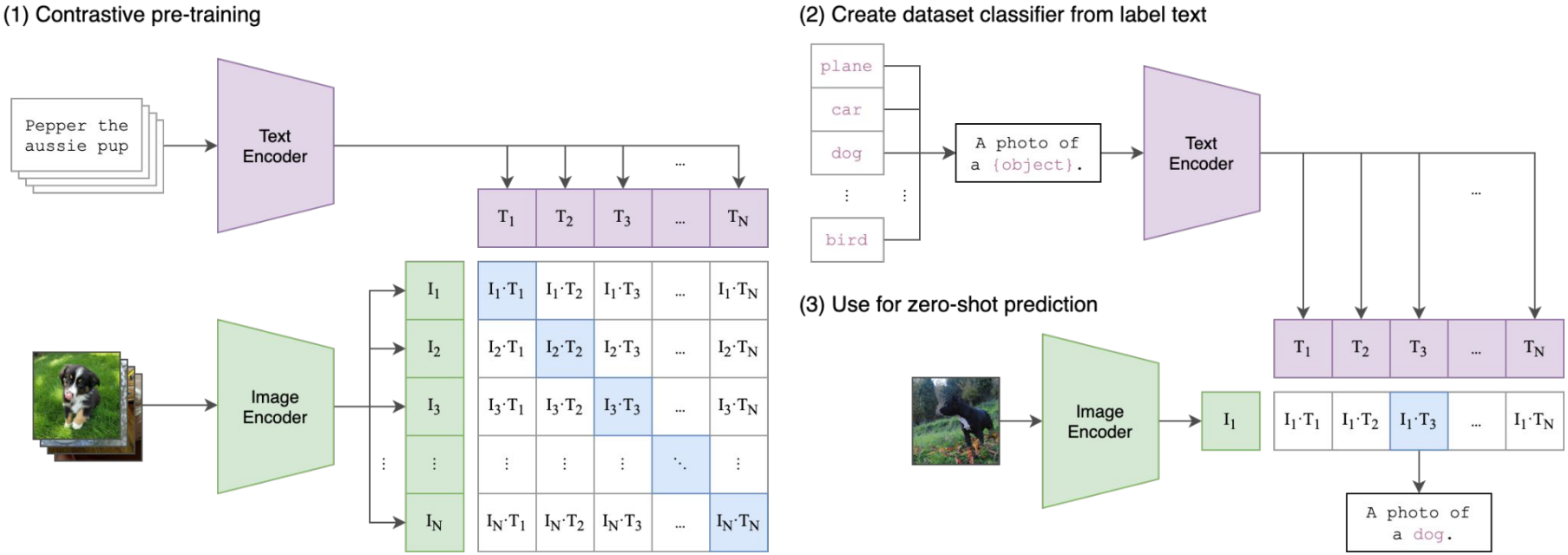


GeoCells : Régions administratives + OPTICS (clustering) + Voronoi

Overview du pipeline :



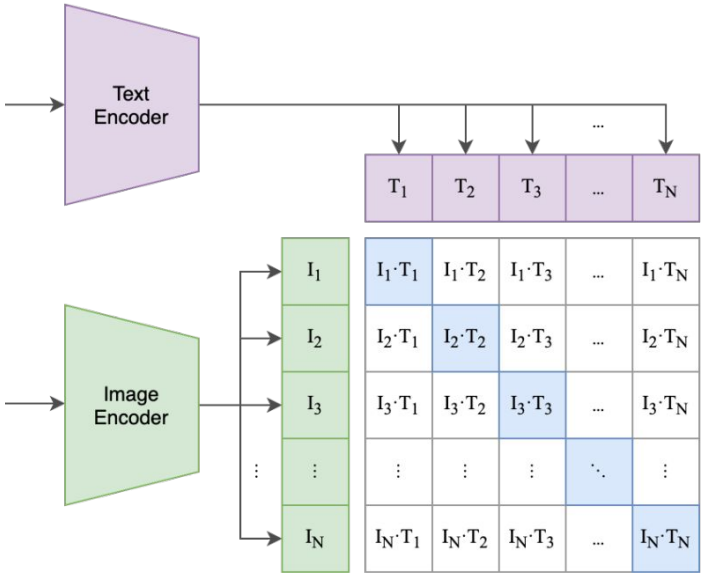
Contrastive Language-Image Pre-Training (CLIP)



