EarthLoc: Astronaut Photography Localization by Indexing Earth from Space

Supplementary Material

7. Empirical Investigation of Challenge Modes

The astronaut photography localization task has various challenge modes, where particular photography conditions make localization more difficult for one image compared to another. In this section, we analyze the correlation between two of these conditions (challenge modes) and performance.

Distance from Nadir As the distance between the ISS nadir point and photo location increases, so too does obliquity and shear due to the imaging geometry. Oblique imagery is often taken through a thicker column of atmosphere, adding blurriness to the image. Thus, we expect some performance drop to accompany increasing distance, as seen beyond 400km in Fig. 8. Further augmentation to satellite images during training can potentially close this gap, particularly if these augmentations are designed to simulate conditions seen in far-from-nadir imagery.

Area We next analyze the correlation between an astronaut photo's geographic area encompassed (area) and recall (Fig. 9). Here, recall decreases as area increases. Based on our training set construction (see Sec. 3.2), there are fewer images with larger areas required to cover the extent (land area between $\pm 60°$ latitude), so fewer such images are included in training. We expect including more such images during training will improve performance for this type of imagery.

8. Effect of 4x90 TTA

We also study the impact of our *4x90TTA* strategy, with results reported in Tab. 4. In each evaluation set, performance improves due to *4x90TTA*. Based on orientation information from the Gateway to Astronaut Photography of Earth, there are approximate the same number of images across orientation angles in [0°,360°). However, the performance does not increase in proportion to the TTA (i.e., by 4x), indicating that EarthLoc learns some rotation invariance during training. This observation is further supported by multiple rotations of the same image present in the top 3 predictions, examples of which can be seen in Fig. 13. Note that even in such cases, the database image with the closest orientation is usually the top prediction, followed by correct database images with other orientations.

Despite EarthLoc's ability to retrieve correct images with other orientations, the ablation experiment in Tab. 4 shows that this capability is limited, and that *4x90TTA* significantly boosts recall on all evaluation sets.

9. Features visualizations

Given the domain gap existing between database and queries images, we show in Fig. 10 the distribution of their features, as extracted by EarthLoc, through a T-SNE.

	Recall@1					
Test Time Aug.	Texas	Alps	California	Gobi	Amazon	Toshka
None	32.6	30.9	31.1	30.4	27.2	39.1
4x90TTA	**54.6**	**53.9**	**55.9**	**46.8**	**45.6**	**67.6**

Table 4. **Ablation on *4x90TTA*.** Performance of EarthLoc on each evaluation set with and without *4x90TTA*. *4x90TTA* approximately doubles Recall@1 across the sets.

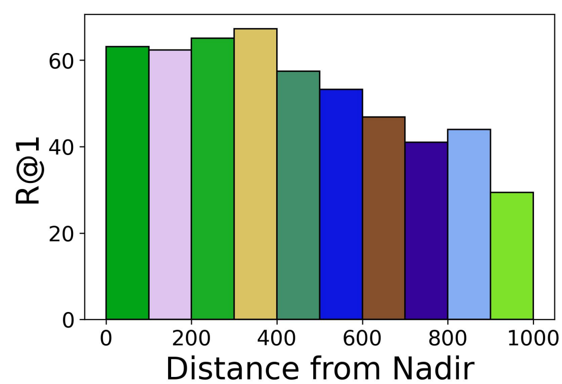

Figure 8. **Recall@1 vs Distance from Nadir (km).** Performance of EarthLoc as distance from nadir increases. Larger distances are more challenging due to obliquity effects, and this is reflected in the drop in performance.

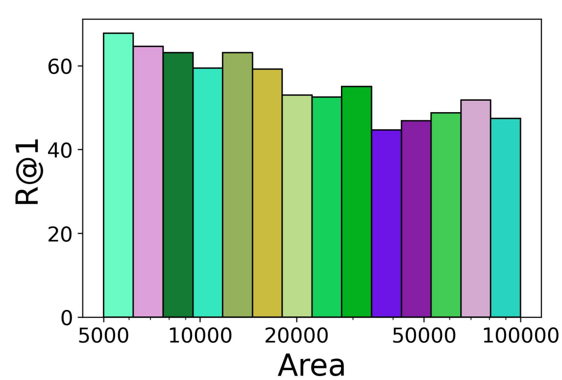

Figure 9. **Recall@1 vs Area (sq. km).** Performance of EarthLoc on astronaut photo queries with different areas. Decrease in performance for larger area photos is attributed to lower quantities of training data for this regime.

