PNAS

Supporting Information for

AI-accelerated Nazca survey nearly doubles the number of known figurative geoglyphs and sheds light on their purpose

Masato Sakai[1]*, Akihisa Sakurai[1], Siyuan Lu[1], Jorge Olano[2], Conrad M. Albrecht[1,3], Hendrik F. Hamann[4], Marcus Freitag[4]*

[1]Yamagata University; Yamagata-shi, Yamagata, 990-8560, Japan.
[2]Archéologie des Amériques, Université Paris 1 Panthéon-Sorbonne; Paris, 75004, France.
[3]German Aerospace Center (DLR); Oberpfaffenhofen, 82234, Germany.
[4]IBM Thomas J. Watson Research Center; Yorktown Heights, New York, 10598, USA.

*Masato Sakai, Faculty of Literature and Social Sciences, Yamagata University; Yamagata-shi, Yamagata, 990-8560, Japan.
*Marcus Freitag, IBM Thomas J. Watson Research Center; Yorktown Heights, New York, 10598, USA.

Email: M.S.: sakai@human.kj.yamagata-u.ac.jp; M.F.: mfreitag@us.ibm.com

This PDF file includes:

Supporting Information Text

Geoglyph construction

The Nazca Pampa is covered with pebbles, which have been exposed to the sun for many years, and their surfaces have turned dark brown, probably due to rock varnish *(1).* Beneath these pebbles is a layer of white sandy soil and pebbles. Therefore, when the dark pebbles on the surface are removed, the ground is exposed generating contrast. Line-type figurative geoglyphs, as well as the linear geometric geoglyphs, were drawn by removing dark brown pebbles from the ground surface in a linear pattern. Relief-type geoglyphs were created by removing dark surface stones from certain parts of the motif, making them appear in a lighter color than the untouched areas. Sometimes removed stones were re-used and piled up on other parts of the motif, creating an even darker and slightly elevated surface (relief).

Archaeological drawings based on drone images of newly discovered geoglyphs

Fig. S1 shows 20 examples of newly discovered relief-type figurative geoglyphs from our AI-assisted survey, each in three versions (Fig. S1A: drone images, Fig. S1B: drone images with outlines as a guide to the eye, and Fig. S1C: archaeological interpretation of the relief). In the archaeological drawings (Fig. S1C) areas with removed surface stones are depicted as white, untouched surfaces as light brown, and elevated, dark areas are illustrated in dark brown.

Giant linear/trapezoidal network

There exists a giant linear/trapezoidal network extending from the Ingenio River Valley at the northern end of the Nazca Pampa all the way to the Nazca River Valley about 15 km to the south *(2).* This network consists of 1,335 straight lines with 165 line centers as nodes *(3).* It has approximately 10 entrances/exits at the northern end, where line-type figurative geoglyphs can be found near trapezoids or straight lines and another approximately 10 entrances/exits at the southern end of the Nazca Pampa. Two of the latter entrances/exits consist of 40-meter-wide roads, each with a line-type figurative geoglyph nearby, so both are considered major entrances/exits.

AI-assisted workflow

Fig. S2 shows a flowchart of the AI-assisted geoglyph detection as described in the Materials and Methods section.

Details of the artificial neural network

Our deep learning model utilizes gridded image classification with relatively small 112x112 pixel image patches (11x11 m^2) and 5 m pitch, rather than object detection, where the model tries to find instances (bounding boxes) of objects (geoglyphs) in larger scenes. We deviated from pure object detection algorithms as applied in *(4)*, by turning the problem of finding new geoglyphs into a gridded classification task, because: (a) archaeological workloads do not require near real-time model inference, thus we can afford slightly longer model runtimes, (b) precise bounding boxes are of little value for geoglyph detection, (c) we are severely restricted by the limited number of known figurative geoglyphs for training. By turning the problem into a classification task, each

training geoglyph is cut into multiple pieces, individually represented in the training set. The approach naturally augments the size of training samples.

The convolutional neural network consists of a ResNet50 feature extractor *(5),* followed by a 2-layer fully connected classifier. We set the batch size to 128. Feature extractor layers are pre-trained on ImageNet *(6)* and their weights are frozen during the initial 190 training epochs (see Fig. S3). During this warmup the geoglyph classifier is trained. The following 50 epochs are dedicated to optimizing both the feature extractor and classifier weights to obtain the final relief-type geoglyph detection model. A focal loss *(7,8),* helps model optimization on imbalanced binary classification datasets. An AdamW optimizer *(9)* with weight decay acts as regularization to counteract model overfitting *(10).* Additionally, we apply a learning rate decay to improve the stochastic gradient descent optimization.

Thirty-three relief-type geoglyphs in the validation set assisted in tuning deep learning hyperparameters *(11)* such as the learning rate, ratio of positive-to-negative training samples, and early stopping based on a given validation accuracy score. Fig. S3 depicts the training loss, training accuracy, learning rate decay, and validation accuracy over number of training epochs. Training and validation accuracies jump after the feature extractor weights are allowed to be updated. The validation accuracy reaches a peak after 11 more epochs. Beyond that, the model starts overfitting on the training set resulting in decay of the validation accuracy. We exploit this characteristic behavior to apply early stopping to yield best model performance.

