

GeoCLIP: Clip-Inspired Alignment between Locations and Images for Effective Worldwide Geo-localization

Vicente Vivanco Cepeda

Gaurav Kumar Nayak

Mubarak Shah

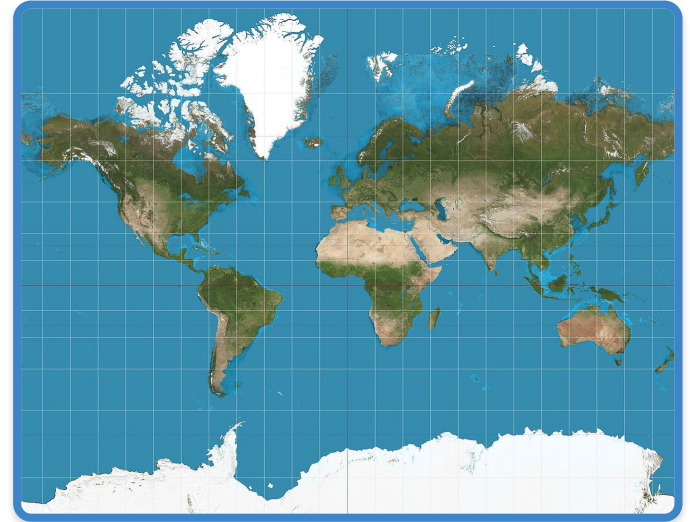


Worldwide Geo-localization

Image Geo-localization at global scale

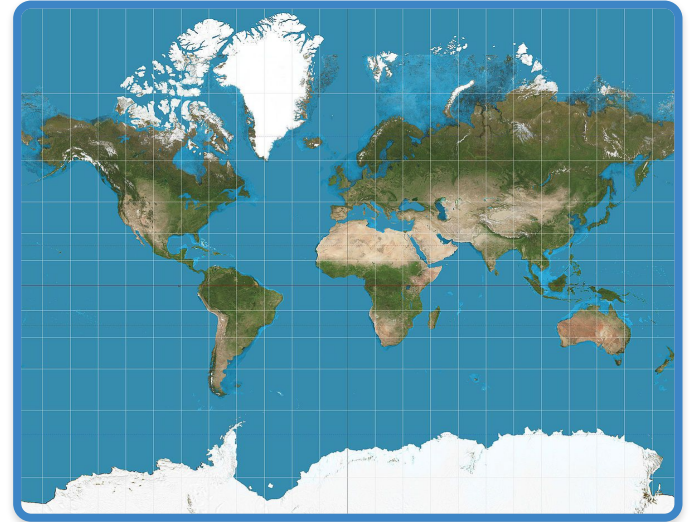
Worldwide Geo-localization

Image Geo-localization at global scale



Worldwide Geo-localization

Image Geo-localization at global scale



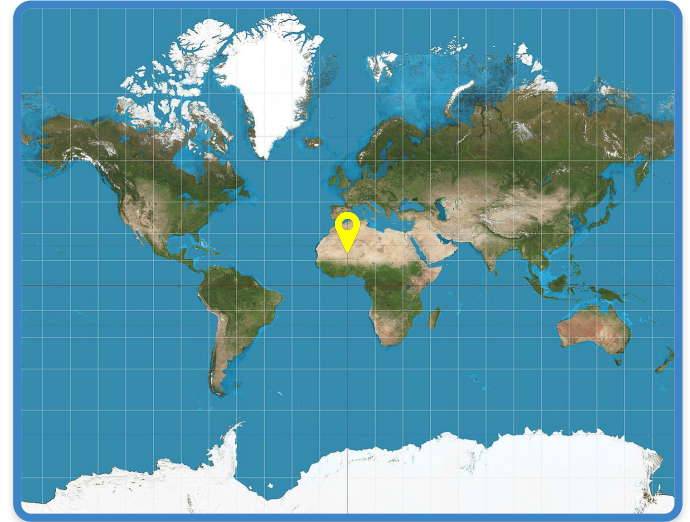
Worldwide Geo-localization

Image Geo-localization at global scale



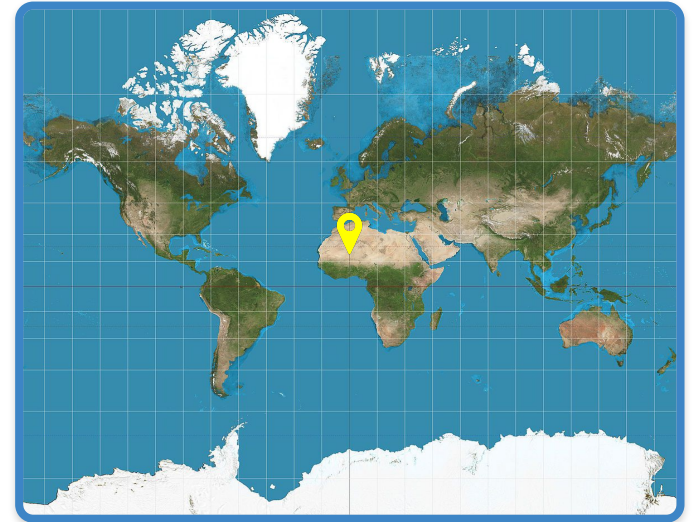
