Interpretable Semantic Photo Geolocation

Jonas Theiner, Eric Müller-Budack, and Ralph Ewerth

5 min summary

@thejtheiner
theiner@l3s.de
https://github.com/jtheiner/semantic_geo_partitioning

Problem Formulation

Task:

- Predicting the location of a photo
- At plane-scale without any restrictions (landmarks, indoor, etc.)

Problem:

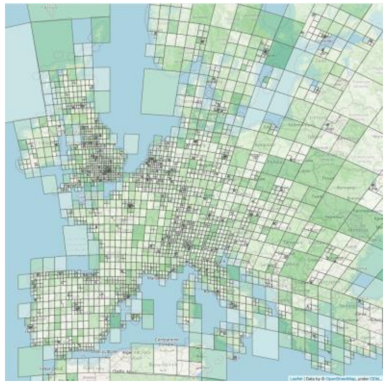
- Humans are far worse than Deep Learning solutions
- Deep Neural Networks (DNNs) lack explainability & interpretability

How to improve the **interpretability** & **explainability** for the geolocation estimation task?

Related Work

- **Base Idea**
  - Divide earth into cells *(=partitioning)*
  - Train a DNN on a classification task
- **Key for success**
  → Construction of the *partitioning*
- *s2* partitioning at multiple spatial scales
  - Leads to state-of-the-art results [Müller-Budack et al., ECCV'18]
  - Jointly learn to localize at multiple spatial scales

quad-tree hierarchy



s2 [Weyand et al., ECCV'16]

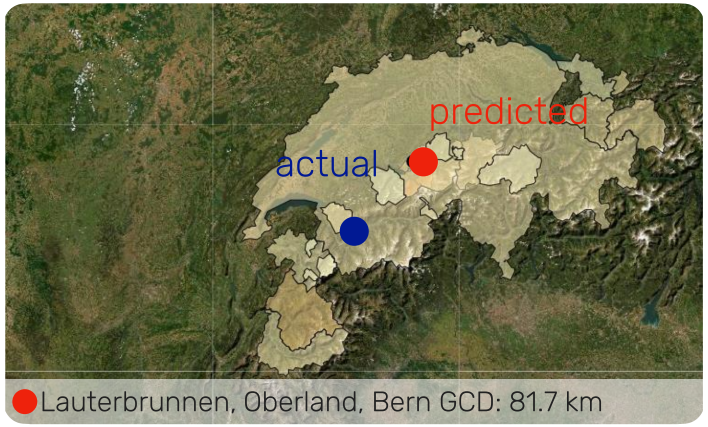
Initialized randomly; formed by data distribution

CPlaNet [Seo et al., ECCV'18]

Contributions

1. Semantic Partitioning

- We incorperate data derived from Open Street Map
  - Territorial boundaries (e.g, streets, cities, countries)
  - Natural boundaries (e.g., rivers, mountains)
  - Man-made boundaries (e.g., buildings, rails)
- More natural model's output & learning task

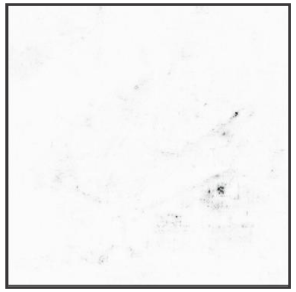


2. Investigate the post-hoc interpretability

- We measure the influence of semantic visual concepts
- Provide insights which features contribute to correct and incorrect predictions