Fine-Tune :

Text meta data :

- A photo in the region of Veneto, Italy
- An image facing north
- In northern hemisphere
- People drive on the left side
- The photo was taken in December

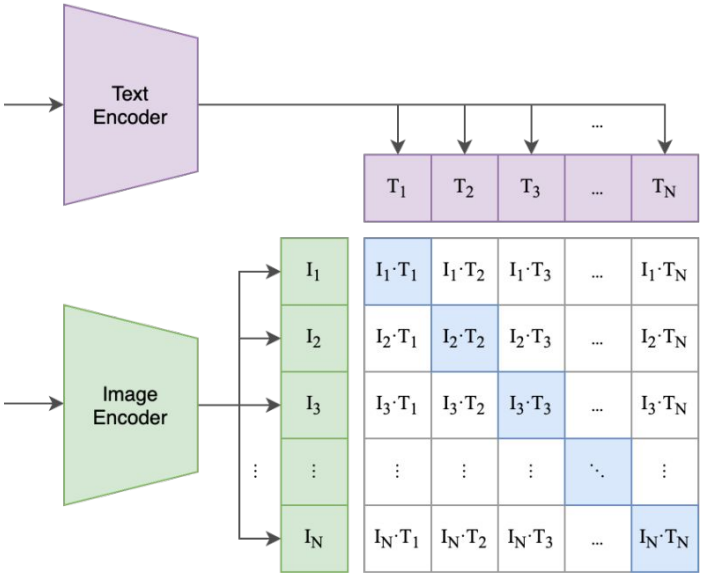


Fine-Tune :

Text meta data :

- A photo in the region of Veneto, Italy
- An image facing north
- In northern hemisphere
- People drive on the left side
- The photo was taken in December

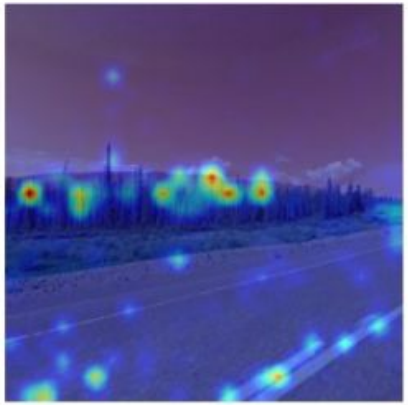
$$\mathcal{L}_n = \mathcal{L}_{n,\text{loc}} + \alpha\mathcal{L}_{n,\text{climate}} + \beta\mathcal{L}_{n,\text{month}} + \gamma\mathcal{L}_{n,\text{reg}}$$

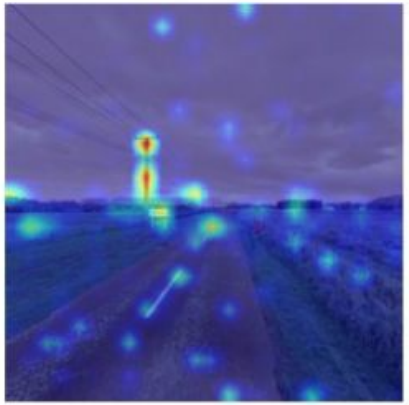


Pourquoi ces informations :

