EarthLoc

Astronaut Photography Localization by Indexing Earth from Space

Gabriele Berton
Politecnico di Torino

Politecnico di Torino
1859

Astronauts take (a lot of) photos of Earth

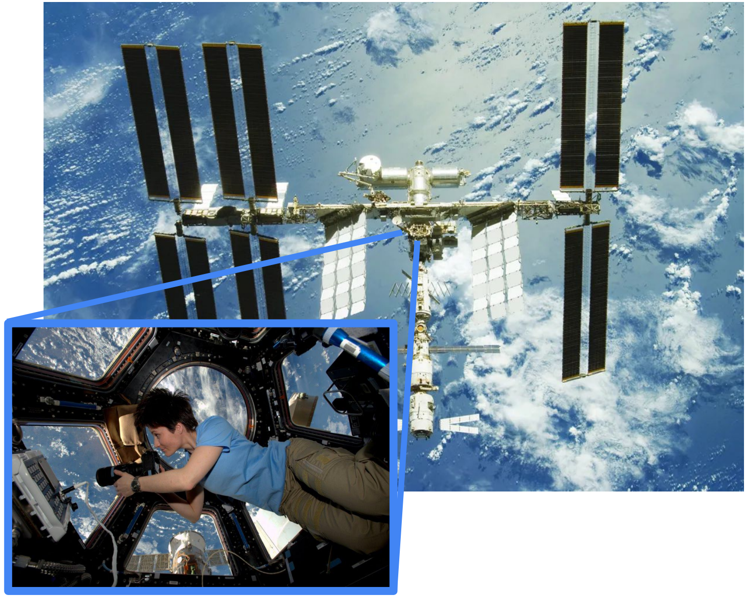
They have taken over 4M photos of Earth in 60+ years

Astronauts undergo photography training before their missions

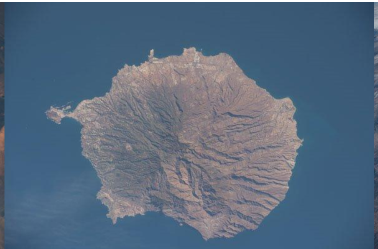
These photos are used for **scientific research** and **emergency response**

Pictures from above of wildfires, floods, volcanos

Photos are manually localized and sent to the respective authorities







Localizing an astronaut photo is difficult!

Camera position know, but orientation unknown

Visibility extends 20M km² (~ North America)

Photos extent can be 100 km² (~ Manhattan)

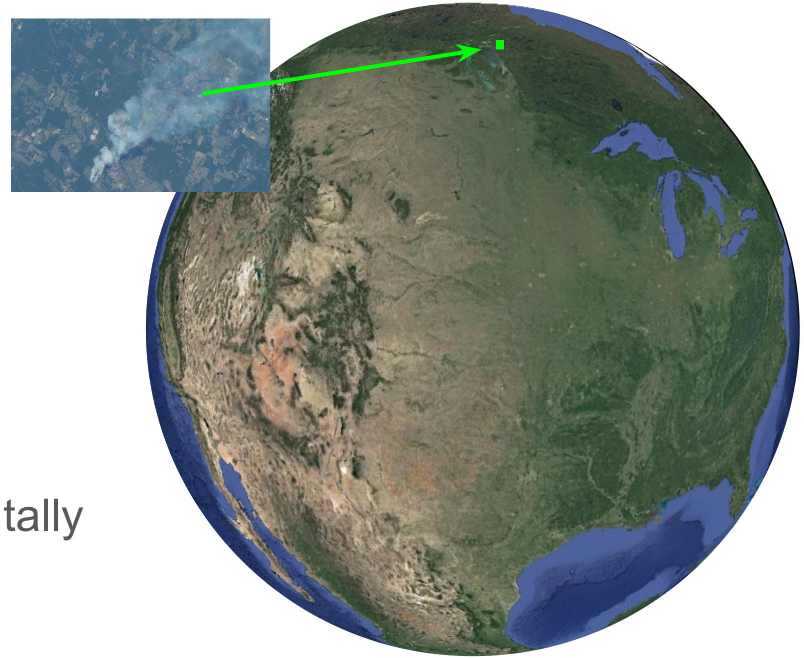
Photos cover 0.0005% of visible area

Finding a needle in a haystack

300k images have been manually localized

"It is a monumentally important, but monumentally time-consuming job" [1]