Because we employ validation accuracy to inform hyperparameter tuning and early stopping, we conducted a separate model run exclusively for testing purposes. Here we not only held out a validation set of 33 known geoglyphs, but also a testing set of 84 known geoglyphs in a 12 km2 area in the central Nazca Pampa. The testing set does not enter the training phase of the model. (Model prediction on the continuous grid is based on the model run without any held-out testing set). Depending on the adjustable parameters N and P, we compute the following geoglyph classification metrics *(12)*: recall for model-missed geoglyphs, precision to quantify model candidates incorrectly identified as geoglyph, and the F1 metric, the harmonic mean of precision and recall (Table S1). Since this “testing” model run is handicapped by the large held-out testing set, the reported metrics are lower bounds for the final model run. The model utilized for the newly discovered geoglyphs built upon a total of 368 known relief-type geoglyphs plus 33 for validation. According to the best F1 metric in the “testing” model run, we fixed the hyperparameters of the geoglyph AI model to N=2 and P=0.55.

Postprocessing of the modeled probability map

The post processing from the grid of classifier-assigned probabilities to geoglyph candidate boxes that can easily be interpreted by humans during screening is illustrated in Figs. 3a, b. We apply a simple, rule-based algorithm with threshold probability P=0.55 to focus on the most promising candidates. For a given patch to classify as geoglyph, at least N=2 neighboring patches must concur in predicting the geoglyph class. Both P and N are adjustable hyperparameters with final choice made after testing results are obtained as detailed above. An example of the post-processing result is provided in Fig. 3B.

To avoid missed detections of geoglyphs our AI model is tuned for a high recall-precision ratio of 14 (Table S1). Due to the enormous size of the Nazca Pampa, there are many more geoglyph candidates flagged by the AI model (47,410 in total) than can be surveyed in the field. However, the AI model eliminates more than 98% of the aerial imagery so that visual inspection can focus on the remaining 2% — a reduction of Big Geospatial Data by a factor of about 50. The limiting factor in achieving better recall and precision is the tiny set of 401 previously known relief-type figurative geoglyphs from the Nazca and Palpa areas used for model training, which are also quite diverse in what they depict. Reasons why the AI may not flag all the geoglyphs are: (a) the optical contrast is lower than the average training geoglyph’s because it is more eroded, or the

color contrast has faded. (b) The angle of the sun was not favorable for capturing a geoglyph. (c) The AI model is expected to miss some geoglyphs even under good viewing conditions due to tuning of the model (Table S1) to a recall of 0.35 on the held-out testing set.

Examples of newly discovered geoglyph groups.

Fig. S4 exemplifies four newly discovered geoglyph groups and their relation to the AI model geoglyph candidate boxes. Geoglyphs that overlap with the AI model candidate boxes are shown in green. Additional geoglyphs that were found during the field survey within the groups are shown in cyan.

Newly discovered geometric geoglyphs.

In addition to the 303 figurative geoglyphs, we discovered 42 geometric and 2 unidentified geoglyphs during the AI-assisted study. Geometric geoglyphs are much more numerous on the Nazca Pampa than figurative geoglyphs, and the 42 discovered here are a small fraction of the approximately 1500 known geometric geoglyphs on the Pampa. Thus, the discussion in this paper focuses mostly on the figurative ones. For completeness, we present the newly discovered geometric geoglyphs in Fig. S5 and Table S3. From 341 candidates targeted for a visit we were able to confirm 40 (41 if accounting for geoglyph groups) AI suggestions as authentic geometric geoglyphs. The 2022/23 field survey identified an additional non-AI new geoglyph for a total of 42 new geometric geoglyphs. Extrapolated success rates of the three candidate ranks suggest 105 new geometric geoglyphs will be discovered by the end of the field surveys.

Supporting Information References.

1 R. I. Dorn, Rock varnish. Am. Sci. 79, 542–553 (1991).

2 M. Sakai, Nasca Geoglyphs: Spatial Distribution Rule, Purpose of Construction, and Protection (in Japanese) (Embassy of the Republic of Peru in Japan, Tokyo, 2023).

3 M. Sakai, J. Olano, H. Takahashi, Centros de líneas y cerámica en las Pampas de Nasca, Perú, hasta el año 2018 (Yamagata University Press, Yamagata, 2019).

4 M. Sakai et al., Accelerating the discovery of new Nasca geoglyphs using deep learning. J. Archaeol. Sci. 155, 105777 (2023), 10.1016/j.jas.2023.105777.

5 K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016), 10.1109/CVPR.2016.90.

6 J. Deng et al., “ImageNet: A large-scale hierarchical image database” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2009), 10.1109/CVPR.2009.5206848.

7 T.-Y. Lin et al., “Focal loss for dense object detection” in Proceedings of the IEEE International Conference on Computer Vision (2017), pp. 2999–3007, 10.1109/ICCV.2017.324.

8 Carwin, Focal loss for dense object detection in PyTorch. GitHub Source Code (2017). https://github.com/clcarwin/focal_loss_pytorch.