Worldwide Geo-localization

Image Geo-localization at global scale



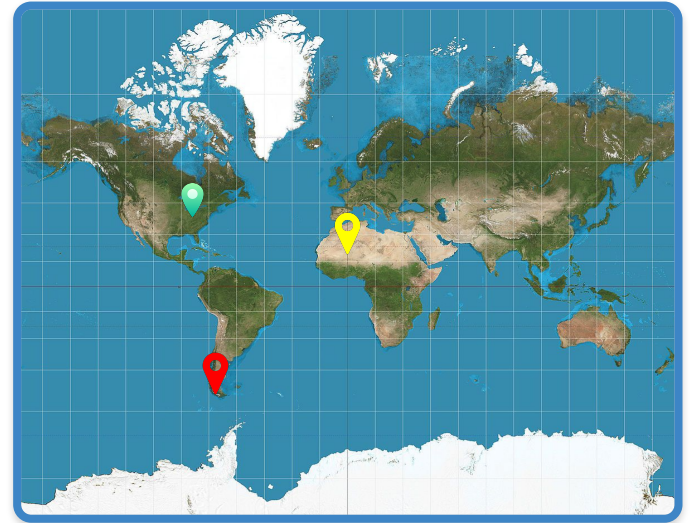
Worldwide Geo-localization

Image Geo-localization at global scale



Worldwide Geo-localization

Image Geo-localization at global scale

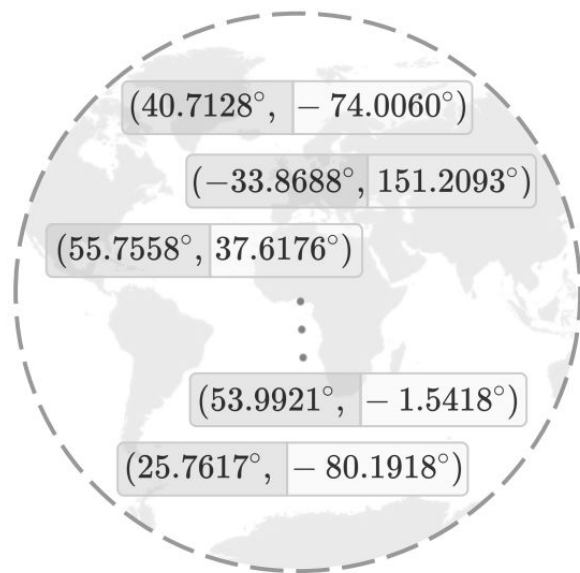


Our Approach (Image-to-GPS Retrieval)

Image



Our Approach (Image-to-GPS Retrieval)



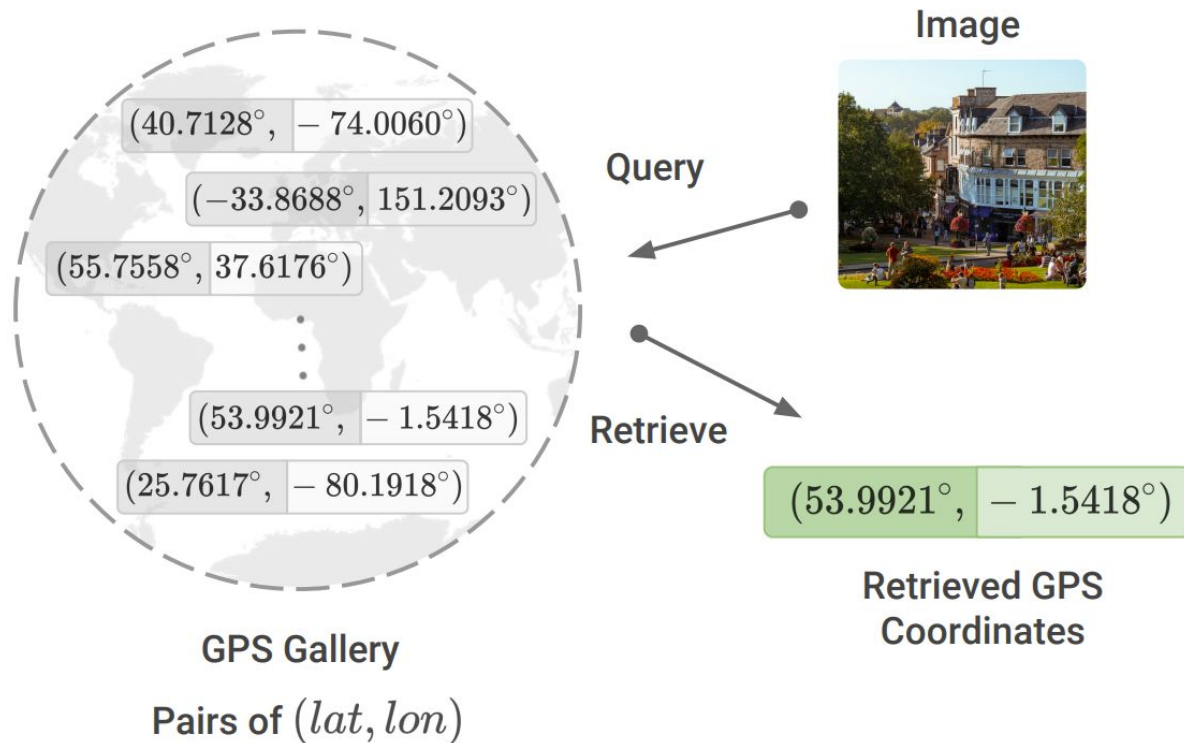
GPS Gallery

Pairs of (lat, lon)

Query



Our Approach (Image-to-GPS Retrieval)



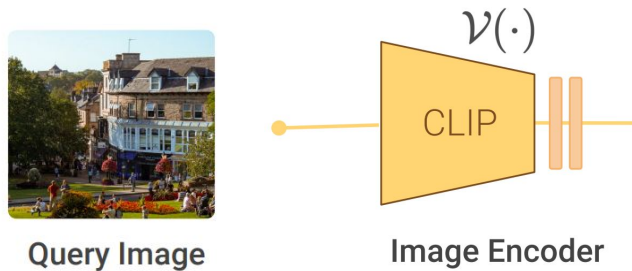
Our Approach (Image-to-GPS Retrieval)

Our Approach (Image-to-GPS Retrieval)

