Outperforming Humans in GeoGuessr with Deep Learning

Valdrin Nimani, Timothy Veigel, Bastian Rothenburger, Markus Fiedler

1. Problem Description

Geolocation based on images and related work

2. Dataset

Description of dataset and preprocessing

3. Methods

Implementation of different approaches

4. Results

Compare results and evaluate models

5. Conclusion

Ideas for future work

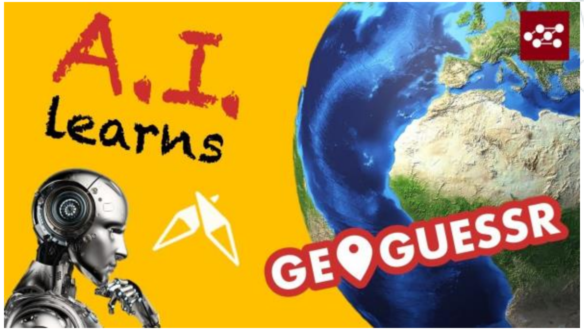
Problem Description

- Create an AI that can play GeoGuessr on the world map
- On average players score ~2000 points per round
- Current best known AI achieves ~3000 points

Improve average score to 4000 points per guess

Related Work - Traversed

- Video about an AI scoring ~3000 points in Geoguessr
- Inspired us to beat him and reach more points
- He provided us with a starting dataset

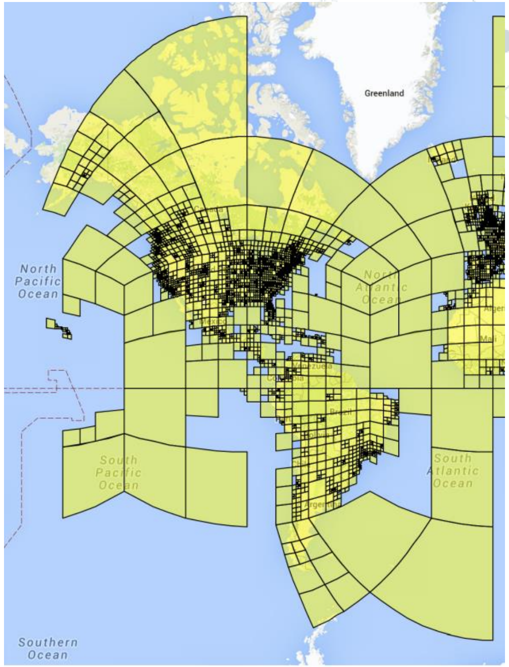


https://www.youtube.com/@TraversedTV

Related Work - PlaNet

- Geolocation of Flickr images
- 126M images mined from web
- Used a CNN, with Googles S2 cells as output
- Use LSTM to interpret images in sequence

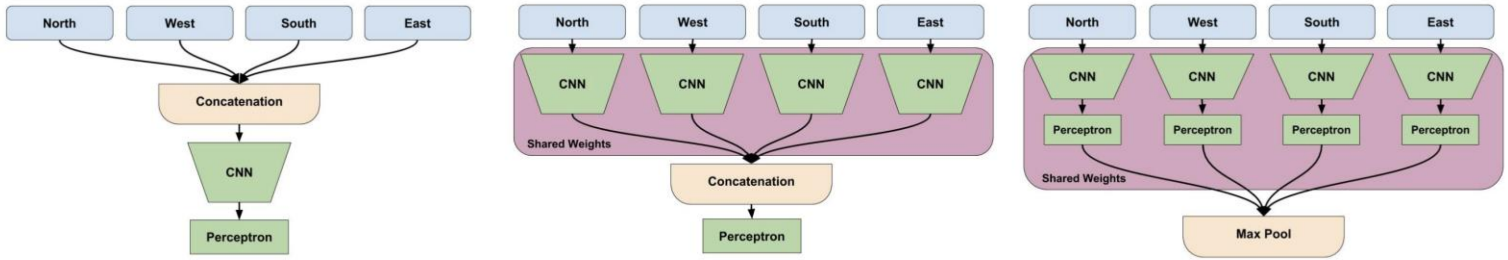
	PlaNet	Human Player
Rounds Won	**28**	**22**
Average Distance	**1131.7km**	**2320.75km**
Points	**~3000**	**~1600**