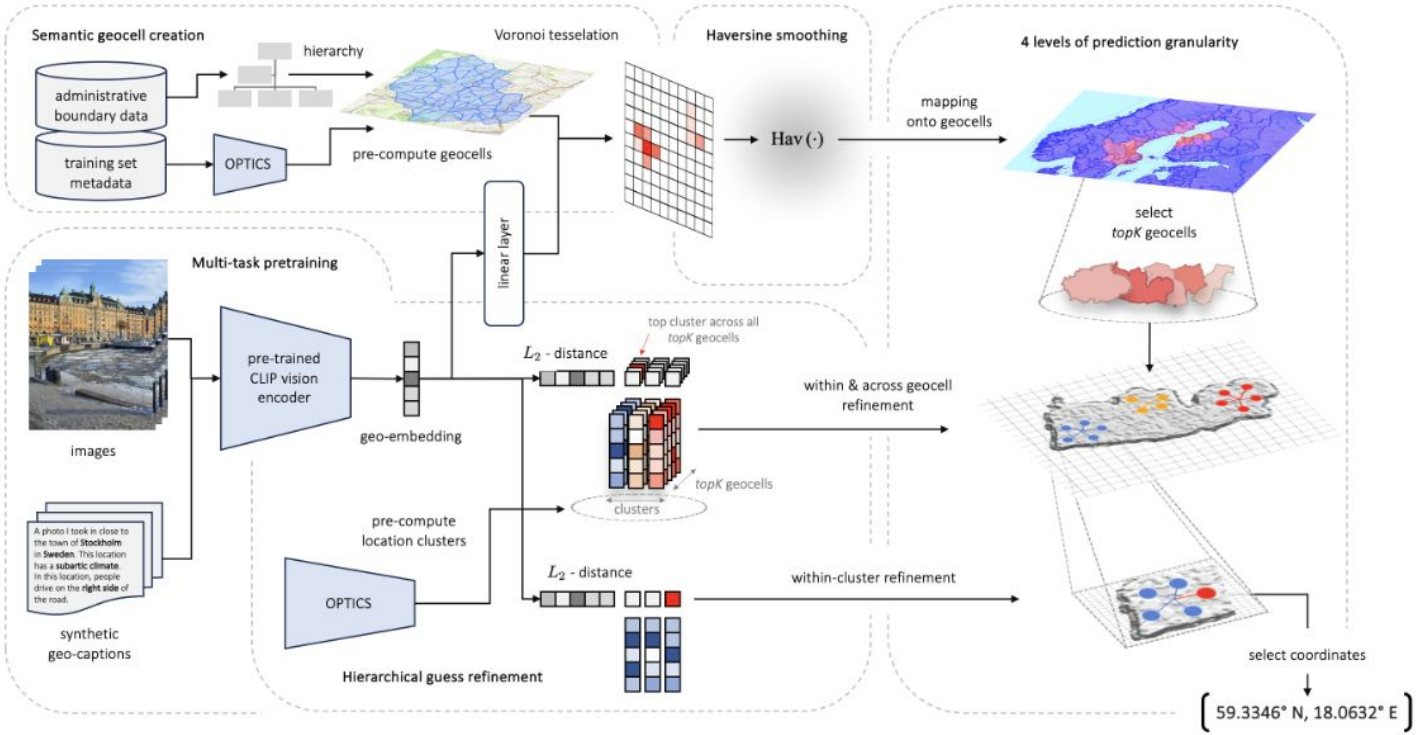
https://geohints.com/

Résultat du pré-entraînement

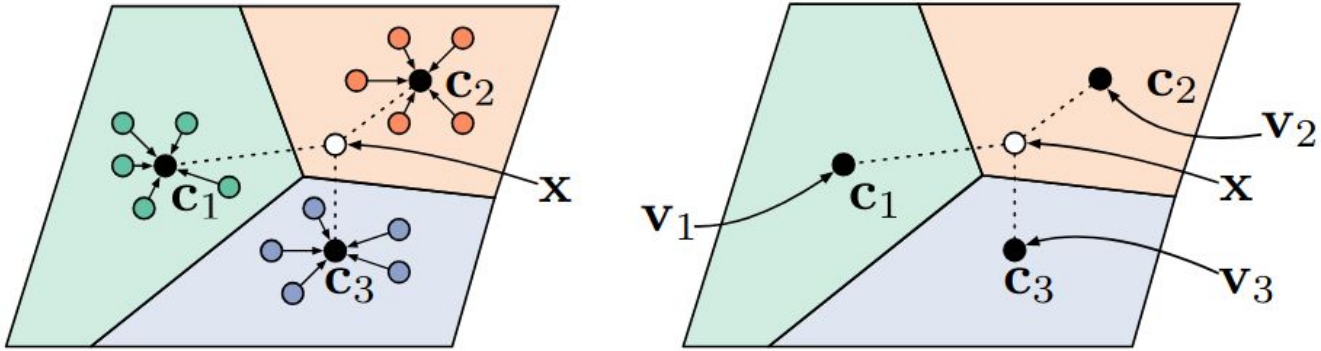





Overview du pipeline :