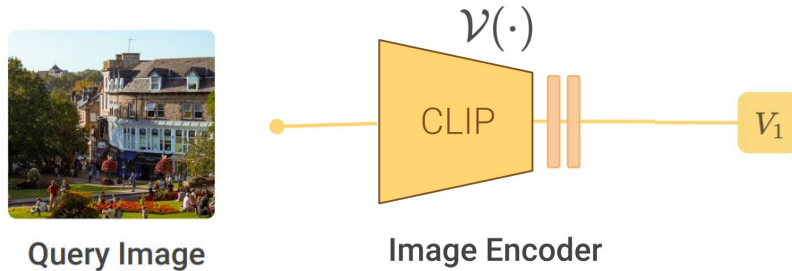


Query Image

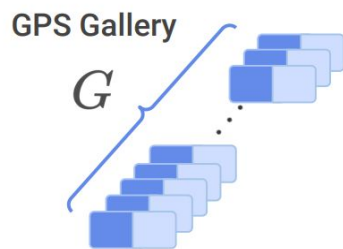
Our Approach (Image-to-GPS Retrieval)



Our Approach (Image-to-GPS Retrieval)



Our Approach (Image-to-GPS Retrieval)



Pairs of (lat, lon)



Query Image

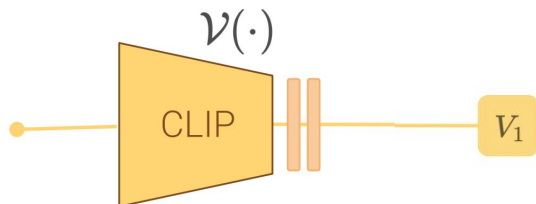
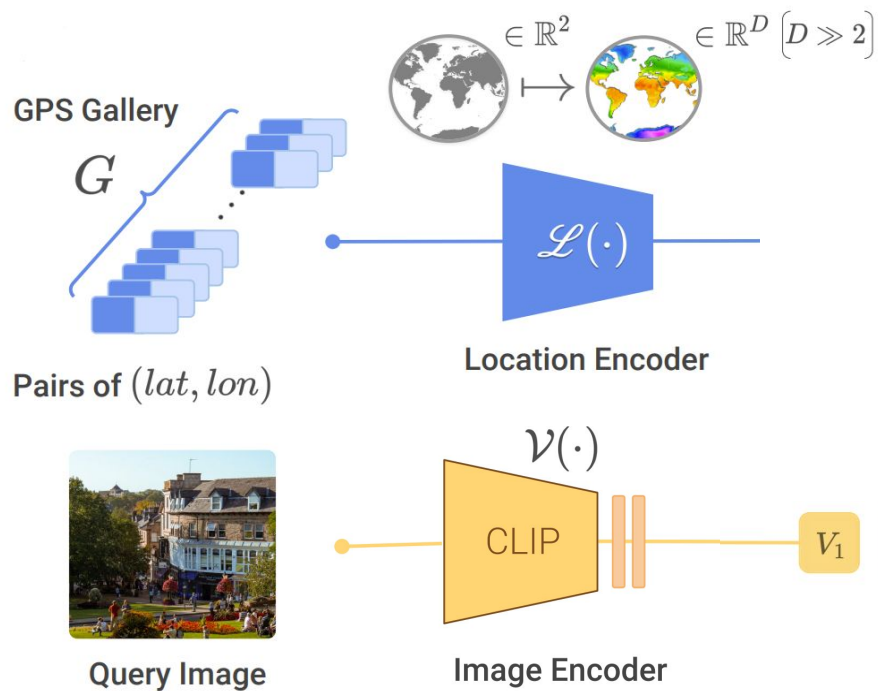
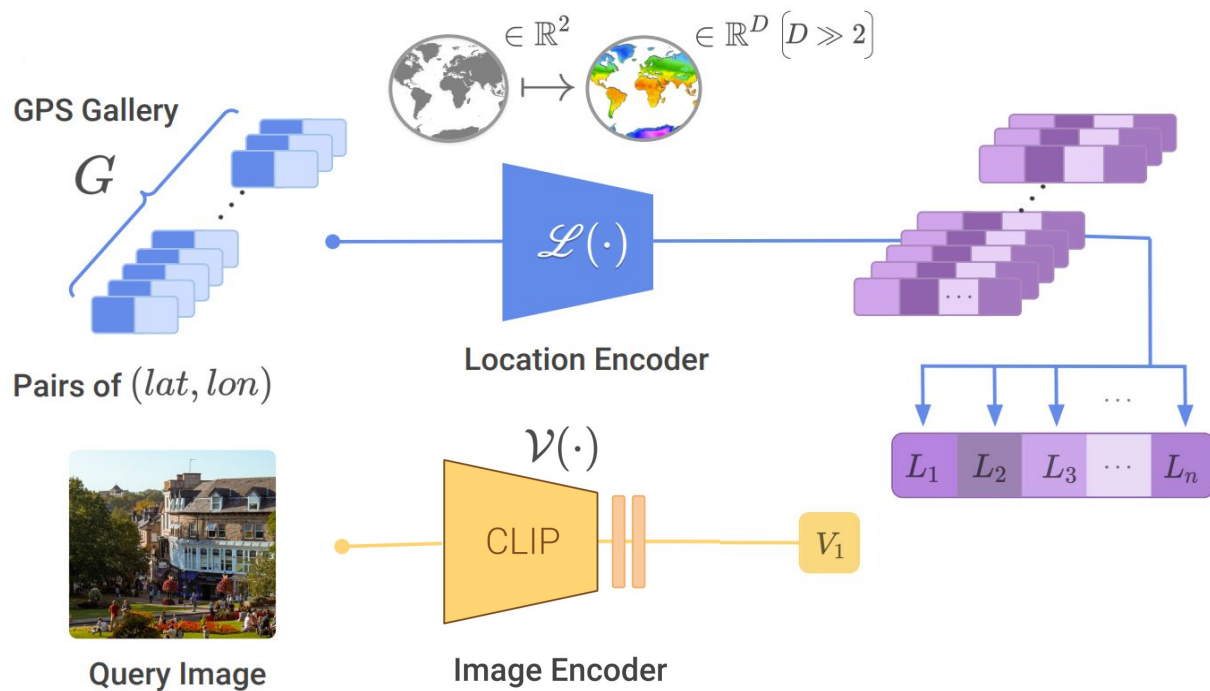