T. Weyand, I. Kostrikov, and J. Philbin, 'Planet-photogeolocation with convolutional neural networks'

Related Work - DeepGeo

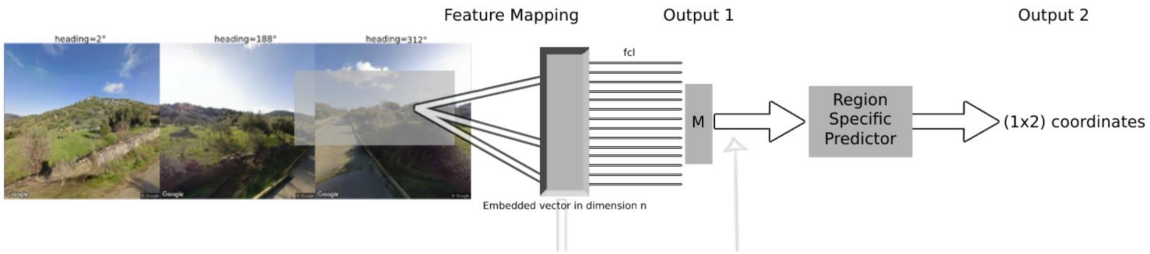
- Predict US states based on Google Streetview images
- Use ResNet with different image integration techniques
- Outperforms humans in correctly identifying US states in 4 out of 5 rounds



S. Suresh, N. Chodosh, and M. Abello, 'DeepGeo: Photo Localization with Deep Neural Network'

Related Work - Classification and Regression Approach

- Predicting the location of images from 15 chosen EU-countries
- Classification and regression on "balanced" Street-View data
- CNN model with data augmentation

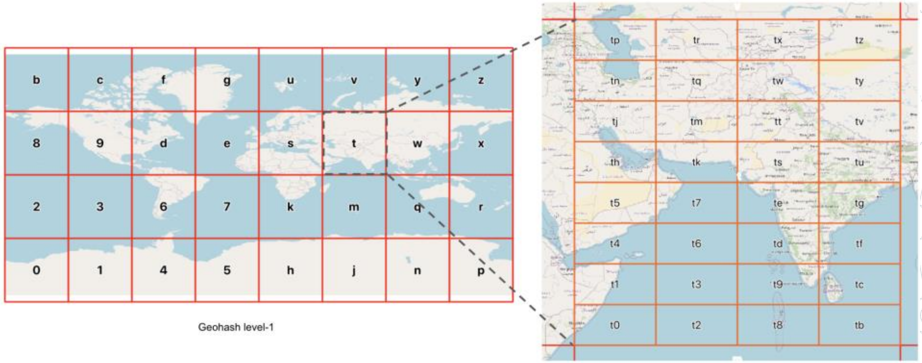


Dataset

- 128,000 images of Google Streetview locations
- Five images from different viewpoints are combined into a single image

Preprocessing

- Images are labeled with longitude and latitude
- Use geohashes to turn it into a classification problem with ~32,000 classes
- With precision 3 the cell has height and width of 156km
  → maximum error 78km



