Revising the timing of arrival of humans in America through deep classification of cut marks on bones

Manuel Domínguez-Rodrigo[1,2,3*], Wonmin Byeon[4], Georgios Arampatzis[4], Luciano Varela[5], Sebastián Tambusso[5], Richard A. Fariña[5], Enrique Baquedano[1], José Yravedra[1,3], Miguel Angel Maté-González[3], Petros Koumoutsakos[4,6]

[1]Institute of Evolution in Africa (IDEA), Covarrubias 36, 28010 Madrid, Spain.
[2]Real Complutense College at Harvard, 26 Trowbridge Street, Cambridge, MA 02138, USA
[3]Human Evolutionary Biology Department, Harvard University, 11 Divinity Avenue, Cambridge, MA 02138, USA
[4]Computational Science and Engineering Laboratory, ETH Zurich, CH-8092, Switzerland
[5]Department of Paleontology, Faculty of Science, University of the Republic, Iguá 4225, Montevideo, Uruguay
[6]Radcliffe Institute, Harvard University, Cambridge, MA 02138, USA

*m.dominguez.rodrigo@gmail.com

The first three authors contributed equally

The earliest uncontroversial presence of humans in America dates to fourteen thousand years ago, at the end of the Last Glacial Maximum (LGM). This presence is evidenced by stone tools and related cut marks and other bone surface modifications (BSM) associated with the consumption of animals by humans. However, this most crucial part of taphonomic analysis of the archaeological human record, has been controversial due to highly subjective interpretations of BSM. Here, we employ advanced machine learning algorithms to increase the accuracy of mark identification on bones. We demonstrate that deep convolutional neural networks (DCNN) and Support Vector Machines (SVM) can recognize marks with accuracy that far exceeds that of human experts (91% over 63%). The application of this method of mark identification to BSM of the 30 Ka site of Arroyo del Vizcaíno (Uruguay) shows with high probability (over 90%) the presence of cut marks imparted by stone tools on bones. This result supports an earlier presence of humans, estimated to 30.000 years ago, in the American continent.

INTRODUCTION

The earliest presence of humans in America has long been controversial, with a consensual interpretation that no convincing evidence existed prior to 15.000 B.P. Recently, a new discovery consisting of the association of a few purported artefacts with some remains of a mastodon, places the oldest evidence of humans in this continent by 130.000 B.P. [1]. However, the evidence provided both for the artefactual nature of the rocks and for the anthropogenic modification of the mastodon remains is yet to be fully accepted. Also, a very small number of sites in both North America (for instance, Bluefish Cave) [2,3] and South America (for example, Monte Verde in Chile and some sites in Serra de Capivara in Brazil) [4,5] claim a slightly older presence of humans in America than 15 Ka. However, the best evidence of pre-LGM human presence in America could potentially be found at Arroyo del Vizcaíno (Uruguay) [6].

This site, located at the bottom of a river stream in southern Uruguay (34°37'2.92" S, 56° 2'32.54" W), has yielded a collection of over 1,500 remains of the South American Pleistocene megafauna (mostly belonging to the giant sloth *Lestodon armatus*), while it is estimated that many thousands are yet to be extracted [7]. The fossils are densely packed in a 60 cm thick level that fines upwards from a muddy sandy gravel to a muddy sand (Fig. 1). About 12 radiocarbon dates have been obtained from purified and non-purified collagen as well as from wood in different laboratories and at different times and they cluster at over 30 kybp [6-10]. The potential existence of cut marks on the Arroyo del Vizcaíno bones advocates an earlier human presence in America, well before the LGM. Indeed, the dates obtained and the possible human modifications are from the very same objects (i.e. the marked bones), thus suggesting a very strong association.

In turn, this association emphasizes the importance of correctly identifying the bone surface modifications that would support such an early human presence in the American continent. In fact, the use of cut marks in the fossil record has long been used for the identification of specific butchering behaviors by early Pleistocene hominins and their carcass acquisition strategies in the earliest stages of evolution of *Homo* [9-13]. Claims have been made about the Pliocene antiquity of carcass processing using purported cut-marked bones in Dikika (Ethiopia) [10] or in the Plio-Pleistocene boundary in Quranwala (India) [14]. The claims made based on these discoveries remain controversial [14], as the purported cut-marked fossils from both sites also bear clear evidence of trampling or sedimentary abrasion [15]. Similar to the findings in Dikika and Quranwala, the Arroyo del Vizcaino site (Uruguay) also contains abundant conspicuous evidence of trampling and/or sedimentary abrasion [7].