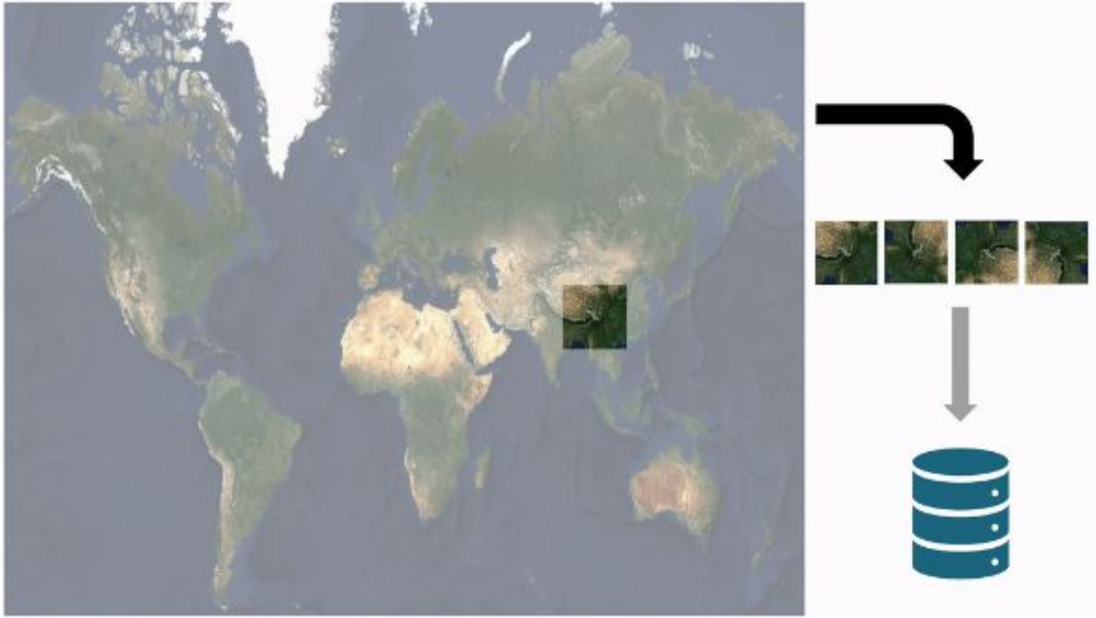
Example of area visible from the ISS



[1] Picturing Earth Behind the Scenes NASAEarthObservatory - https://www.youtube.com/watch?v=drrP_Iss0gA

Creating a database

APL requires a database that is (1) global, (2) spans the resolution range of astronaut photos, and (3) accounts for the lack of canonical orientation (no gravity in space, so no "up"!)



Database samples

We collect cloudless images at:

- 3x resolutions

- 4x rotations

- 4x years

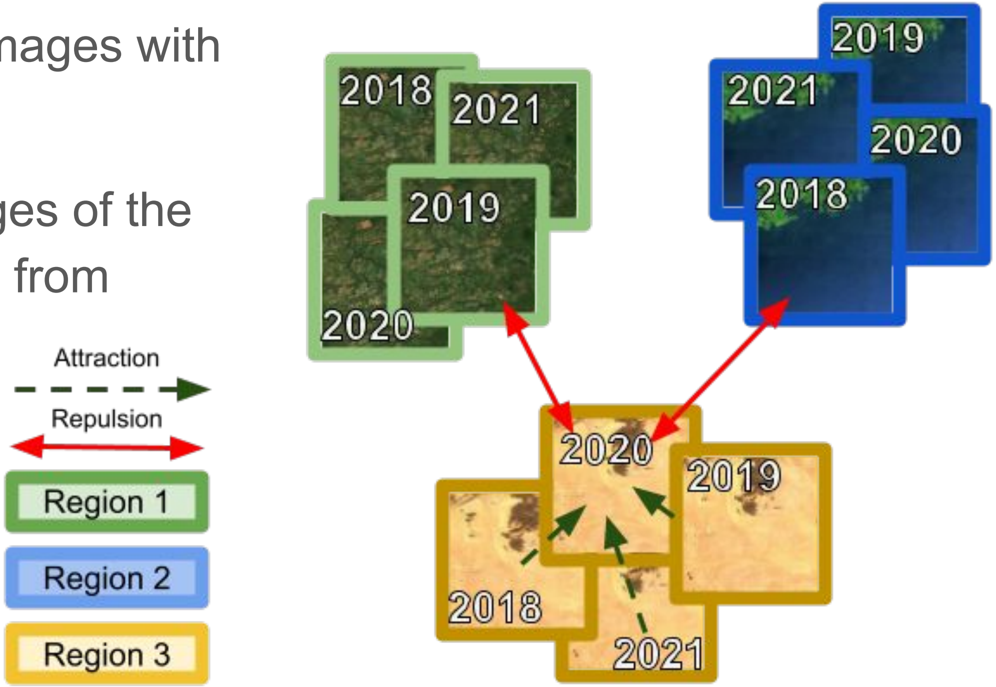
Cloud resources from ESA NoR sponsorship