Image Encoder

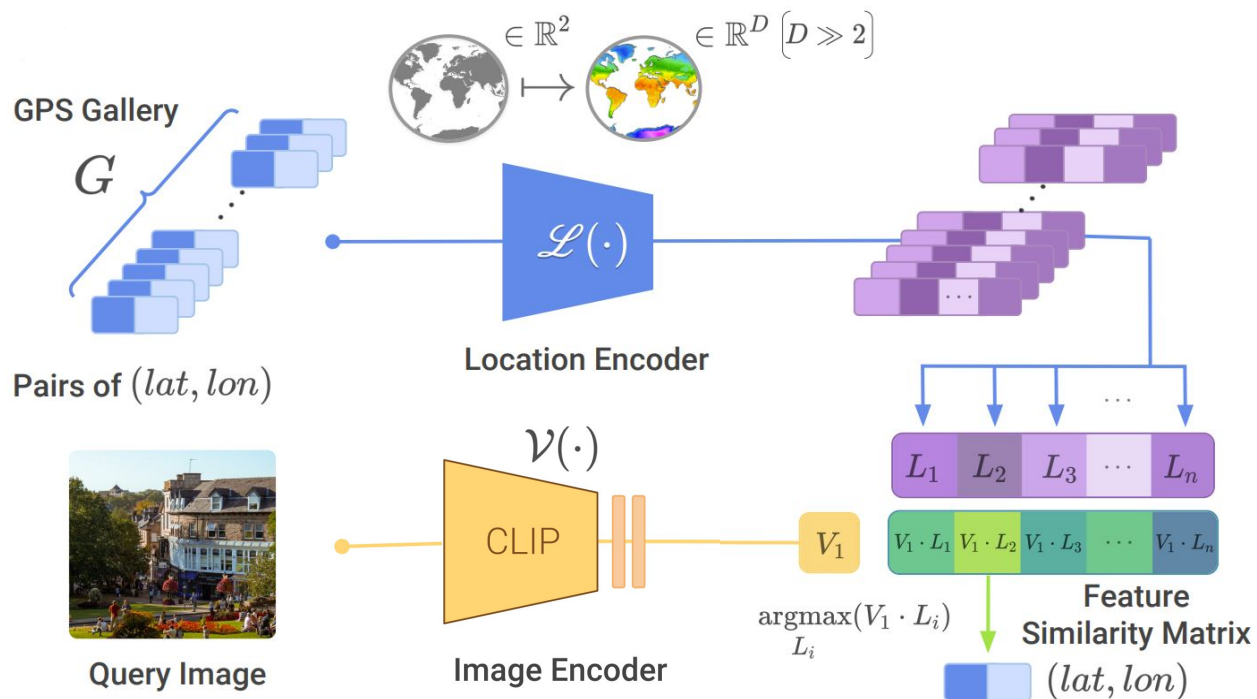
Our Approach (Image-to-GPS Retrieval)



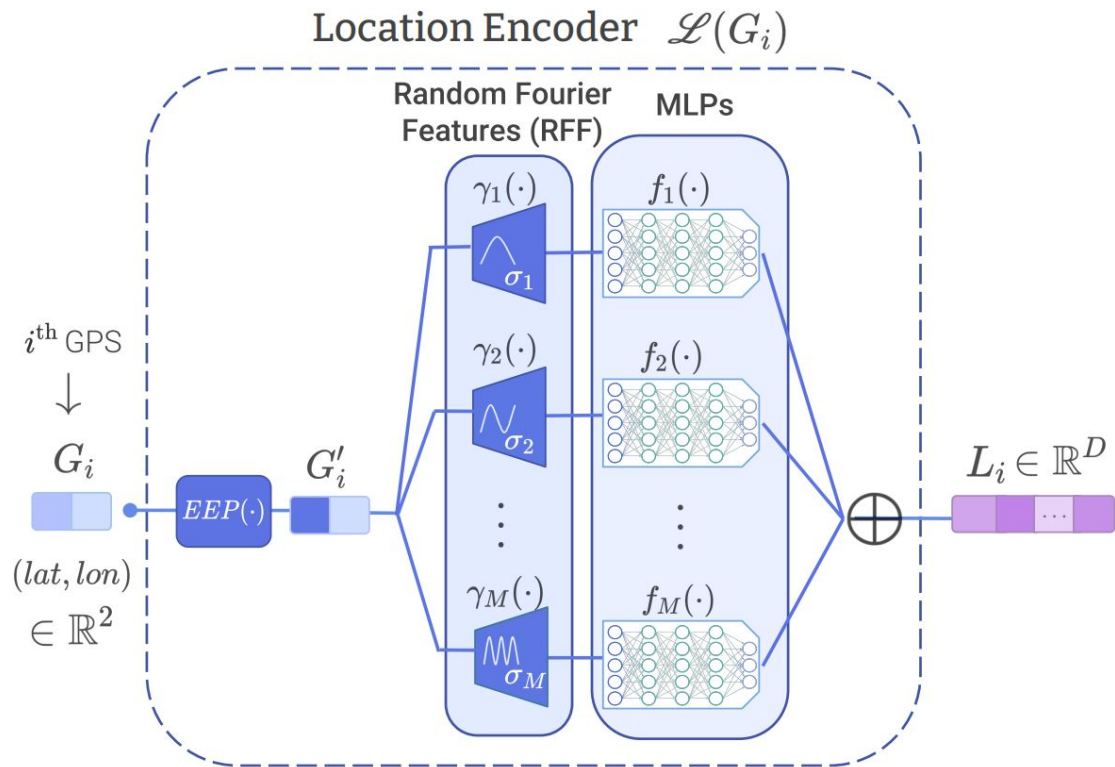
Our Approach (Image-to-GPS Retrieval)