Taphonomists broadly disagree on how cut marks should be properly identified. A series of microscopic criteria were experimentally defined [16–18], but there is substantial disagreement on how these criteria should be interpreted by individual researchers [17,19]. The lack of objective methods underscores the fact that the present identification of BSM (namely, cut marks) depends strongly on the subjective assessment and knowledge of each researcher. Cut mark identification is presently conducted outside replicable hypothesis-testing frameworks.

In order to increase the level of objectivity in cut mark identification, we introduce automatic image classification by machine learning algorithms. We consider in particular Convolutional Neural Networks (CNN) [20–25] and Support Vector Machine (SVM) [26–29] that have been shown to consistently exceed human capabilities in image and pattern recognition. Their use for image processing and classification has been recently greatly facilitated by access to open source software packages such as Neuroph of TensorFlow [30-32].

The experimental sample size that was used in the present work (79 marks) intentionally kept short for two reasons: to maximize human expert scores (we had experimentally tested that the larger the sample the higher the identification failure rates by humans) and to minimize the computer´s accuracy (the higher the sample size the better the computer is trained and the higher the identification rates produced) [26-32]. Despite this initial advantage for human experts, the differences in successful identifications of marks reported here between humans and computer algorithms are large in favor of the latter.

The CNN and SVM methods enable taphonomist to perform bone surface mark identification in a more objective way than has ever been possible. We present results that demonstrate the superiority of machine learning algorithms in identifying BSM over “subjective” assessment by several human experts. At the heart of this comparison lies the question of whether cut marks can be reliably identified and not if a machine can be a better classifier than a human. The positive answer to this question, using modern bones, opens the door to the consideration that cut marks, like other taphonomic entities, are subjected to morphological evolution, through a palimpsestic multiple-agent processes and an interplay between stasis and change [33–35]. Taphonomists would then have to examine whether fossil cut marks could be as reliably identifiable as modern experimental cut marks. This would enable the objective resolution of many cut mark-related controversies, including whether hominin butchering behaviors are identifiable in Pliocene fossils and whether the Americas were reliably occupied by humans more than 30.000 years ago.

METHODS

Methodological description of the sample

Experimental Marks

A selection of 79 experimental marks was used as the training set (SI). These were composed of 42 trampling marks and 37 cut marks. Trampling marks were created by using four types of sediments: fine-grained (0.06–0.2 mm), medium-grained (0.2–0.6 mm and coarse-grained (0.6–2.0 mm) sand, as well as a combination of the previous sand types over a clay substratum, and granular gravel (**>**2.0 mm). These marks were selected from the trampling experiment reported by [18] and they include all the variety of abrasive sediment particles (other than large pebble gravel grains ranging between 4-6 mm) potentially creating trampling marks in natural settings.

The set of cut marks was made using quartzite flakes as reported in Domínguez-Rodrigo et al.´s experimental sample [18]. The 37 cut marks were made with simple flakes (n=10) and retouched flakes (n=27). Although these marks are somewhat dissimilar they are much more similar compared to other non-anthropogenic marks. The use of retouched flakes was purposely chosen because the resulting marks are statistically non-differentiable from cut marks created by natural rock flakes given that both are caused by irregular edges as opposed to the more straight edge of simple flakes [36]. Since one of the most polemic interpretations of the earliest purported cut marks (e.g., Dikika) is that they may have been made with natural rocks, by using marks made with retouched flakes we model that assumption as well as that of cut marks created with modified tools for later periods. For this study, both sets of cut marks were lumped together for targeting discrimination between trampling and cut marks in general.

The selected marks were photographed in a standardized way under the binocular microscope under 40x. Images were then modified in Photoshop by transferring them to grayscale. Further transformation of the original files was carried out as explained in the computing methods below (SI).

Archaeological Marks

A total of six marks from the archaeological assemblage of Arroyo del Vizcaíno were selected on two *Lestodon* rib fragments (CAV 451 and 452) (Fig. 2). The marks, estimated by experts to be cut marks, were observed and photographed using an Olympus SZ61 stereomicroscope under a magnification of 40x. Helicon Focus software was used to make a complete in-focus image.