EarthLoc: baseline training

EarthLoc is trained on satellite images with multi-similarity loss[2]

In each batch, pull together images of the same region, push away images from different ones

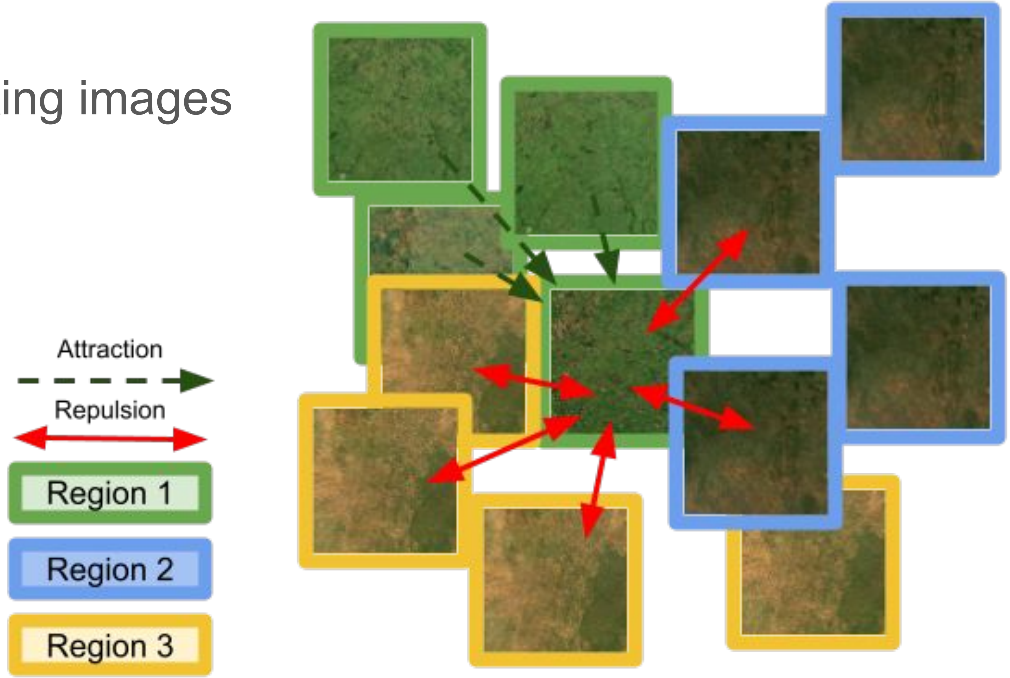
This is trivial and the loss quickly goes to 0



[2] Xun Wang et al, Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning, CVPR 2019