Our Approach (Image-to-GPS Retrieval)



Location Encoder



Im2GPS3K Dataset

| Method | Street 1 km | City 25 km | Region 200 km | Country 750 km | Continent 2500 km |
|---------------------------|------------------------|-----------------------|--------------------------|---------------------------|------------------------------|
| [L]kNN, $\sigma = 4$ [26] | 7.2 | 19.4 | 26.9 | 38.9 | 55.9 |
| PlaNet [27] | 8.5 | 24.8 | 34.3 | 48.4 | 64.6 |
| CPlaNet [18] | 10.2 | 26.5 | 34.6 | 48.6 | 64.6 |
| ISNs [12] | 10.5 | 28.0 | 36.6 | 49.7 | 66.0 |
| Translocator [14] | 11.8 | 31.1 | 46.7 | 58.9 | 80.1 |
| GeoDecoder [5] | 12.8 | 33.5 | 45.9 | 61.0 | 76.1 |
| Ours | 14.11 | 34.47 | 50.65 | 69.67 | 83.82 |

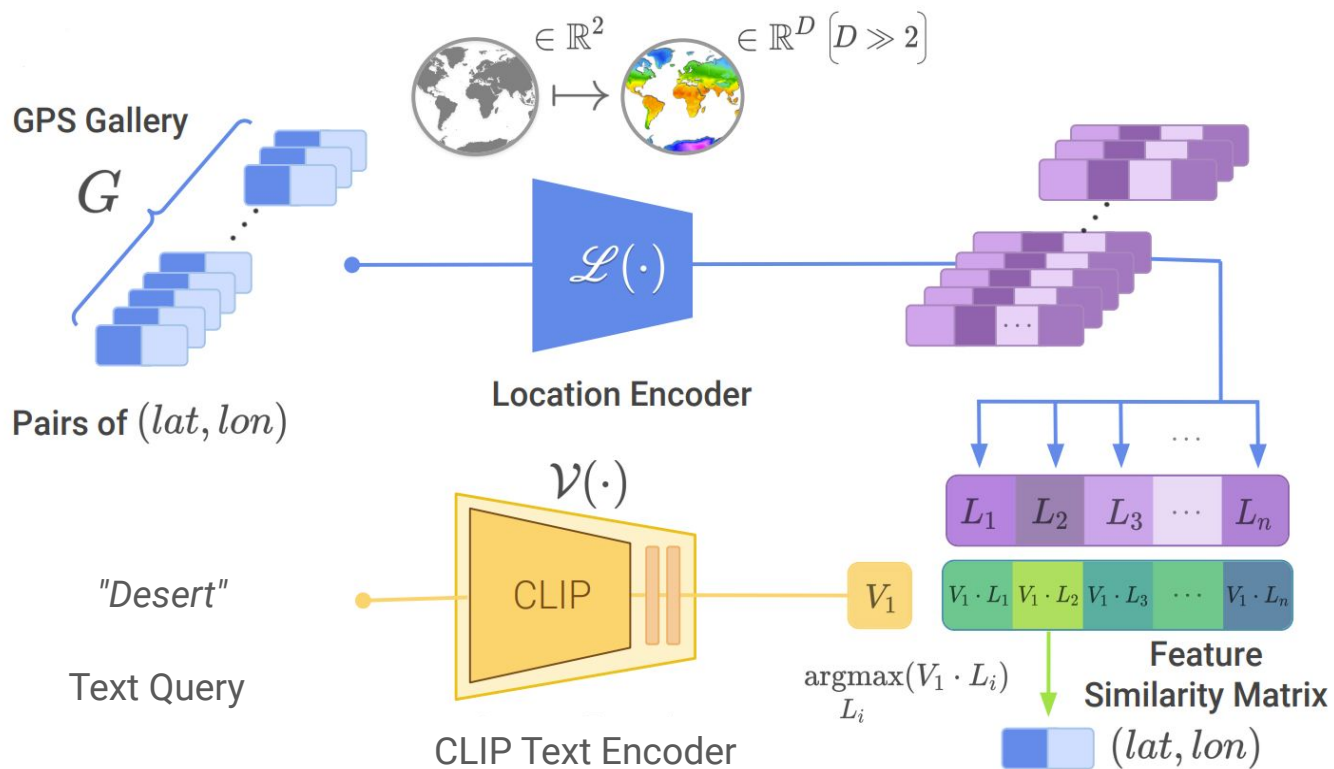
Google World Streets 15K Dataset

| Method | Street 1 km | City 25 km | Region 200 km | Country 750 km | Continent 2500 km |
|-------------------|------------------------|-----------------------|--------------------------|---------------------------|------------------------------|
| ISNs [12] | 0.05 | 0.6 | 4.2 | 15.5 | 38.5 |
| Translocator [14] | 0.5 | 1.1 | 8.0 | 25.5 | 48.3 |
| GeoDecoder [5] | 0.7 | 1.5 | 8.7 | 26.9 | 50.5 |
| Ours | 0.6 | 3.1 | 16.9 | 45.7 | 74.1 |