Raffinement : “prototype-neural networks”



(a) Few-shot (b) Zero-shot

Figure 1: Prototypical networks in the few-shot and zero-shot scenarios. **Left**: Few-shot prototypes $\mathbf{c}_k$ are computed as the mean of embedded support examples for each class. **Right**: Zero-shot prototypes $\mathbf{c}_k$ are produced by embedding class meta-data $\mathbf{v}_k$. In either case, embedded query points are classified via a softmax over distances to class prototypes: $p_{\phi}(y = k|\mathbf{x}) \propto \exp(-d(f_{\phi}(\mathbf{x}), \mathbf{c}_k))$.

Raffinement : “Finding nearest neighbors”

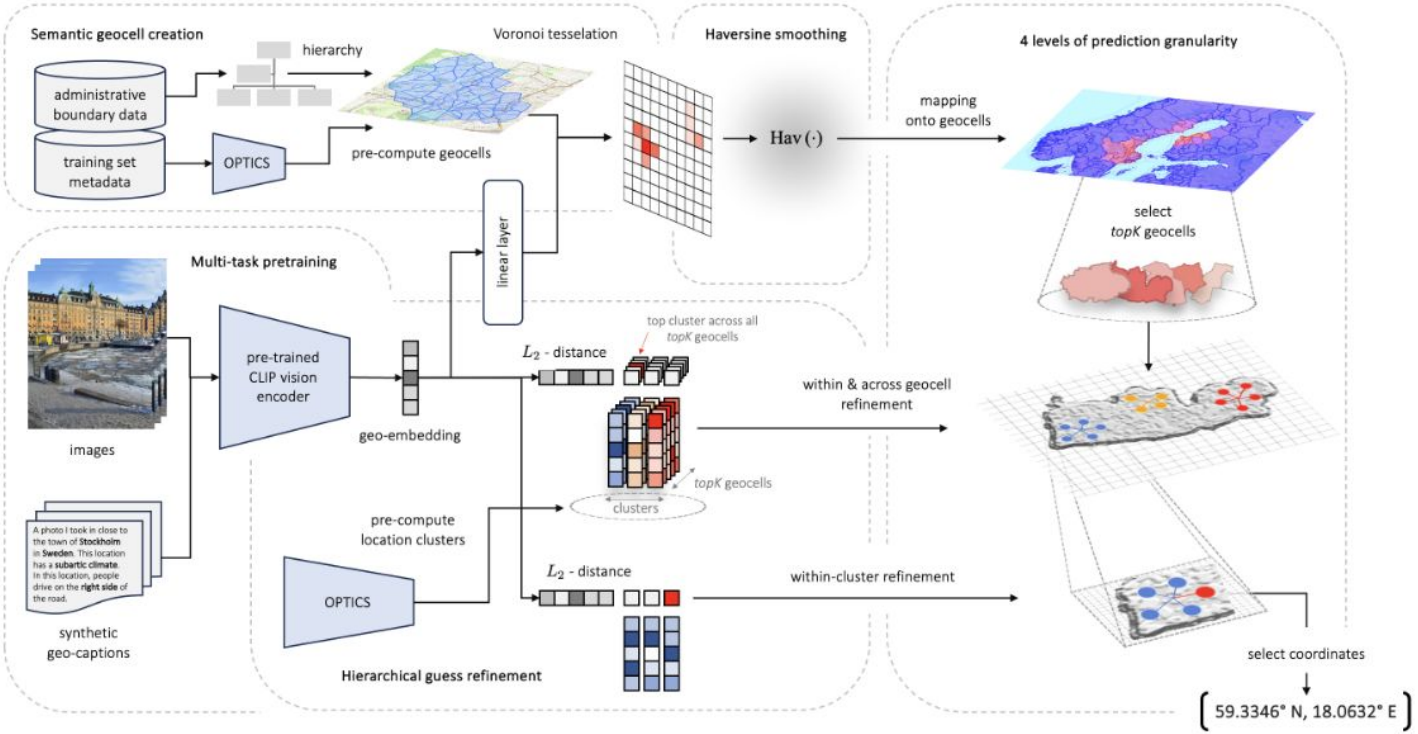
$$c^* = \arg\min_{c \in C_g} \left\| f(x) - \frac{1}{|c|} \sum_{l \in c} f(l) \right\|_2$$

f : CLIP

C_g : Candidates clusters

l : Locations within a cluster

Overview du pipeline :