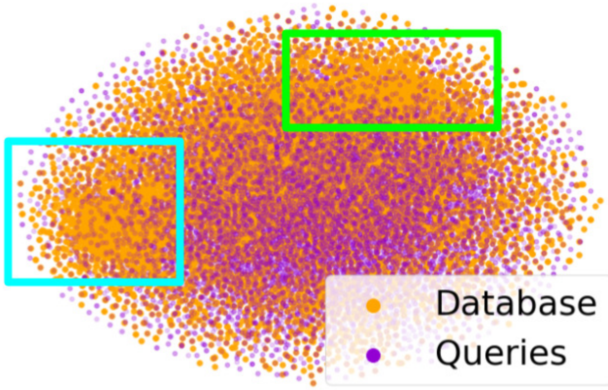

Figure 10. **T-SNE representation.** The two colored boxes represent geography types with few queries (areas rarely photographed by astronauts), namely open sea (light blue box) and plains (green).

10. Further Examples of Queries and Database

We present additional examples of astronaut photo queries (Fig. 11) and satellite database images (Fig. 12).

These example queries illustrate the variety inherent in astronaut photography. Some images have occlusion due to spacecraft hardware, some images contain clouds, and all have different orientations with respect to North. There is significant geographic and scale variation, with some photos highlighting neighborhoods of large cities and others showing entire lakes or vast mountain ranges.

The satellite imagery is more regular. In addition to uniform orientation, these images are algorithmically post-processed and thus share similar characteristics, which are somewhat different from those of astronaut photography. Though not all satellite imagery is cloudfree, this particular set is constructed to minimize clouds, and consequently few clouds are seen in the example images.

11. Qualitative Results

11.1. Examples of Queries and Predictions

Example queries and associated top 3 predictions from EarthLoc are in Fig. 13. Often, multiple orientations of the same database image are within the top 3. In other cases, correct database images with different areas (scale) are retrieved.

In examples where no correct database image is retrieved, similar looking images are often returned (left, third from bottom). In some failure cases, however, predictions do not show significant similarity to the query (bottom left). We have not found any unifying characteristics in such scenarios.

11.2. Examples of Clustered Batches

As described in Sec. 4.2, we construct batches from clusters to facilitate training. Samples of such clusters are in Fig. 14. Clusters are built by collecting *regions* that have similar representations in feature space, despite being from potentially far reaching places on Earth. These similar representations often correspond to shared characteristics, and we can assign high level labels to clusters, like “rivers in the forest” to the top cluster and “mountainous deserts” for the second.

During training, the loss function works to separate the representations for different classes/*regions* that are batched together, so building batches from clusters provides a much more challenging optimization task than random batching, as intra-cluster (across the row) images are much more similar than inter-cluster images (down the columns).

11.3. Examples of Year-Wise Data Augmentation

We show a subset of a training batch as it is presented to the model (i.e., after augmentation) in Fig. 15. This illustrates our Year-Wise Data Augmentation. Each half-row (4 images) is a *region* quadruplet, with one image from each of the years 2018, 2019, 2020, and 2021. Columns are arranged by year (i.e., each column contains images from the same year). According to our Year-Wise Data Augmentation scheme (see Sec. 4.3), images from the same year (i.e., in each column) have received the same augmentation. Augmentations are color jittering, random perspective, and random rotation.