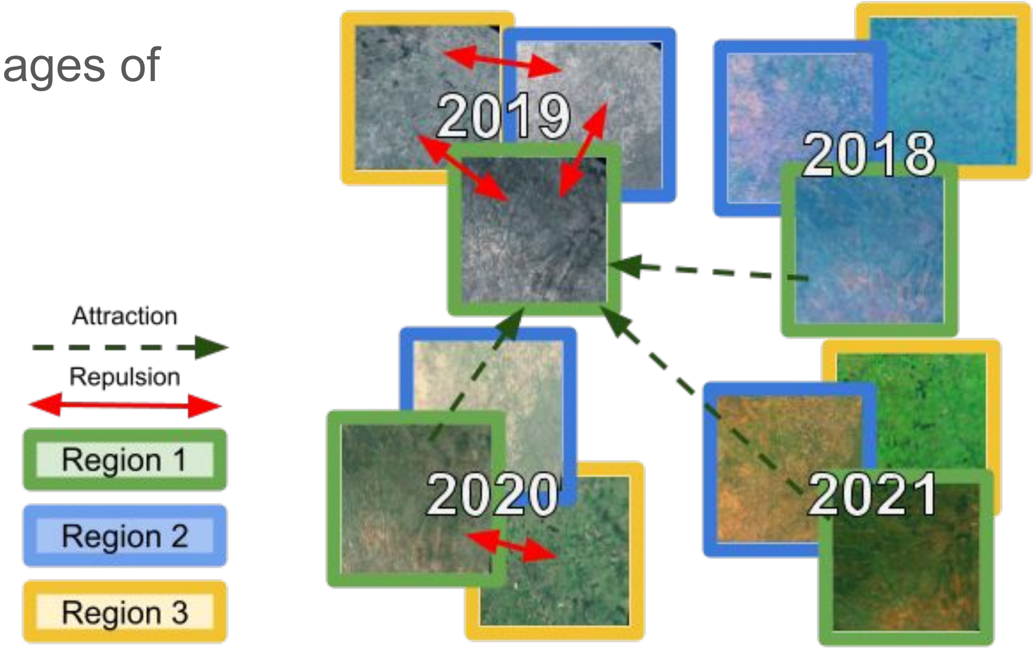
1. Training with clustered batches: + 4.4%

Each batch samples similar looking images



2. Training with year-wise augmentation: + 4.3%

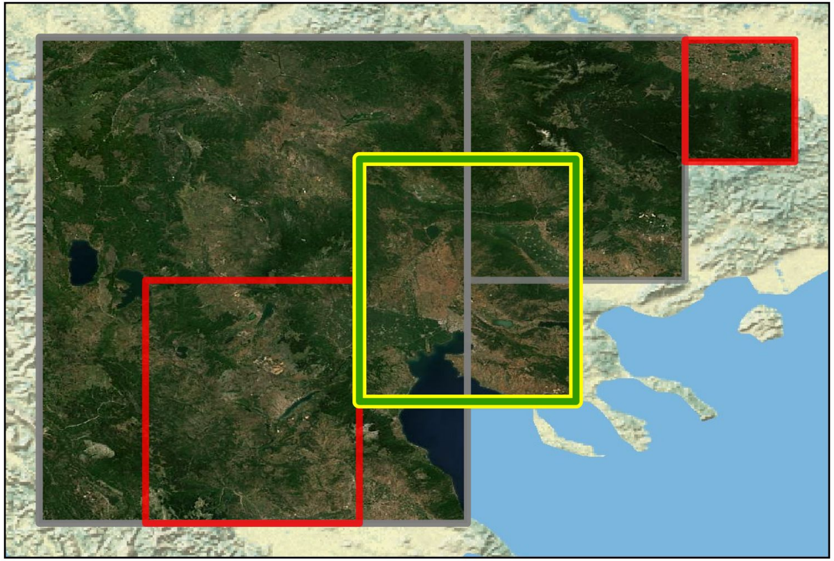
Apply same augmentation to images of different regions but same year