Input :

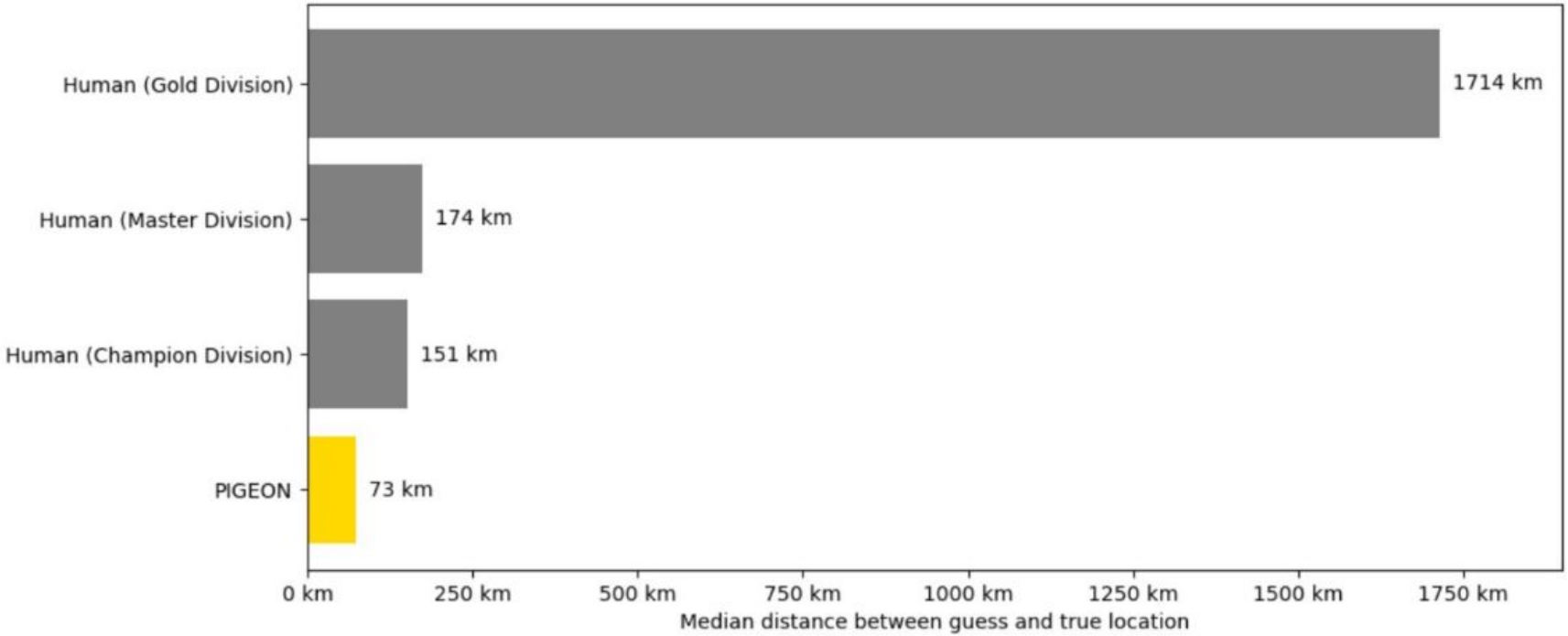
Figure 6: Four images comprising a 360-degree panorama from a location in Pegswood, England, in our dataset.