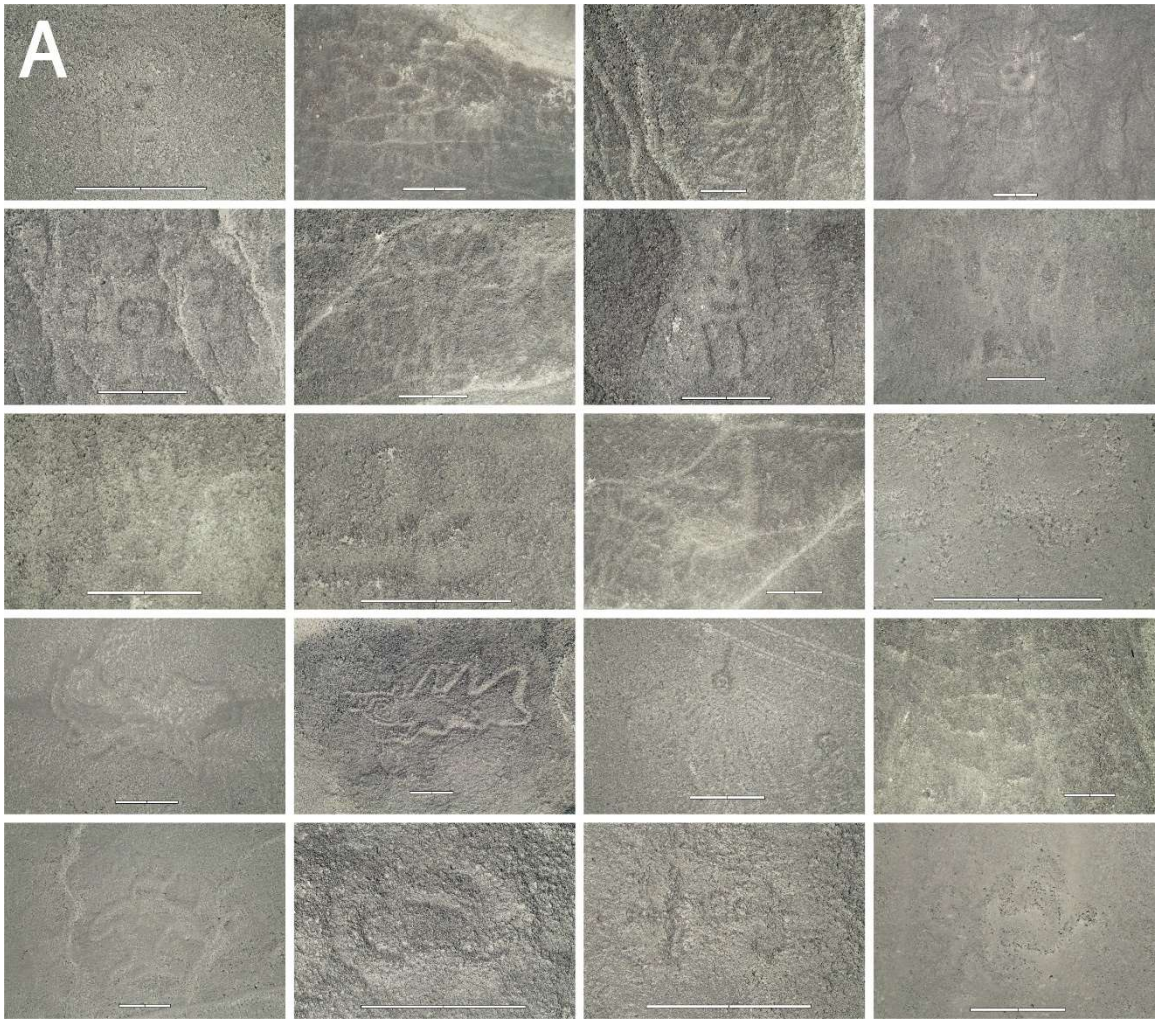
9 I. Loshchilov, F. Hutter, Decoupled weight decay regularization. arXiv [Preprint] (2017). https://doi.org/10.48550/arXiv.1711.05101.

10 M. M. Bejani, M. Ghatee, A systematic review on overfitting control in shallow and deep neural networks. Artif. Intell. Rev. 54, 6391–6438 (2021), 10.1007/s10462-021-09975-1.

11 L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316 (2020), 10.1016/j.neucom.2020.07.061.

12 M. Hossin, M. N. Sulaiman, A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process 5, 1–11 (2015), 10.5121/ijdkp.2015.5201.

Fig. S1.
Examples of newly discovered relief-type geoglyphs from the AI-assisted survey.
(A) Drone images taken during the field survey that confirmed the geoglyphs as authentic. **(B)** Images with outlines as a guide to the eye. **(C)** Archaeological interpretation of the relief (white: stones removed; dark brown: stones piled up; light brown: original surface). Scale bars: 5m.
(This figures spans 3 pages)