[9] https://eol.jsc.nasa.gov
[10] https://s2maps.eu

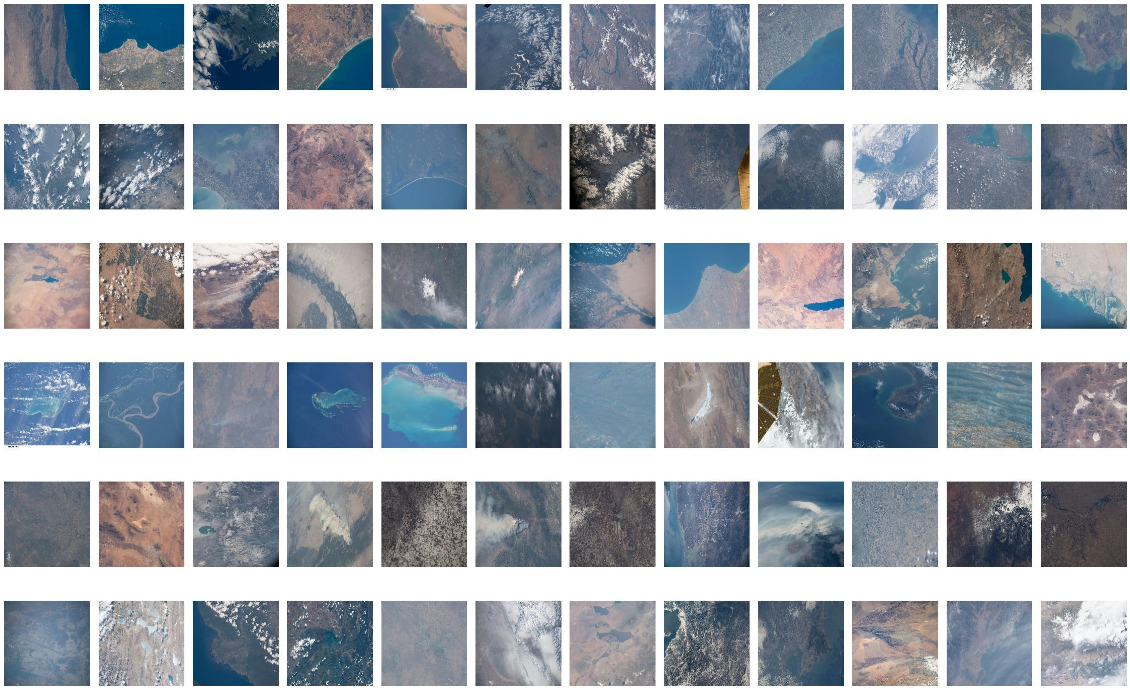

Figure 11. **Random examples of queries.** All queries come from the Gateway to Astronaut Photography of Earth collection [9]. Each row is a randomly selected set of queries from each test set, respectively being Alps, Texas, Toshka Lakes, Amazon, California, Gobi.

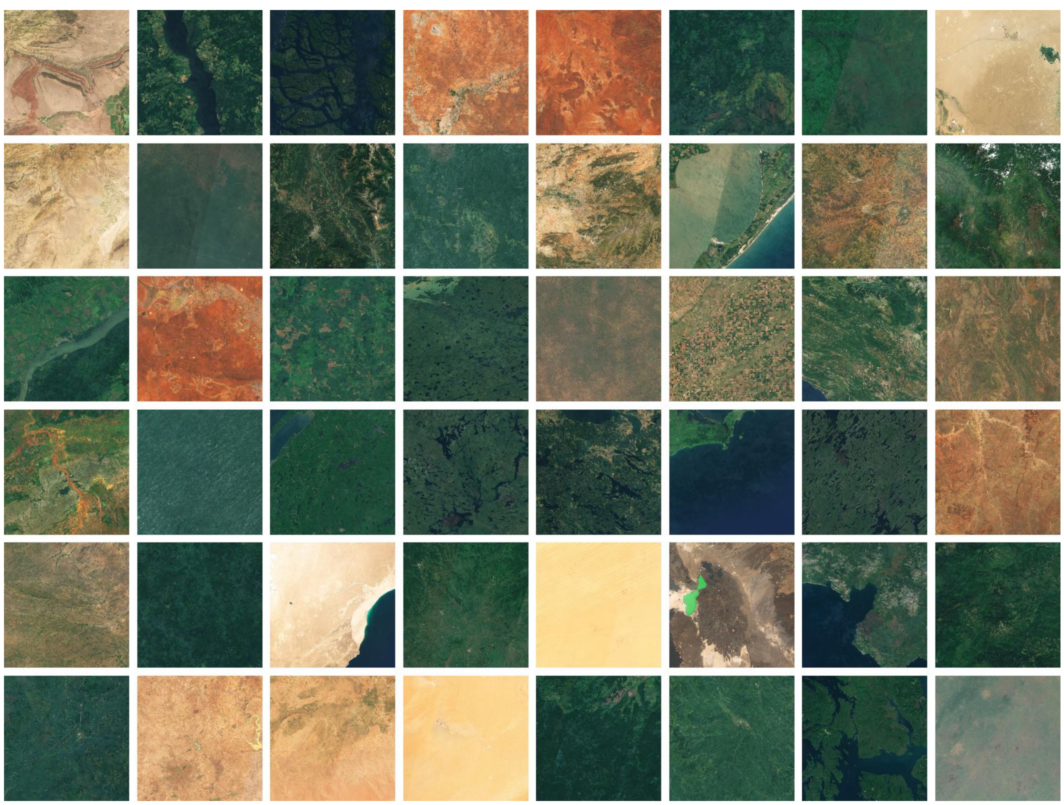

Figure 12. **Random examples of database images.** All database images come from the EOX Sentinel-2 cloudless collection.[10]

Figure 13. **Qualitative Examples of EarthLoc Predictions.** Query image and top 3 predictions. Green indicates a correct prediction, red an incorrect prediction. Each half-row is a separate example.