Beyond Geo-localization: Geo-tagged NUS-WIDE Dataset

| Method | Classifier | mAP |
|---------------------------|-------------------|--------------|
| Visual (V) | visual | 0.234 |
| OneHot [3] | geo | 0.066 |
| HashTag [22] | geo | 0.163 |
| TagFeature [10] | geo | 0.161 |
| GPS2Vec onehot [30] | geo | 0.130 |
| GPS2Vec [30] | geo | 0.182 |
| GPS2Vec+ [31] | geo | 0.210 |
| GeoCLIP (Ours) | geo | 0.249 |
| OneHot [3] + V | fusion | 0.238 |
| HashTag [22] + V | fusion | 0.261 |
| TagFeature [10] + V | fusion | 0.260 |
| GPS2Vec Onehot [30] + V | fusion | 0.277 |
| GPS2Vec [30] + V | fusion | 0.300 |
| Liao et al. [10] | fusion | 0.347 |
| GPS2Vec+ [31] + V | fusion | 0.348 |
| GeoCLIP (Ours) + V | fusion | 0.362 |

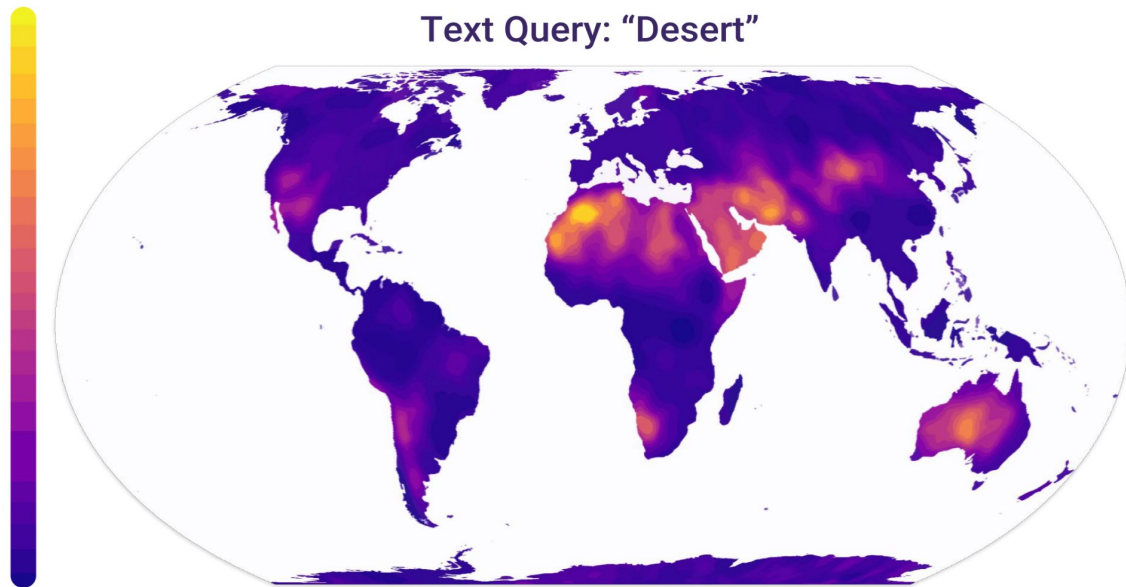
Geo-localization using Text Query



Geo-localization using Text Query

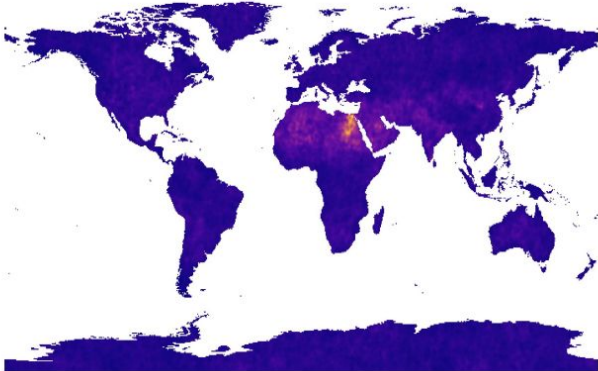
Text Query: "Desert"

Geo-localization using Text Query



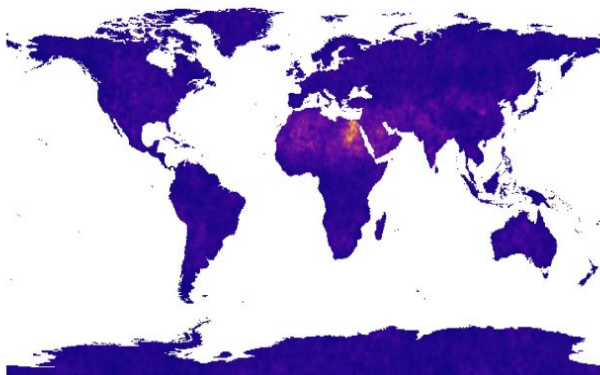
Geo-localization using Text Query

Query: "Pyramids of Giza"

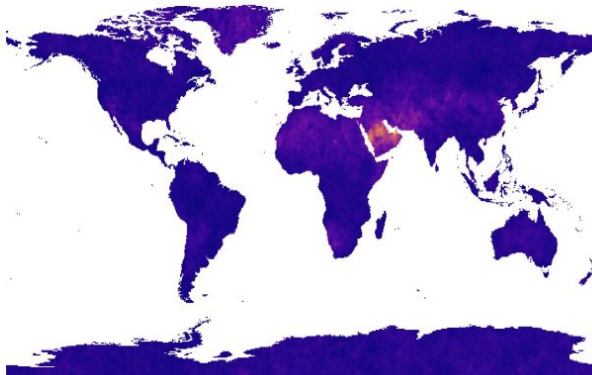


Geo-localization using Text Query

Query: "Pyramids of Giza"