B


C

Figure S2

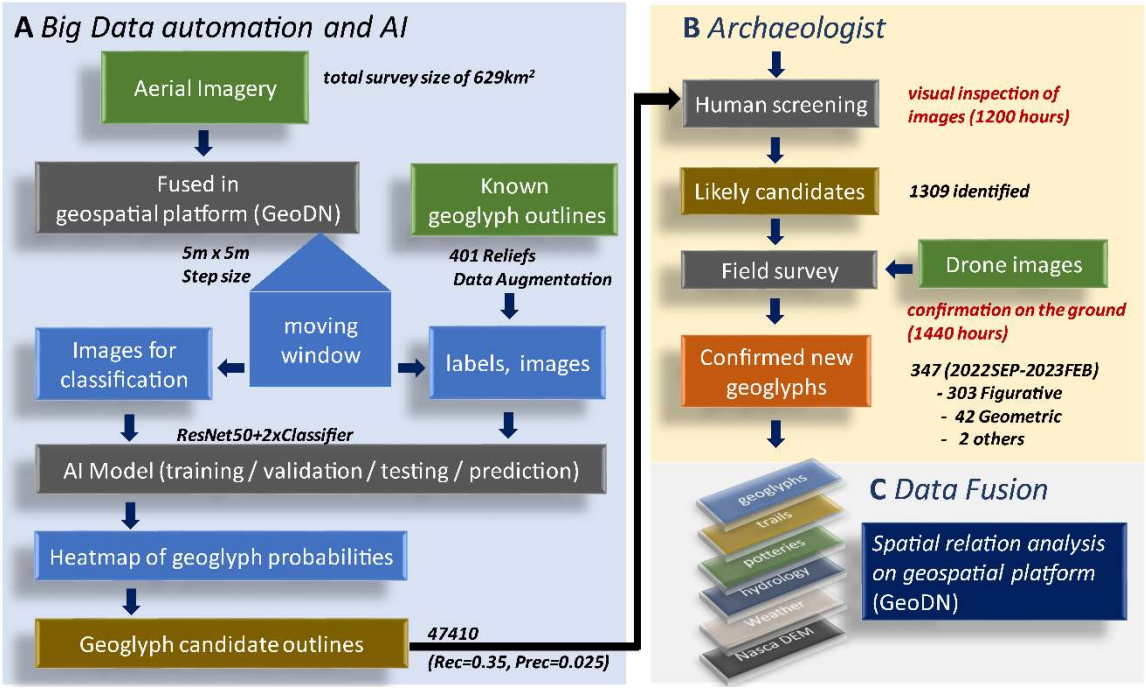


Fig. S2. Flowchart of the geospatial analysis system to detect new geoglyphs. (A) Big Data automation and AI drive the generation of geoglyph candidates. **(B)** Subsequently, geoglyph candidate images are visually inspected to focus on the most likely candidates, which are confirmed or rejected in a field survey by archaeologists. **(C)** Spatial relation analysis of geoglyphs with respect to trails and linear geoglyphs/trapezoids.