Image pattern recognition methods

Deep learning-based method

Convolutional Neural Networks (CNNs) are among the most potent deep learning-based methods for image classification [24,37]. The networks receives an image as input and transforms it via several hidden layers. Each hidden layer includes convolutional, Rectified Linear Unit (ReLU), Max-Pooling (MP), and Fully-Connected (FC) layers. The convolutional layer computes the weights of a grid of neurons which are connected to the local regions (kernels) of each pixel. The ReLU layer contains the nonlinear activation function: $f(x)=max(0,x)$. The MP layer takes the maximum value from the rectangular region which down-samples and compresses the information along the spatial dimensions. Finally, the FC layer has fully-connected neurons from the activations of the previous layer. These connections are weight matrix multiplications with biases followed by the ReLU activation function. The final layer is the softmax function which computes the normalized posteriors. The networks is trained to minimize the cross-entropy loss between the output and the true distribution. The architecture of Convolutional Neural Networks is summarized in Figure 3.

Support Vector Machine (SVM) method

In this approach we classify the images by training a binary SVM classifier [38]. First a vocabulary of visual words is created from images in the training set using the Bag of Words (BoW) technique [39]. In this method features are extracted from images and the feature space is clustered. The vocabulary is composed by the different clusters. The images in the data set are encoded using the BoW vocabulary and an SVM is trained on the encoded set.

Testing by human experts

Three experienced taphonomist with 7-20 years of practical experience on bone surface modifications using modern (i.e., controlled) and fossil assemblages were selected to independently identify mark types in the image set described above. Their results were compared first among themselves and then with those provided by the computer vision approach. The three experts were trained by one of the senior authors (MDR) of this paper.

Experiments

Dataset.

The experimental marks dataset consists of 79 gray scale images and is divided into a training and a test set. The test set consists of 10 images from each category (cut and trampling) and the rest of images were used for training. The accuracy of the classifier is estimated by averaging the accuracy over 60 independent trainings and the mean, as well as, the standard deviation of the accuracy is reported. Finally, the classifier is trained using the whole experimental dataset and the six archaeological marks are being classified.

Convolutional Neural Networks (CNNs)

Pre-processing**.** The size of original images in the dataset were heterogeneous in origin; the width ranged between 573 and 1350 pixels and the height ranged between 71 to 375 pixels. To speed up the training and provide a fixed size input to the network, the image resolution was resized to 180x520 pixels. The image values were normalized between 0 and 1 by the min-max scaling: $X_i' = \frac{X_i - min(X)}{max(X) - min(X)}$, where X_i is the original value and X_i' is the normalized value at the position i . No other pre-processing was performed, although the illumination and the quality of the images in the dataset are very diverse (see Fig. 4).

Architecture. The input layer receives an image with the size $520 \times 180 \times 1$. To find the optimal architecture, two to four convolutional layers with MP layers and one to two FC layers with various sizes of neurons, and filter size between three to seven were tested. Exponential Linear Units (ELU) activation [40] was also tested in addition to ReLU, but it yielded no performance difference.

Finally, two convolutional layers and one FC layer were used as a final model. The first and second convolutional layers have 64 and 128 convolutional kernels. The process was followed by ReLU activation and MP. The size of the kernels for convolutional layers was $5 \times 5 \times 1$ with a stride 1 and for max-pooling layers it was $2 \times 2 \times 1$ with a stride 2 . In the last layer, one fully-connected layer with 512 neurons were connected to the output layer consisting of a binary classification.

Training. The network was trained by Adaptive Moment Estimation (Adam) [35]. The learning rate was initialized to 0.001 and it used the exponential decay with a rate of 0.99. L2-weight decay and dropout [41] were used as regularizers in the FC layer. The constant for the L2-

weight decay was $5e\text{-}4$, including 20 of dropout. The weights were randomly initialized with zero mean and a 0.1 standard deviation, and the biases were initialized to the constant value 0.1 . Tensorflow GPU framework [39] was used for the CNN experiments.

Support Vector Machine (SVM)

Pre-processing. The images were normalized such that they all have zero mean and standard deviation one.

Architecture. The implementation of the SVM approach was done in Matlab using the Computer Vision System Toolbox. The key-points in the images were selected from points either on a regular grid or detected with a SURF detector. In both cases, the vocabulary was created from all the images in the training set using features extracted with the SURF extractor [42,43]. A grid step size of 12×12 and a vocabulary size of 600 was yielding results comparable to those of the SURF detector with a vocabulary size of 400 . For the SVM, a gaussian kernel was used in conjunction with hyperparameters, scaling and box constraint. The later is related to the penalization of misclassifications in the case of non-separable data sets.

Training. The hyperparameters were optimized using Bayesian optimization [44] and the Covariance Matrix Adaption Evolution Strategy (CMA-ES) algorithm [45] produced similar results.