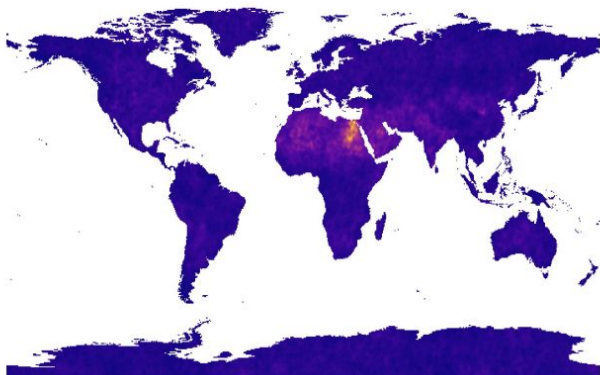


Query: "Burj Khalifa"

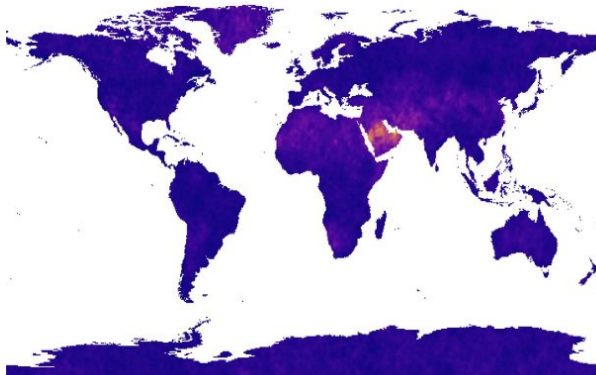


Geo-localization using Text Query

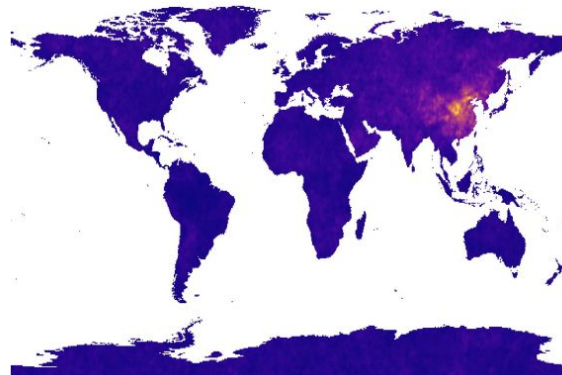
Query: "Pyramids of Giza"



Query: "Burj Khalifa"

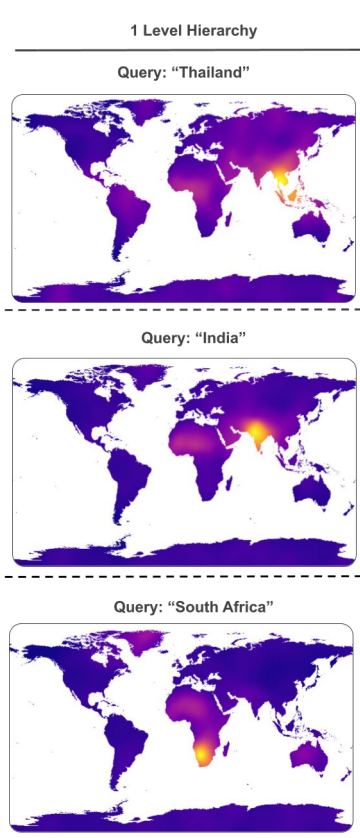


Query: "Great Wall of China"