Geohash level-1

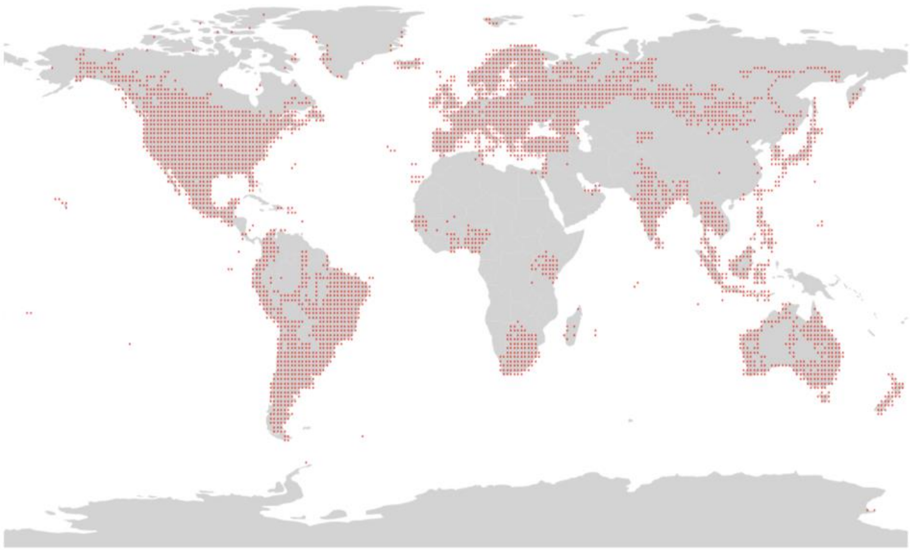
Geohash level-2

Geohash level-3

Geohash level-2

Preprocessing

- Filter out geohashes without any samples to reduce model size (3,200 classes)
- Get additional data from Google Streetview API (50,000 images)
- Add continent labels for sequential model



Regression Model

- Idea: Input image, output coordinates
- ResNet 18, Adam optimizer and Haversine-distance based loss function
- Only reached an average distance of 2500km(~1500Points) on the testset
- Suffered strongly from overfitting and converged to one single "Average guess"

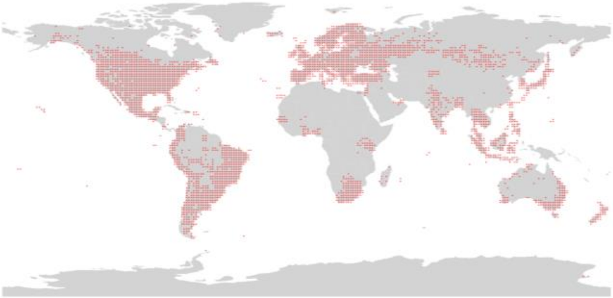
→ Abandoned due to bad inital results

End-to-End Model – Finding a suitable architecture

- Own architectures → Hard to train
- Pretrained ResNet family from pytorch
  - ResNet 18, 50, 101
- Pretrained Visiontransformer
  - ViT-Base, 16x16 patch size
- ResNet 50 had best training time for accuracy ratio
  → Benchmark for model comparison

Sequential Model

Trained on: All data
Labels: continents
Classes: 7
Accuracy: 91,3%



Trained on: North America
Num. Samples: 39767
Accuracy: 28%

Trained on: Europe
Num. Samples: 41215
Accuracy: 25,99%

Trained on: South America
Num. Samples: 11365
Accuracy: 31,55%

Trained on: Asia
Num. Samples: 25195
Accuracy: 34,37%

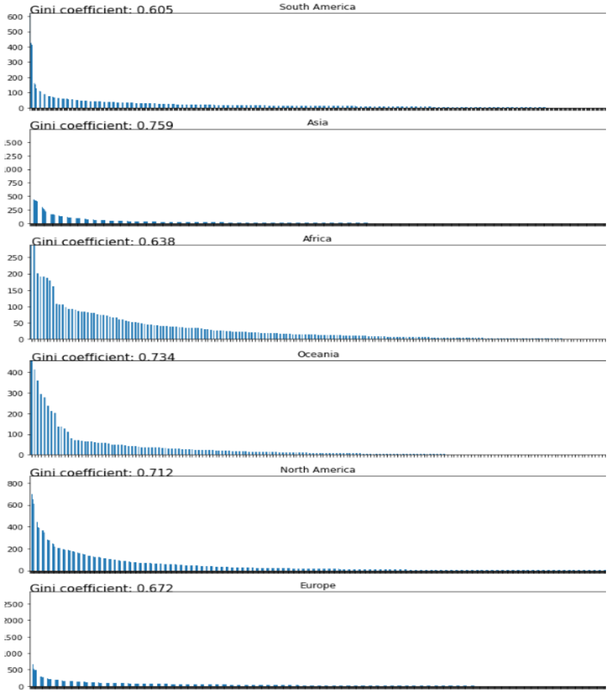
Trained on: Africa
Num. Samples: 5745
Accuracy: 51,09%