Concept Influence

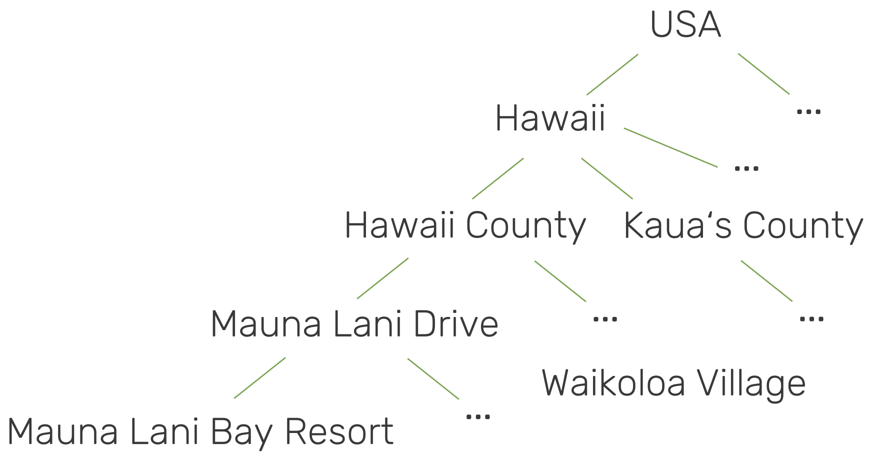
Building: 2.86
Mountain: 0.62
Hill: 0.22
Grass: 0.01

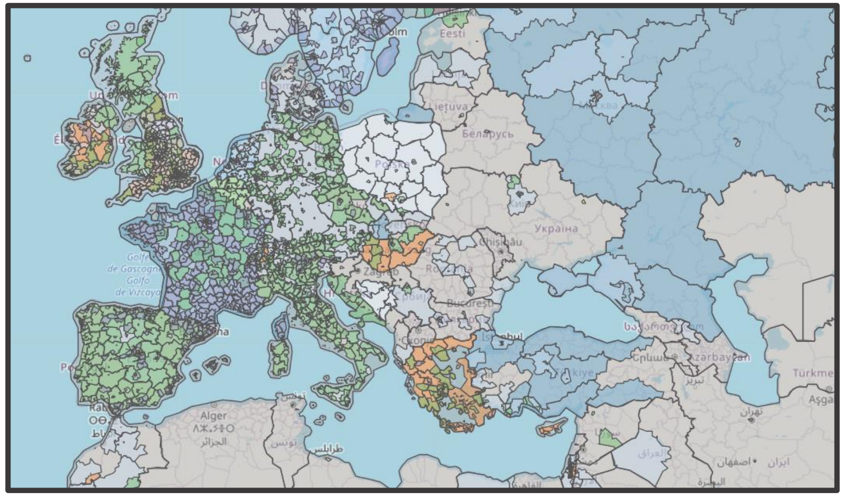


Semantic Partitioning

Construction

- Given: Dataset with GPS coordinates (+ images)
- Reverse Geocoding
  → One adress vector for each coordinate
- Construct a hierarchy of locations
  →Each location can be assigned to one parent
- Partitioning: Subset from hierarchy based on the image distribution