Results :

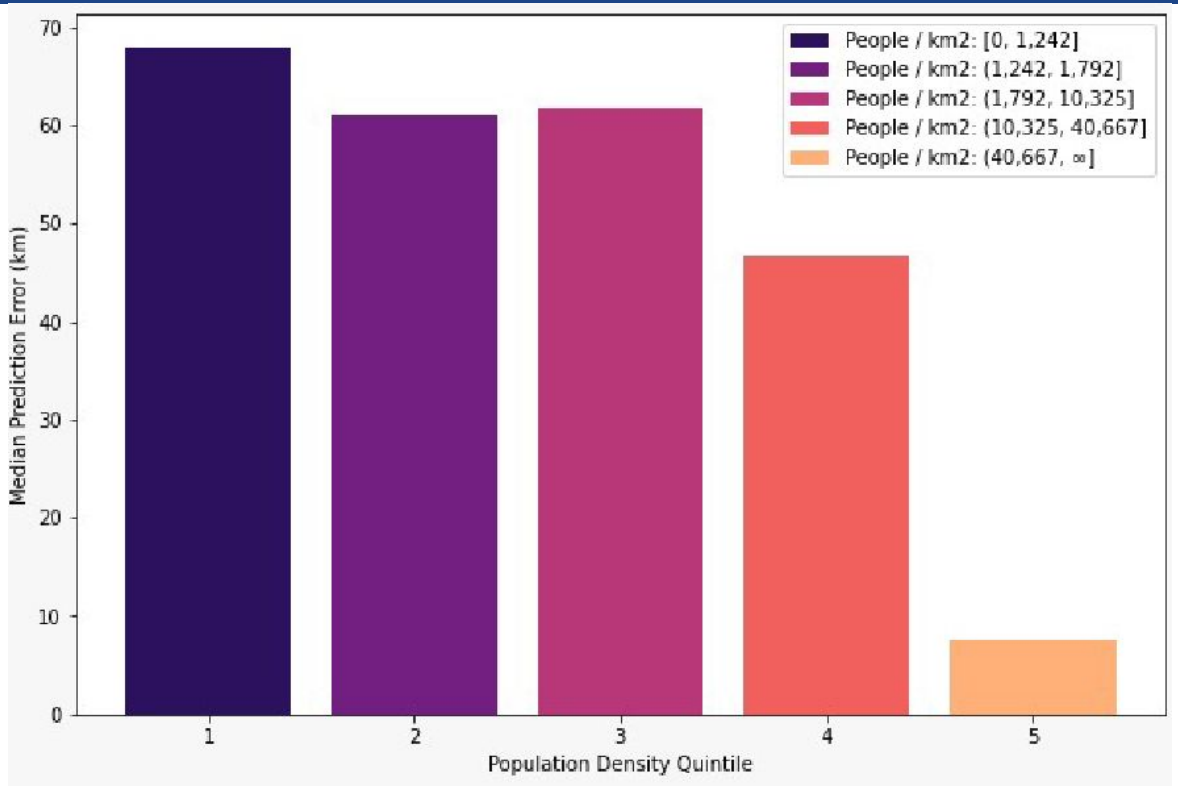
Ablation	Country Accuracy %	Mean Error *km*	Median Error *km*	Geoguessr Score *points*
PIGEON	**91.96**	**251.6**	**44.35**	4,525

Ablation	Distance (% @ km)				
	Street 1 km	*City* 25 km	*Region* 200 km	*Country* 750 km	*Continent* 2,500 km
PIGEON	**5.36**	**40.36**	78.28	94.52	**98.56**

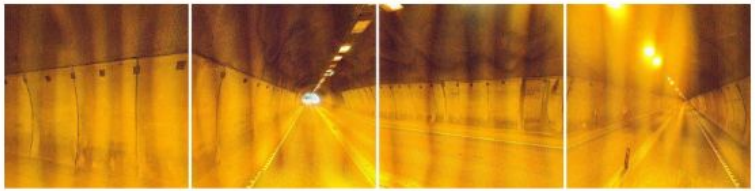
Results :


Human (Gold Division)
1714 km
Human (Master Division)
174 km
Human (Champion Division)
151 km
PIGEON
73 km
0 km
250 km
500 km
750 km
1000 km
1250 km
1500 km
1750 km
Median distance between guess and true location

Urban vs rural :



Worst-case :



(a) Image from a tunnel.

(b) Image from a body of water.

(c) Image from a dark area.