For both, CNN and SVM classifiers, the accuracy of BSM identification was averaged over 60 models trained by randomly partitioning the experimental dataset. This evaluation technique gives reliable validation of the model; the performance was not biased by a particular partition of training and testing sets. In our experiments, SVM was less stable than CNNs depending on the partition of the samples.

Bone mark identification by machine learning algorithms and human experts

Experimental marks

The best CNNs architecture identified marks correctly with a 91 % mean accuracy (sd=5.3). Note that cut marks and trampling marks have balanced classification accuracy unlike SVM and human experts: 90 mean accuracy (sd= 10.6) for cut marks and 90 mean accuracy (sd= 8.0) for trampling marks.

The best accuracy obtained by SVMs is 81.5 (sd=7.5) using the grid method to select points and 82.5 (sd=8) using the SURF selector. For the SURF detector, the mean accuracy of trampling mark identification is 80 (sd=80) and the mean accuracy of cut mark identification is 84.3 (sd=13).

In contrast, human experts correctly identified a substantially lower frequency of marks. The three taphonomists produced similar identification rates. A chi-square test showed that their identifications are statistically indistinguishable (X2= 1.7577; p= 0.780). They identified correctly a higher percentage of cut marks (mean=66%; range= 64%-70%) than of trampling marks (mean=60%; range= 52%-65%). Their overall correct identification rate averaged 63% of marks.

Archaeological marks

For the classification of the archaeological cut marks the classifiers were trained on the whole experimental dataset. Both CNN and SVM classified them as cut marks. This classification is in agreement with the human experts' classification. However, there is a difference in the certainty of the results: CNNs are between 90.5% and 99.9% and SVMs between 57% and 72%.

The unreliability of SVM-based approach stems from the fact that the features from the images are extracted before the training phase of SVM. These hand-designed extractors extract textures by considering neighboring information. Under the diverse conditions of the experimental dataset these extractors become unstable. On the other hand, CNN classifiers learn the features from the images in during the training stage. This gives the ability to the model to generalize well and especially with a large amount of data. Therefore, the CNN model trained by experimental marks is very reliable for the new (unseen) archaeological marks.

DISCUSSION

The accuracy in experimental mark identification by CNN and its high probability in the correct identification of the selected marks from Arroyo del Vizcaíno supports that this site´s modified bones could potentially constitute the oldest direct evidence of humans especially in South America but possibly in the Americas as a whole.

The present study shows that expert taphonomists may be only moderately successful in identifying BSM in high (40x) magnification (accuracy=63%). This is an insufficient success rate for elaborating reliable interpretations pending on correct identification of marks, such as those on the Arroyo del Vizcaíno, the Dikika fossils or the Quranwala purported cut-marked bones. *Ad hoc* experimentation in the latter case may be as spurious as for the Dikika case [13]. Similarities between particular types of experimental and fossil marks, both in the case of the Quranwala and Dikika, are overshadowed by the plethora of other associated marks on those fossils that remain ignored and by other available experimental marks that are morphologically similar and objectively undifferentiated from those reported on those fossil bones [14].

The present study shows that machine learning algorithms can help overcome the subjectivity and human bias in the interpretation of bone surface modifications. This "objective" approach depends on how well the computer algorithm is trained. This quality is also dependent not only on the within-sample diversity of each type of mark but also on the variety of marks types that are part of the training sample. The accuracy achieved in the present work (>91%) to differentiate cut marks from trampling marks in unprecedented. Future studies should build on the referential works used here and the results obtained. All trampling marks used in this study are made with fine-grained, medium-grained and coarse-grained sand as well as small gravel [18]. Future analyses should include trampling marks created with pebble gravel, such as those reported by [14]. Most cut marks used were made with flint and quartzite tools. Given the diversity in cut mark morphologies created by same tools made on different raw materials [46,47], marks created with other raw material types and using different tool types should also provide a wider panorama of the morphological diversity that each of these agents introduce.

A larger image dataset will contribute to refine the results presented here. For instance, the most successful work for image classification using CNNs [24] used 1.3 million images for training, and the model in [25] for traffic sign recognition is trained on 25,350 images. The samples used in this work was very difficult to identify. The data samples have high variation in illumination, contrast and resolution. Identifying them is even hard with human eyes (see Fig. 4). Minimal pre-processing (only image resizing and normalization) and no data augmentation were performed in model training. A wide variety of image processing and transformation methods are frequently used in the literature to increase the robustness of the model, especially for deep learning approaches [20-25]. The common image processing algorithms as pre-processing are histogram equalization or contrast normalization, which enhance the quality of images. The common data augmentation includes random transformation, rotation, multi-scaling, and flipping, which diversify the dataset. Such techniques usually help to increase the classification performance. Despite not having applied any of these improvement sample methods, our results outperform human experts by a wide margin (almost 50%). This fact, regardless of the sample size used, underscores the drastic improvement in mark identification by machine learning methods compared to human experts. The present work demonstrates that there is great potential in employing machine learning to advance the taphonomic discipline and we envision that it will be broadly adopted by the community.