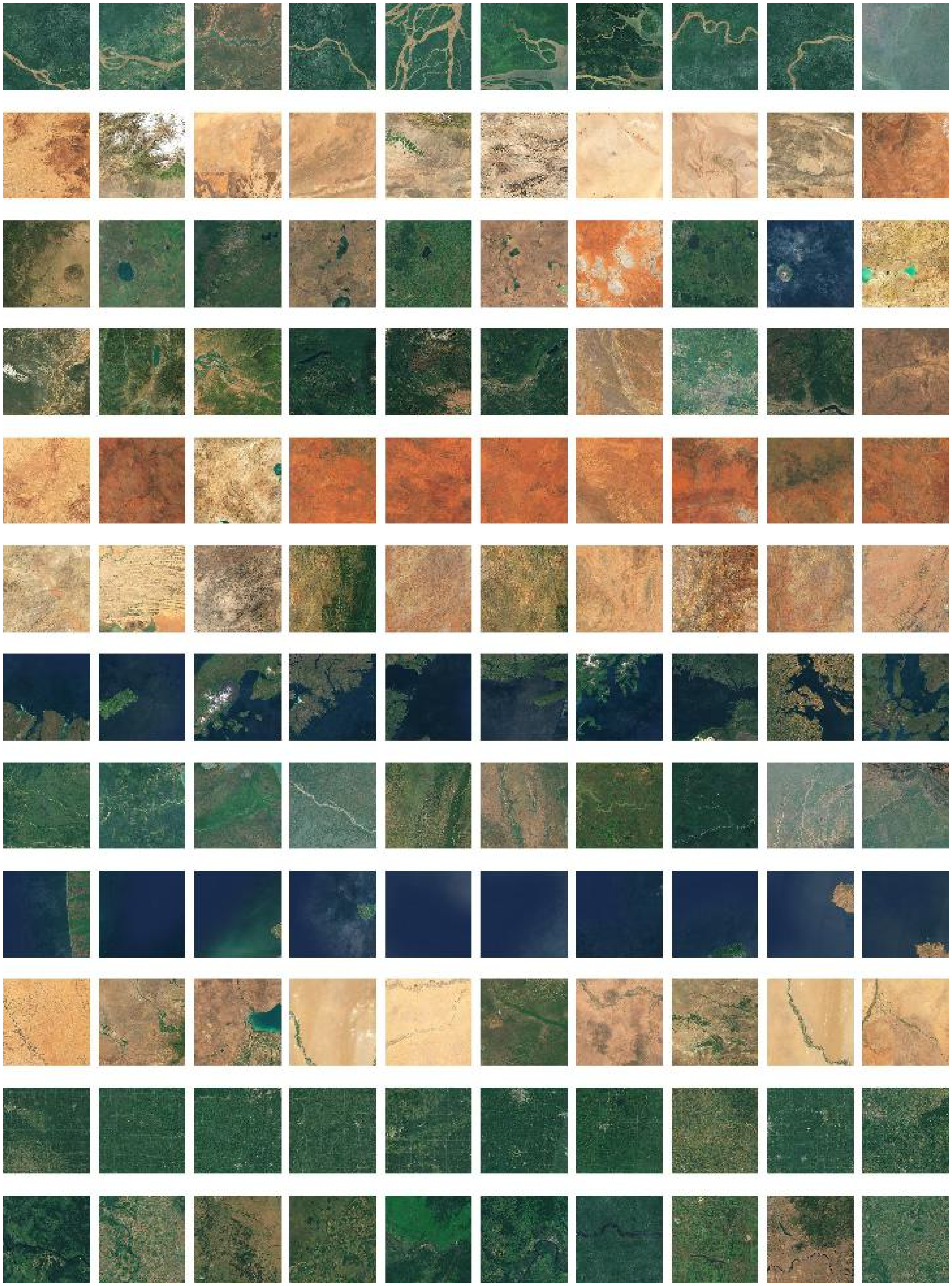

Figure 14. **Examples of clusters, one cluster per row.** Clusters are formed from images with similar representations. Training batches are selected from images within one cluster (see Sec. 4.2).