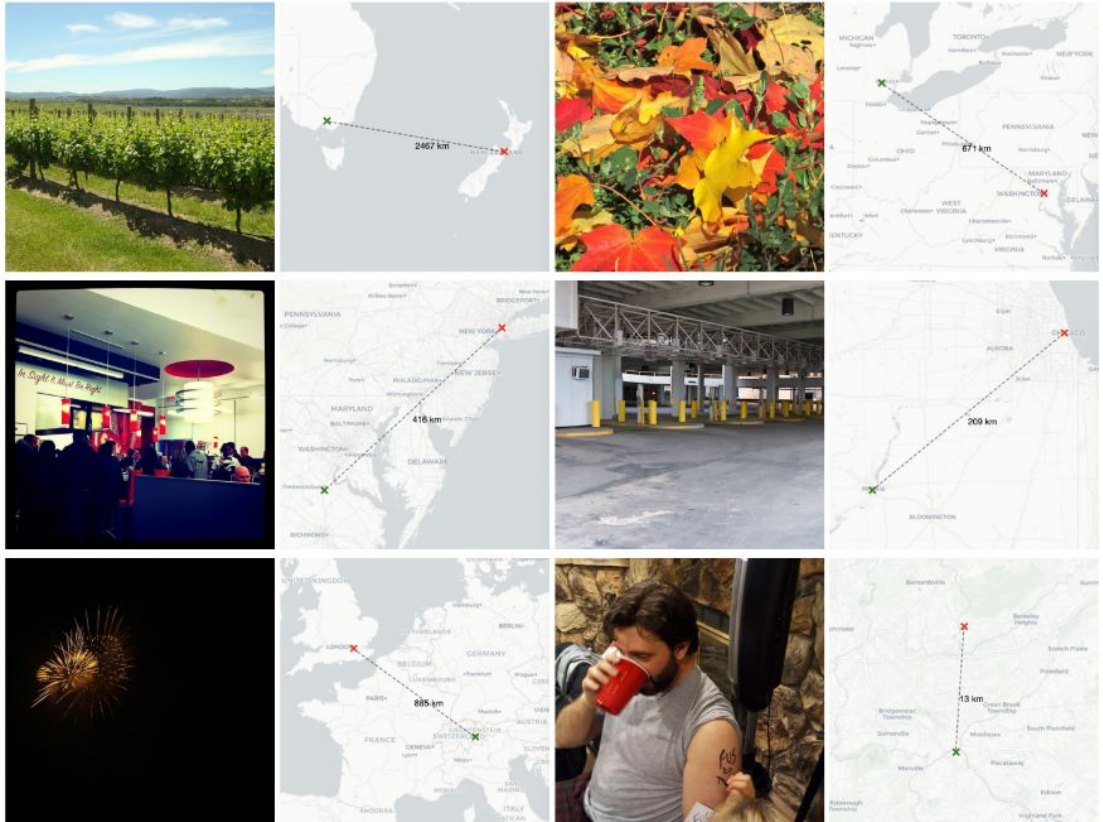
(d) Image from a forest.

(e) Image from an indoor area.

(f) Image from a soccer stadium.

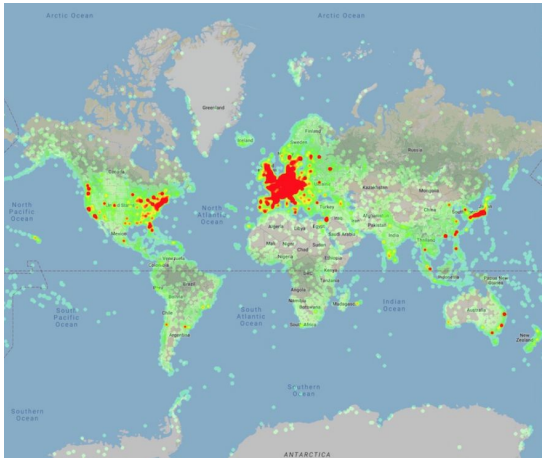
Figure 11: Examples of images for which PIGEON was the most uncertain about the correct location.

Results - PIGEOETTO :



“General image locator”:

- MediaEval database (Geolocated Flickr images)
- Google Landmarks v2



Ethics

FIN