Figure S3

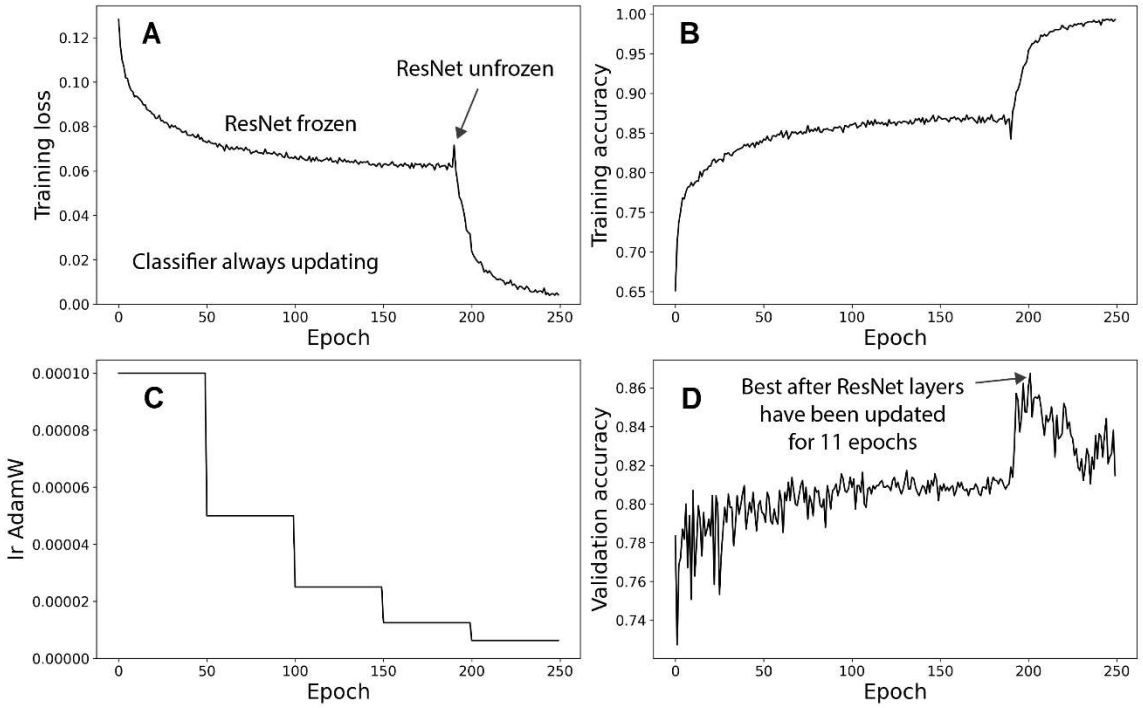
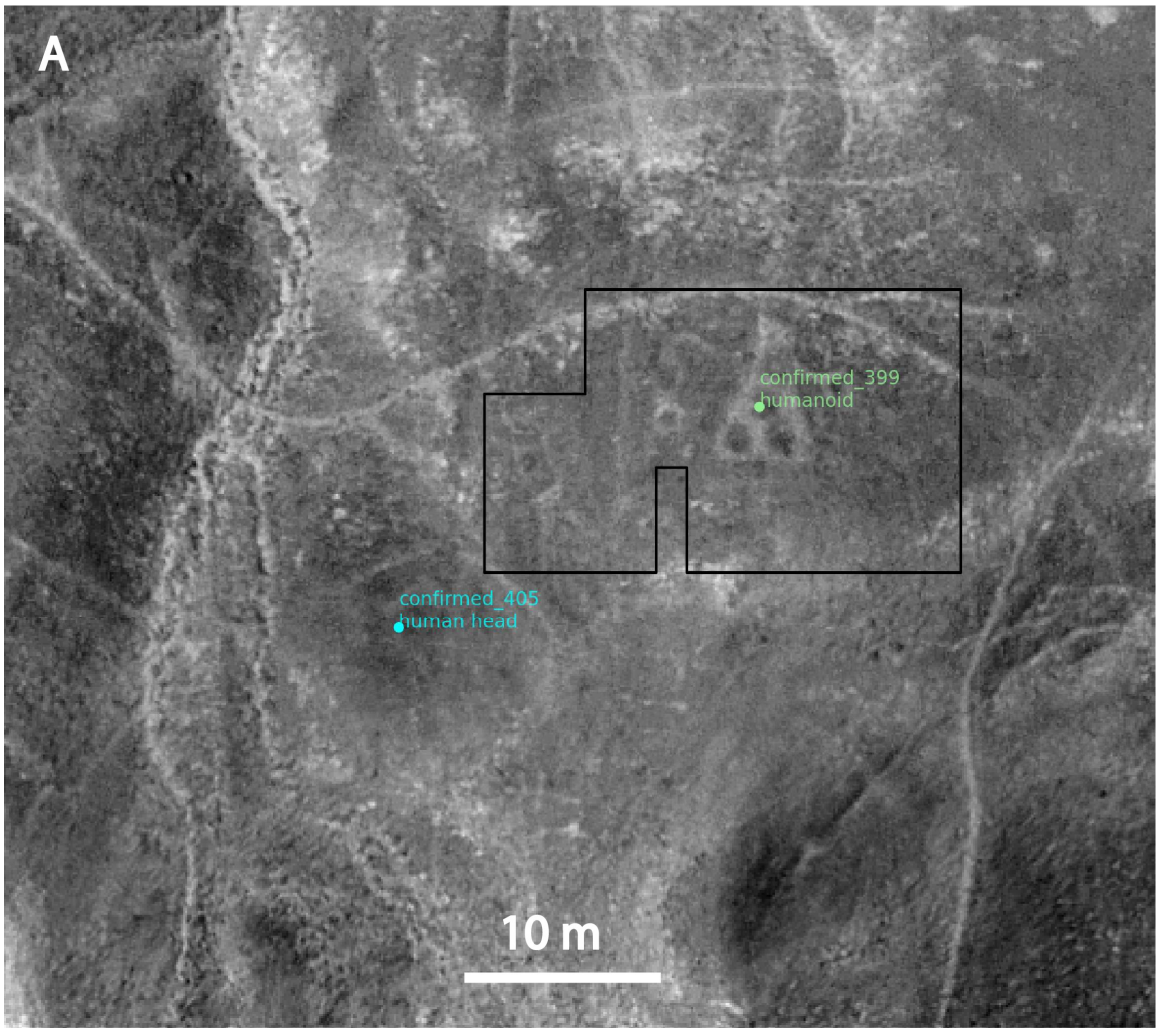