Semantic Partitioning

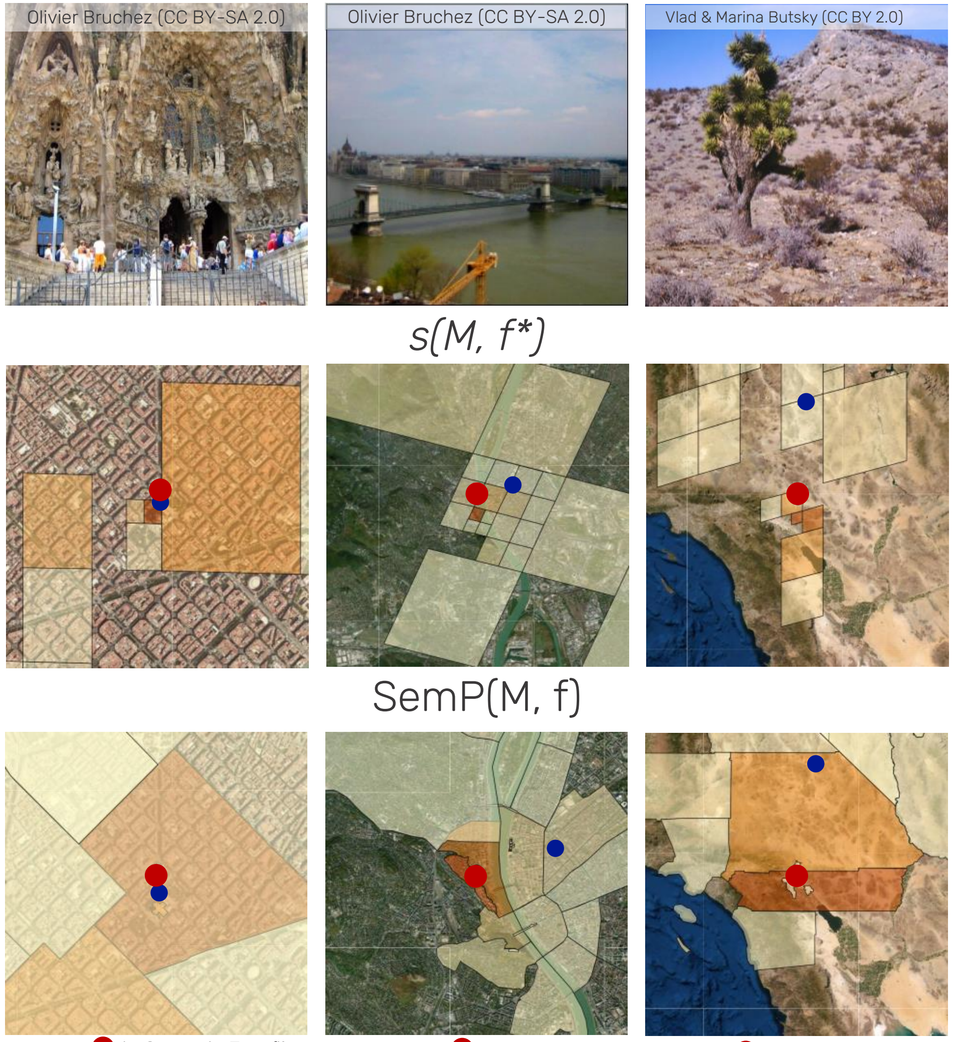
Evaluation

- Comparison with *s(M, f*)* [Müller-Budack et al., ECCV'18]
    →Yields state-of-the-art results
    →Fixed setup (same data, backbone, training, etc.)

- Similar or better results than state-of-the-art-approaches

Method	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
s2(M,f*) (repr.)	11.5 %	30.8 %	41.0 %	55.7 %	70.8 %
SemP({100, 125, 250},f)	12.5 %	**31.4 %**	**42.7 %**	**57.3 %**	**72.0 %**
SemP({50, 75, 100},f)	**13.5 %**	30.8 %	41.2 %	54.7 %	70.2 %

Results on the Im2GPS3k dataset. Percentage of images localized within given radius. Only the partitioning differs.



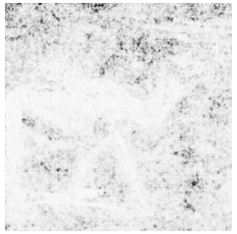
la Sagrada Família, Barcelona, Catalonia, Spain

Castle District, Budapest, Central Hungary, Hungary

Riverside County, California, USA

Concept Influence

- Given: A segmentation and explanation map for an input image



- We measure the attribution of individual concepts on model's prediction

Model Analysis

- We identified concepts which give visual cues
  - to a concrete location (e.g., tower, building, sky)
  - to a rough region (e.g., window pane, tree, flower)
- Border between specific concepts is of interest (skyline, mountain range)

References

- Weyand, T., Kostrikov, I., & Philbin, J. Planet-photo geolocation with convolutional neural networks. ECCV’16 (pp. 37-55). Springer, Amsterdam
- Seo, P. H., Weyand, T., Sim, J., & Han, B. Cplanet: Enhancing image geolocalization by combinatorial partitioning of maps. ECCV’18 (pp. 536-551). Springer, Munich
- Müller-Budack, E., Pustu-Iren, K., & Ewerth, R. Geolocation estimation of photos using a hierarchical model and scene classification. ECCV‘18 (pp. 563-579). Springer, Munich