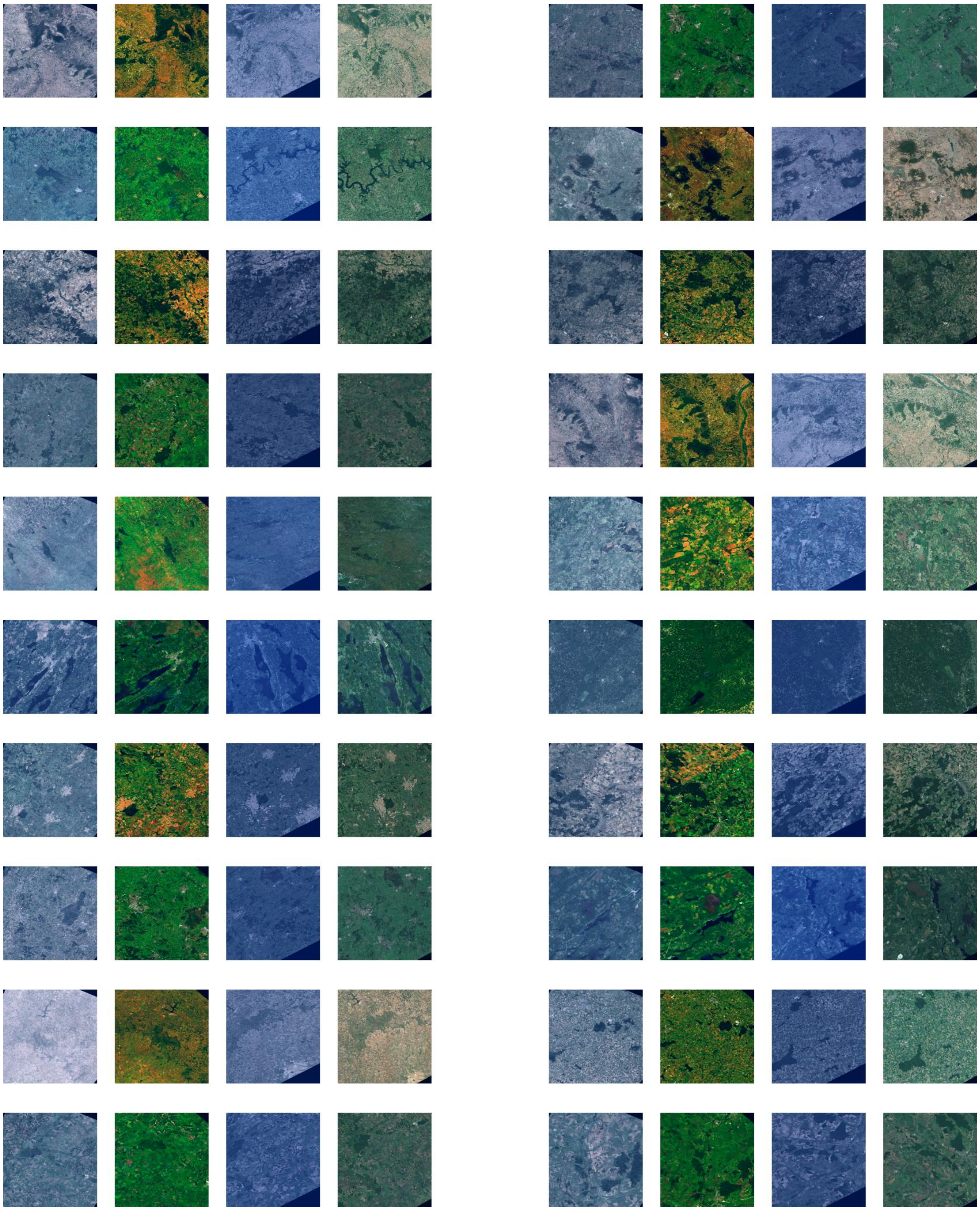

Figure 15. **An example of a batch**, showing 20 out of 32 quadruplets due to space limits. Each half-row of 4 images represents the quadruplet from one *region*. Each *region* has images from 2018, 2019, 2020 and 2021. The same augmentation is applied to all images from a given year, e.g. the images from 2020 receive a blueish color transformation in this batch.