Trained on: Oceania
Num. Samples: 5472
Accuracy: 43,01%

Trained on: Antarctica
Labels: geohashes
Accuracy: /

Class Imbalance

Model	Samples	Output Cluster	Accuracy	Gini
South America	11365	540	0.3155	0.605
Asia	25195	734	0.3437	0.759
Africa	5745	183	0.5109	0.638
Oceania	5472	172	0.4301	0.743
North America	39767	779	0.28	0.712
Europe	41215	745	0.2599	0.672

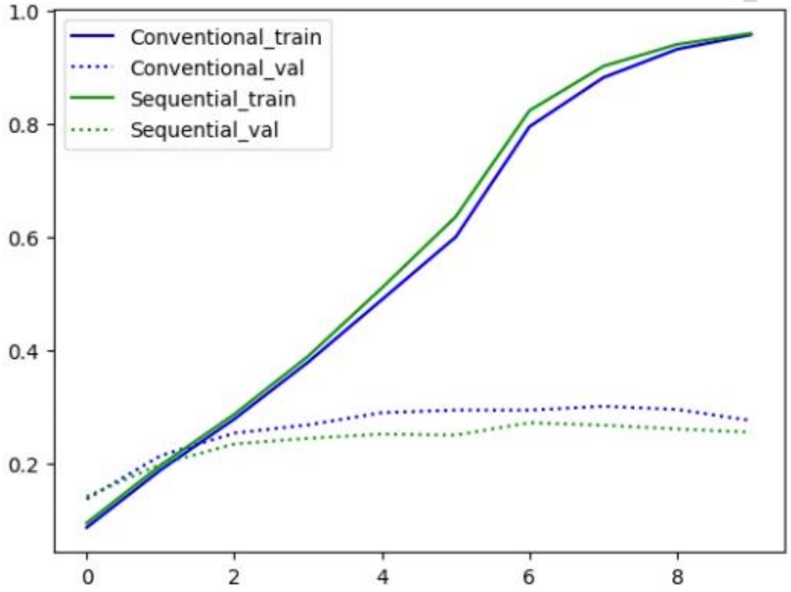


Performance

Regarding
- Accuracy
- Geoguessr Score
- Overfitting

the sequential model did not perform any better than a conventional end to end model.

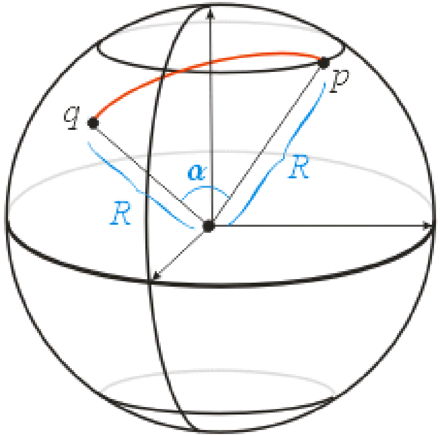
→ The approach was then discarded



Improving Results

- Problems:
  - Class imbalance
  - Overfitting
  - Cross Entropy loss ignores distance
- Solution Proposals:
  - Stratified Sampling
  - Dropout, Weight Decay, Augmentation
  - **Haversine Loss**

Modifying the Loss function: Haversine Loss



$$L = YD - \gamma ln(y_k)$$

Y are the normalized model Outputs, D is the Haversine Distance between GT and every cluster center, γ is a hyperparameter