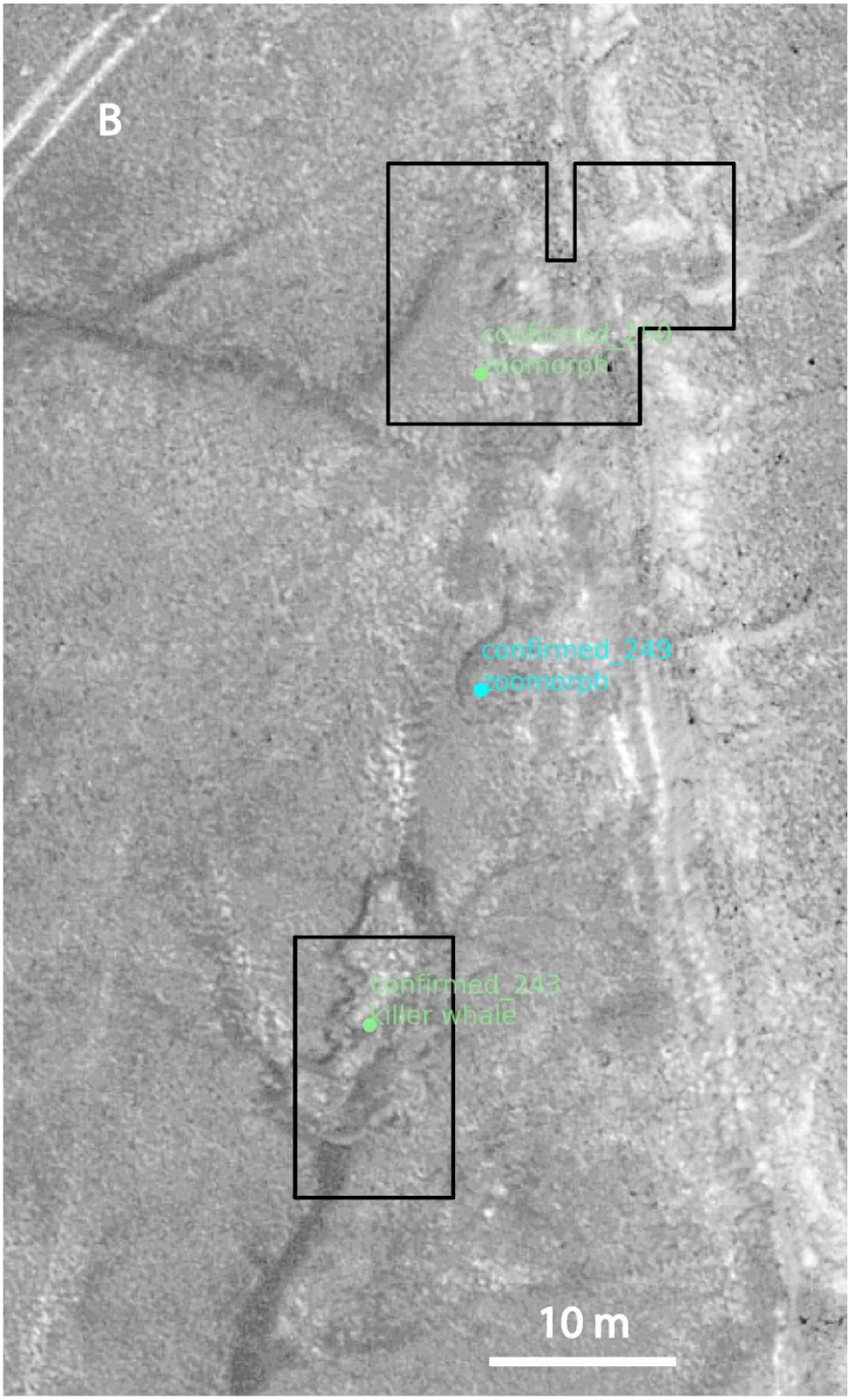
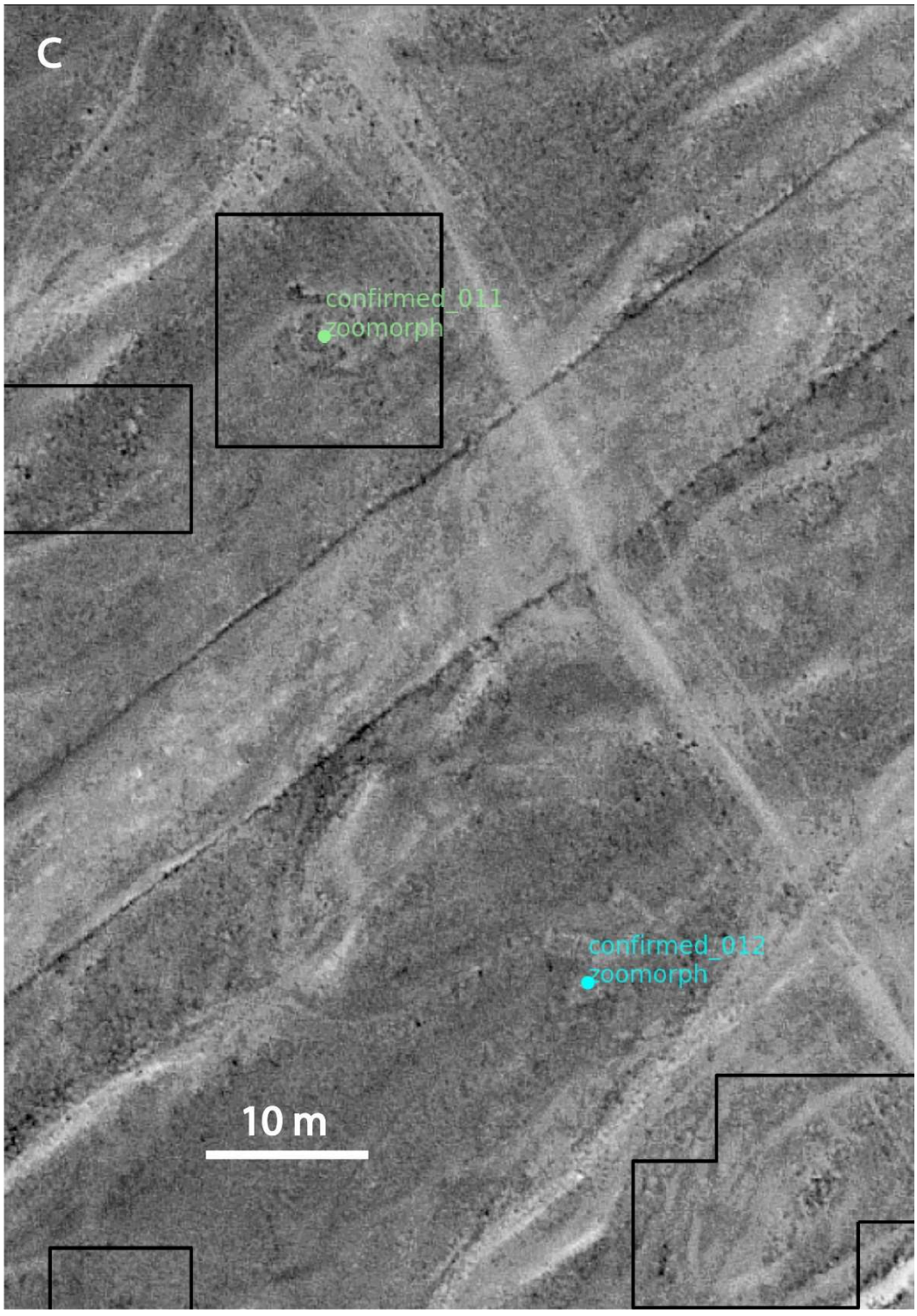