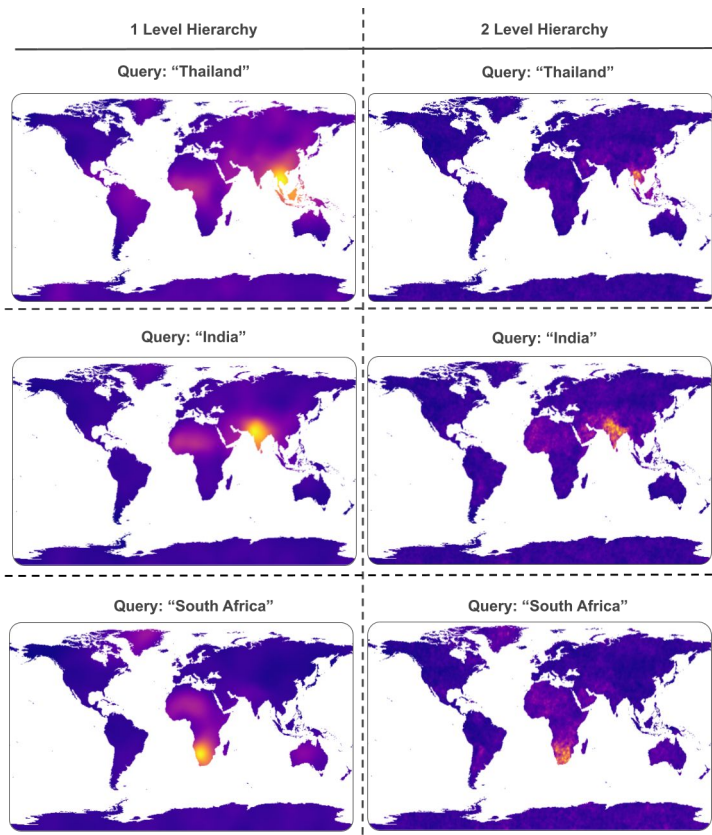


Geo-localization using Text Query

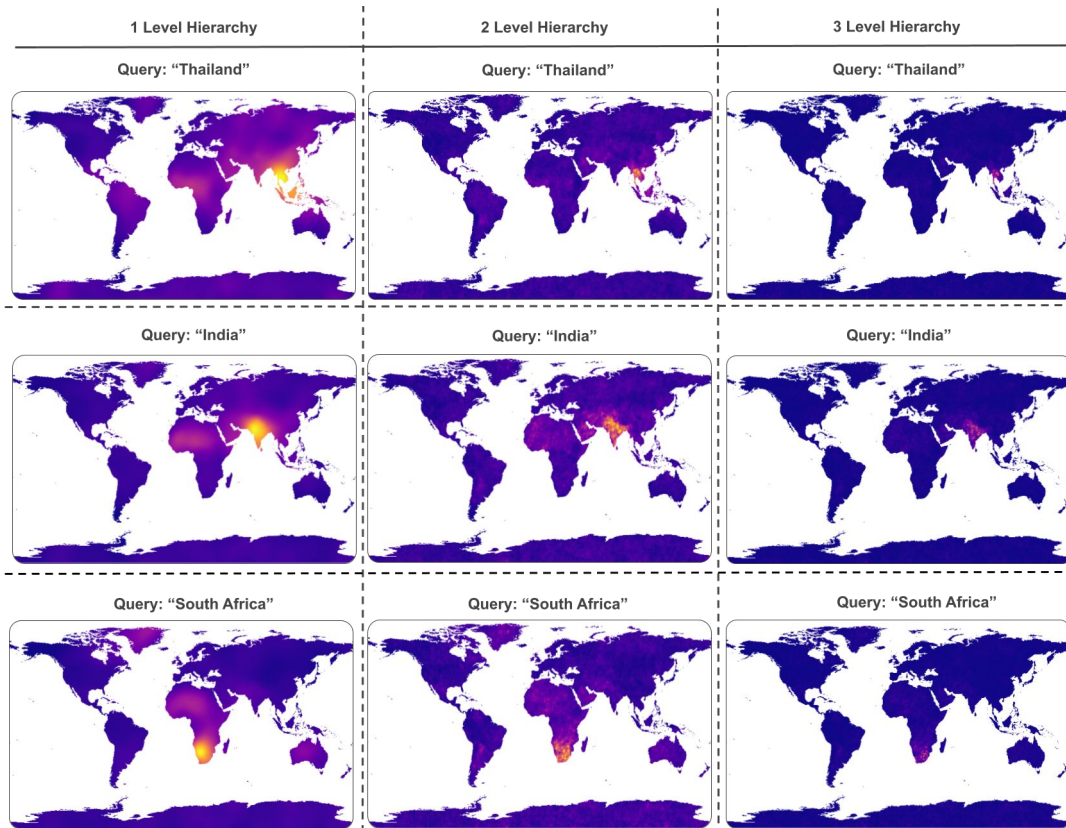
Geo-localization using Text Query



Geo-localization using Text Query

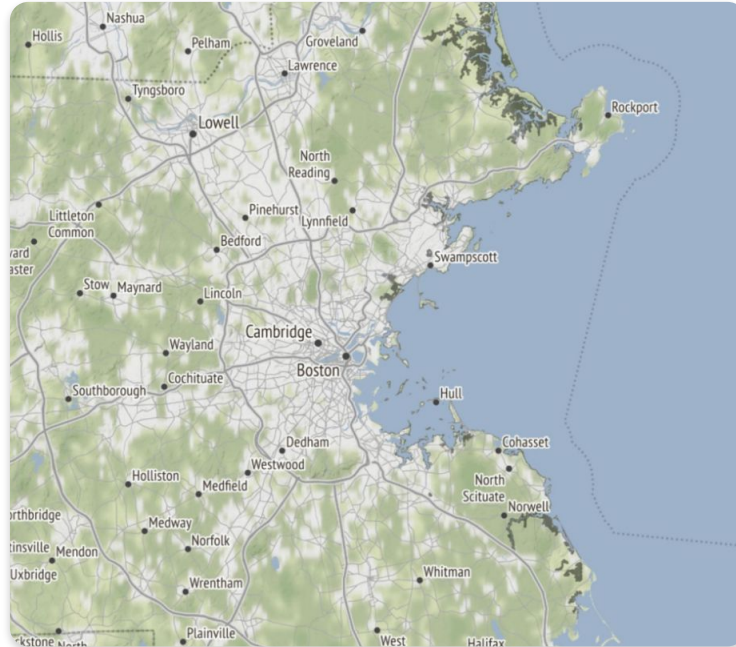


Geo-localization using Text Query



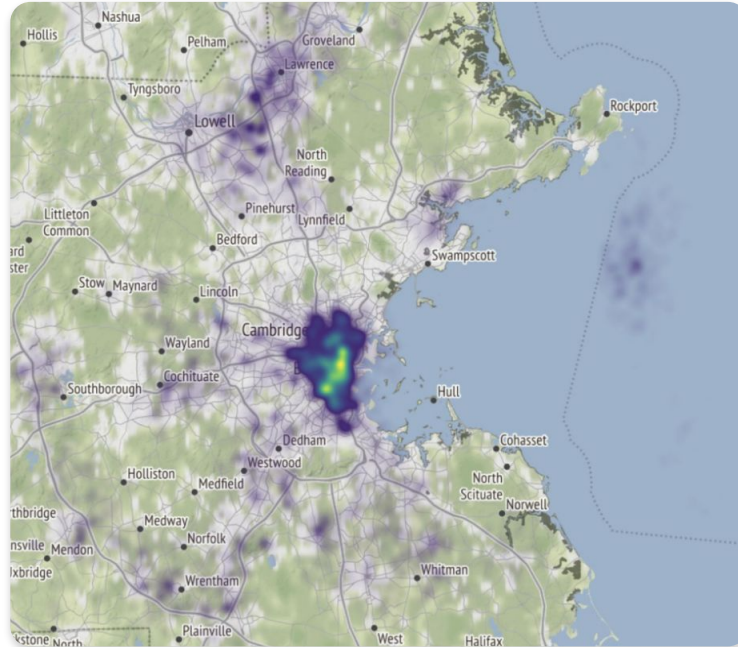
Zooming In: Geo-localize at City Scales

Text Query: Boston, Massachusetts



Zooming In: Geo-localize at City Scales

Text Query: Boston, Massachusetts



Thank you!