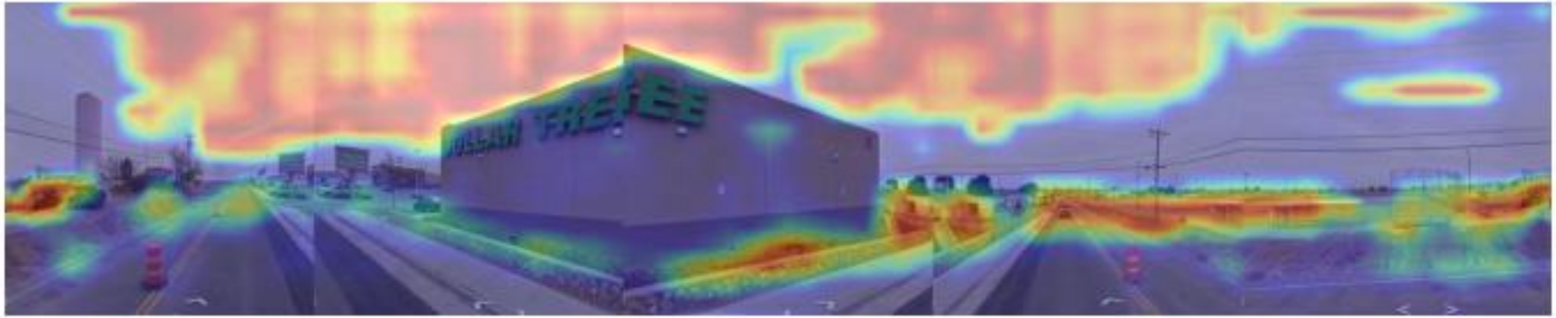
- Base: Cross Entropy Loss
- Add real distance similar to a regularization term
- Increase punishments for clusters that are far away

Results Overview

	ResNet50	Sequential ResNet50	VIT_B_16	Haversine ResNet50	Augment ResNet50	Google Data ResNet50	Haversine Augment Google ResNet50
Score Level							
0 – 1000	23,0%	22,5%	27,1%	20,7%	20,6%	19,4%	12,1%
1000 – 2000	6,5%	7,8%	5,4%	5,3%	5,6%	5,3%	5,0%
2000 – 3000	9,6%	9,7%	8,9%	9,1%	10,2%	8,7%	8,9%
3000 – 4000	16,4%	15,5%	14,8%	16,3%	15,3%	17,3%	15,2%
4000 – 5000	44,4%	44,5%	43,8%	48,6%	48,3%	49,4%	58,8%
Average Score	3010	3006	2900	3180	3163	3237	3600

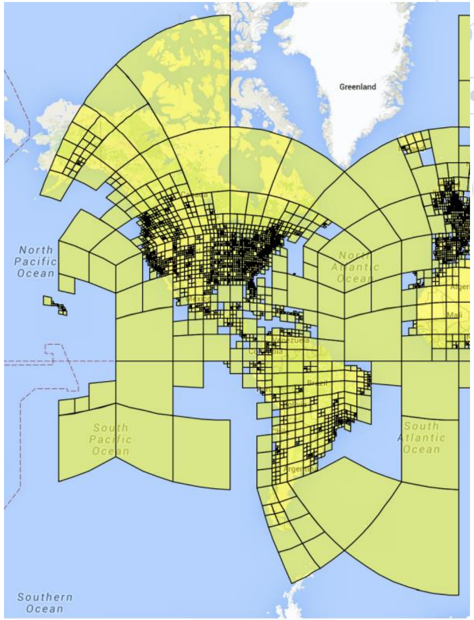
Visualizing Network - GradCAM

Visualizing Network - GradCAM



Future Work/Improvements

- Use Recurrent Neural Networks to interpret the set of images from a given location in a sequential manner
- Solving Imbalance and adaptive Cluster sizing:
  - Advanced K-Means
  - S2 Partitioning
  - Balanced Sampling from Google (~800k panoramas)
- Use regression to make more fine-grained prediction of coordinates



Sources and References

- A. Vaswani et al., ‘Attention is all you need’, Advances in neural information processing systems, vol. 30, 2017.
- T. Weyand, I. Kostrikov, and J. Philbin, ‘Planet-photo geolocation with convolutional neural networks’, in Computer Vision--ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14, 2016, pp. 37–55.
- S. Suresh, N. Chodosh, and M. Abello, ‘DeepGeo: Photo Localization with Deep Neural Network’. arXiv, 2018.
- R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, ‘Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization’, CoRR, vol. abs/1610.02391, 2016.
- Geohashes [Image]. Retrieved from:
  https://www.geospatialworld.net/blogs/polygeohasher-an-optimized-way-to-create-geohashes/

Thanks!

Wanna play a round?