Fig. S3. Model training and validation. (A) Training Loss, **(B)** Training Accuracy, **(C)** Learning Rate Decay, and **(D)** Validation Accuracy as a function of Training Epoch. The classifier weights are updated continuously, while the pretrained feature extractor weights (on ImageNet) are frozen until epoch 190. The best validation accuracy is achieved at epoch 201.

Fig. S4.
Examples of confirmed geoglyph groups in relation to the AI-suggested boxes.
(A-D) Geoglyphs that intersect one of the boxes are classified as AI-assisted discoveries (green). Geoglyphs that are part of a group with at least one AI-assisted geoglyph discovery are classified as part of AI-groups since they are less than 50m apart (cyan).
(This figures spans 4 pages)




B
confirmed_249
zoomorph
confirmed_243
killer whale
10 m


C
confirmed_011
zoomorph
confirmed_012
zoomorph
10 m

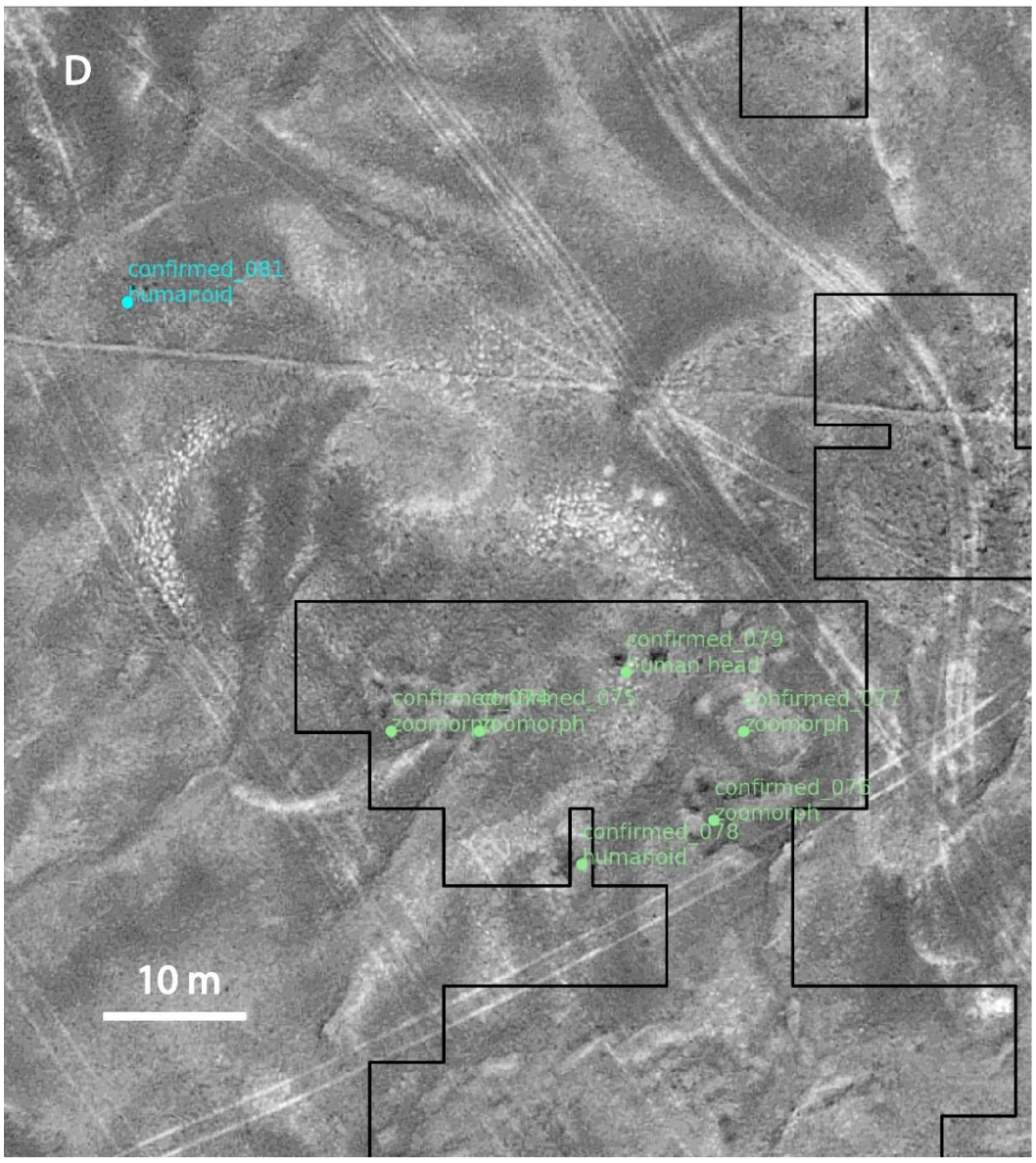

D
confirmed_081
humanoid
confirmed_079
human head
confirmed_074
confirmed_075
zoomorph
zoomorph
confirmed_077
zoomorph
confirmed_076
zoomorph
confirmed_078
humanoid
10 m

