Overview

- **Goal:** Determine the geolocation of any photo.
- **Geolocation by classification** using CNNs.
  - Input: photos, output classes: geographical cells.
  - Classification allows multimodal location predictions.
- **Localizes landmarks, landscapes, locally typical objects, and even some plants and animals.**
- Competitive with Im2GPS, outperforms humans.

Geographical Partitioning Scheme

- **Adaptive partitioning using S2 cells.**
  - Recursive splitting until no cell contains over 150 photos.
  - Discard cells with less than 10 photos.
  - Even class distribution, adaptive spatial resolution.
- Up to **street level resolution in densely populated areas.**

Training

- **Data:** 91M photos from the web with Exif locations.
- 26,263 classes (10k, 50). Target is 1-hot vector.
- Inception architecture, 97M parameters, trained 2.5 months.

Results

- Correctly localized photos
- Incorrectly localized photos
- Top photos by region

PlaNet vs. Im2GPS

**Im2GPS:** Localizes query by nearest neighbor matching against Flickr photos. Im2GPS (new) adds geo-clustering, 1-vs-all SVM.

<table>
<thead>
<tr>
<th>Method</th>
<th>Index / Training Images</th>
<th>Street 1km</th>
<th>City 25km</th>
<th>Region 200km</th>
<th>Country 750km</th>
<th>Continent 250km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Im2GPS (orig)</td>
<td>6.6M</td>
<td>12.0%</td>
<td>15.0%</td>
<td>23.0%</td>
<td>47.0%</td>
<td>51.9%</td>
</tr>
<tr>
<td>Im2GPS (new)</td>
<td>6.5M</td>
<td>2.5%</td>
<td>21.9%</td>
<td>32.1%</td>
<td>35.4%</td>
<td>51.9%</td>
</tr>
<tr>
<td>PlaNet</td>
<td>900k</td>
<td>0.4%</td>
<td>3.8%</td>
<td>7.6%</td>
<td>21.6%</td>
<td>42.5%</td>
</tr>
<tr>
<td>PlaNet</td>
<td>6.2M</td>
<td>6.3%</td>
<td>18.1%</td>
<td>30.0%</td>
<td>45.6%</td>
<td>65.8%</td>
</tr>
<tr>
<td>PlaNet</td>
<td>91M</td>
<td>8.4%</td>
<td>24.5%</td>
<td>37.6%</td>
<td>53.6%</td>
<td>71.3%</td>
</tr>
</tbody>
</table>

Model comparison

<table>
<thead>
<tr>
<th>Features Prediction</th>
<th>PlaNet</th>
<th>Im2GPS (new)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>377 MB</td>
<td>577 GB</td>
</tr>
</tbody>
</table>

Playing GeoGuessr against Humans

- **Goal:** Guess the location of a random street view panorama.
- 10 humans played 50 rounds against PlaNet. PlaNet won 28.
- Median error: 1131.7km (PlaNet), 2320.8km (humans)

Model Analysis

Comparison to retrieval

Reducing model size

Geolocating Photo Albums

- **Idea:** Context can help disambiguate hard-to-locate photos.
  - Using LSTM (long short term memory) architecture to predict the sequence of geolocations in photo albums.
- **Data:** 29.7M albums (616M photos), split 80% / 20%.