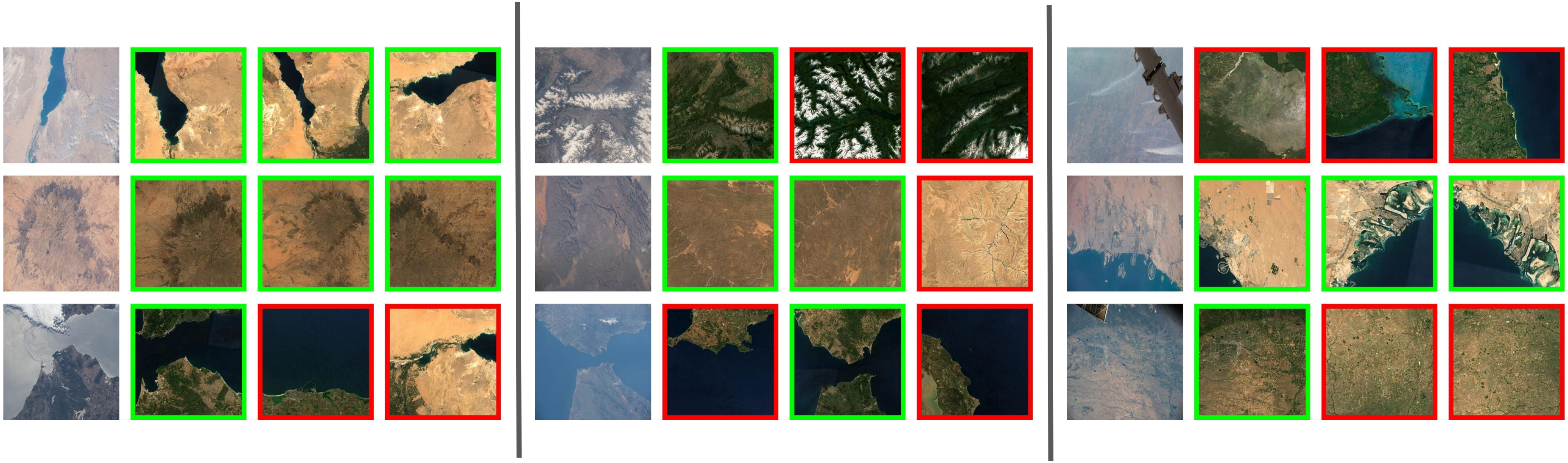
3. Training with neutral-aware multi-similarity: + 5.1%

Multi-similarity loss assumes images to be either positives or negatives.

We introduce the concept of neutrals



EarthLoc: Qualitative Results



- Typically **more than one rotation** of the same region in top 3 - shows EarthLoc learned some rotation invariance
- Most **incorrect predictions** are of visually similar regions

EarthLoc: Results

Method	Type of Training Imagery	Texas			Alps			Gobi Desert			Amazon		
		R@1	R@10	R@100	R@1	R@10	R@100	R@1	R@10	R@100	R@1	R@10	R@100
Nadir	-	2.4	-	-	1.2	-	-	1.8	-	-	3.1	-	-
Random Choice	-	0.2	1.7	15.5	0.1	1.1	11.6	0.1	1.0	13.2	0.1	1.1	11.5
AnyLoc (DINOv2 + NetVLAD)	Universal VPR	44.1	68.7	87.8	40.7	70.8	**92.0**	28.7	57.0	81.7	38.6	63.8	**86.2**
TorchGeo (ResNet50 w MOCO)	Satellite	1.0	3.2	11.6	0.3	1.3	5.7	1.9	6.6	19.4	0.7	2.9	9.9
TorchGeo (ResNet50 w SeCo)	Satellite	6.1	15.6	41.7	7.4	20.2	49.2	4.6	13.3	32.9	5.7	15.6	38.5
TorchGeo (ResNet50 w GASSL)	Satellite	9.7	22.8	46.4	9.1	23.1	50.5	6.3	17.5	45.4	8.3	20.3	40.1
OGCL UAV-View (ConvNeXt-XXLarge)	UAV	16.2	35.3	65.8	14.4	34.1	64.7	10.0	26.3	54.7	21.1	38.4	63.4
OGCL UAV-View (ViT-L/14)	UAV	17.6	33.2	55.9	14.6	33.2	63.9	7.6	22.8	50.1	20.4	39.1	62.5
MBEG (ViT-L/14)	UAV	7.0	17.6	35.1	6.6	19.3	45.9	4.4	15.0	38.0	6.4	17.1	39.3
EarthLoc (Ours)	Satellite	**55.9**	**73.0**	**88.3**	**58.4**	**76.8**	89.5	**51.1**	**67.5**	**86.5**	**47.2**	**67.9**	84.6

- achieves SOTA
- 50x faster than second best
- 10x smaller features than second best