Figure S5

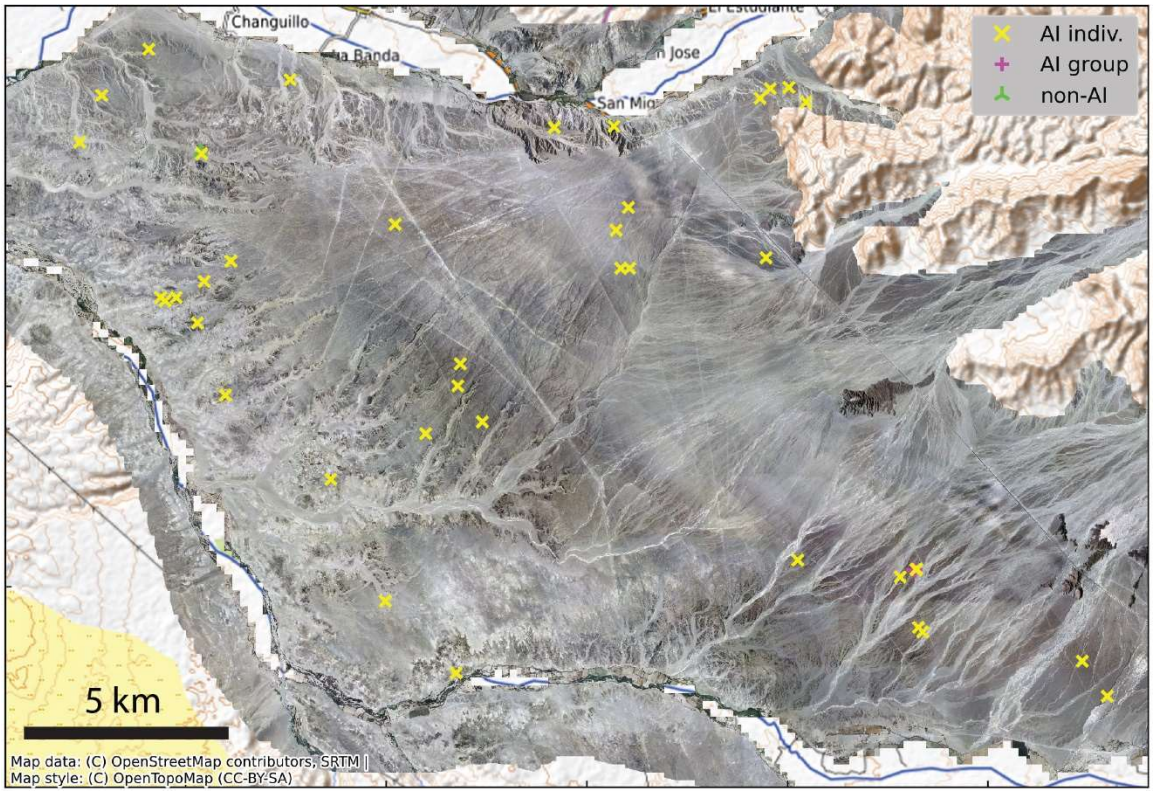


Fig. S5. Distribution of the 42 newly discovered geometric geoglyphs. (40 AI, 1 AI-group, and 1 non-AI) from the 2022/23 field survey.

Table S1

	Recall		Precision		F1	
	N=1	N=2	N=1	N=2	N=1	N=2
P=0.50	0.68	0.56	0.007	0.012	0.013	0.024
P=0.55	0.42	0.35	0.011	0.025	0.021	0.047

Table S1. Testing using held-out geoglyphs. Recall, precision, and F1 metrics as a function of tunable parameters N (minimum size of geoglyph candidate) and P (cutoff probability), to quantify model performance.

Table S2

	(a) screened candidates	(b) visited field survey 2022/23	visited %	confirmed figurative geoglyphs (method c1)	confirmed figurative geoglyphs (method c2)	(d) AI success rate [confirmed /visited]	(e) additional figurative geoglyphs expected	extrapolated
Rank I	227	155	68.3	109	77	0.5	37	114
Rank II	361	95	26.3	27	24	0.25	66	90
Rank III	721	91	12.6	30	21	0.23	145	166
total ranked	1309	341		166	122		248	370
unranked proposed by AI				78	56			
total AI				244	178			
non-AI				59	125			
total				303	303			

Table S2. Newly discovered Figurative Geoglyphs. AI-suggested geoglyph candidates of the three ranks and how many of them have been (a) screened by us and determined to be likely genuine by inspecting the imagery, (b) visited by us during the 2022/23 field survey, (c) confirmed as new figurative geoglyphs. (d) Calculated AI success rate. (e) Additional figurative geoglyphs expected from the three ranks in future field surveys and extrapolation. Confirmed new figurative geoglyphs are accounted for in two different ways: (c1) Counting as AI-discovered those geoglyphs that were individually predicted by AI and those associated with AI-predicted geoglyphs in groups. (c2) Counting as AI-discovered only those geoglyphs that were individually predicted by the AI. The AI success (d) rate is based on the latter. Unranked geoglyphs that were predicted by the AI but not included in the likely geoglyphs of the three ranks, as well as figurative Geoglyphs discovered without the help of AI (“non-AI") are shown at the bottom.

Table S3

	(a) screened candidates	(b) visited field survey 2022/23	visited %	confirmed geometric geoglyphs (method c1)	confirmed geometric geoglyphs (method c2)	(d) AI success rate [confirmed /visited]	(e) additional geometric geoglyphs expected	extrapolated
Rank I	227	155	68.3	25	25	0.16	11	36
Rank II	361	95	26.3	13	12	0.13	35	47
Rank III	721	91	12.6	3	3	0.03	19	22
total ranked	1309	341		41	40		65	105
unranked proposed by AI				0	0			
total AI				41	40			
non-AI				1	2			
total				42	42			

Table S3. Newly discovered Geometric Geoglyphs. AI-suggested geoglyph candidates of the three ranks and how many of them have been (a) screened by us and determined to be likely genuine by inspecting the imagery, (b) visited by us during the 2022/23 field survey, (c) confirmed as new geometric geoglyphs. (d) Calculated AI success rate. (e) Additional geometric geoglyphs expected from the three ranks in future field surveys and extrapolation. Confirmed new geometric geoglyphs are accounted for in two different ways: (c1) Counting as AI-discovered those geoglyphs that were individually predicted by AI and those associated with AI-predicted geoglyphs in groups. (c2) Counting as AI-discovered only those geoglyphs that were individually predicted by the AI. The AI success (d) rate is based on the latter. Unranked geoglyphs that were predicted by the AI but not included in the likely geoglyphs of the three ranks, as well as geometric Geoglyphs discovered without the help of AI (“non-AI") are shown at the bottom.