While it may still be too early to assert that humans (and, if so, which species) were present in America 130.000 years ago [1] the present BSM deep classification method to the Arroyo del Vizcaíno site provides the most unambiguous direct evidence of early human presence in America, dating to 30'000 years ago. This finding will encourage the search for this kind of evidence (as well as others) in similarly early sites like those in southern Chile and northern Brazil. This discovery also supports a human presence previously documented at Bluefish Caves (Canada) 24,000 years ago, through the presence of a cutmarked equid mandible [48]. The results presented here provide compelling evidence that humans were directly engaged in megafaunal butchery~~,~~ supporting previous interpretations of the human impact in the megafaunal extinction at the end of the Pleistocene in America [49]. This does not necessarily support that human presence in America extends back to the Middle Pleistocene [1], but certainly provides compelling support to a human presence prior to LGM.

Method online

A step-by-step description and the code used will be provided upon acceptance.

Acknowledgements

This collaborative work was carried out with support from a Research Salvador Madariaga grant to MDR (Ministry of Education, Culture and Sport, Spain. Ref PRX16/00010). WB, GA and PK acknowledge support by the ERC Advanced Investigator award No 2-73985-14. MDR thanks D. Lieberman and the Human Evolutionary Biology Department at Harvard and the Royal Complutense College at Harvard, where this research was conducted. RAF, LV and PST thanks to CSIC (UdelaR) and ANII for financial support. We also thank Lucía Cobo and Julia Aramendi for their help during the design of this pilot study. Roberto Bracco-Boksar made important contributions. Martín Batallés provided the photographs in Figure 1. We also thank three taphonomic experts, who will remain anonymous in the present paper.

References

1. Holen SR, Deméré TA, Fisher DC, Fullagar R, Paces JB, Jefferson GT, Beeton JM, Cerutti RA, Rountrey AN, Vescera L, Holen KA. 2017. A 130,000-year-old archaeological site in southern California, USA. Nature 544: 479-483.
2. Bourgeon L, Burke A, Higham T. 2017. Earliest human presence in North America dated to the Last Glacial Maximum: New radiocarbon dates from Bluefish Caves, Canada. PLoS One. 12 (1): e0169486. doi:10.1371/journal.pone.0169486
3. Pitblado BL. 2011. A tale of two migrations: Reconciling recent biological and archaeological evidence for the Pleistocene peopling of the Americas. Journal of Archaeological Research 19: 327–375.
4. Dillehay TD, Ocampo C, Saavedra J, Sawakuchi AO, Vega RM, Pino M; Collins MB, Cummings LS, Arregui I, Villagrán XS, Hartmann GA, Mella M, González A, Dix G. 2015. New Archaeological Evidence for an Early Human Presence at Monte Verde, Chile. PLoS ONE 10 (11): e0141923
5. Boëda E, Rocca R, Da Costa A, Fontugne M, Hatté C, Clemente-Conte I, Santos JC, Lucas L, Felice G, Lourdeau A, Villagrán X, Gluchy M, Ramos MP, Viana S, Lahaye C, Guidon N, Griggo C, Pino M, Pessis A-M, Borges C, Gato B. 2016. New data on a Pleistocene archaeological sequence in South America: Toca do Sítio do Meio, Piauí, Brazil, PaleoAmerica, 2:4, 286-302
6. Fariña RA, Tambusso PS, Varela L, Di Giacomo M, Musso M, Gascue A, Bracco R. 2014b. Among others, cut-marks are archaeological evidence: Reply to "Archaeological evidences are still missing: Comment on Fariña et al. Arroyo del Vizcaíno Site, Uruguay" by Suárez et al. *Proceedings of the Royal Society B* 281 (1795): 20141637.
7. Fariña RA, Tambusso PS, Varela L, Czerwonogora A, Di Giacomo M, Musso M, Bracco-Boksar R, Gascue A. 2014a. Arroyo del Vizcaíno, Uruguay: A fossil-rich 30-ka-old megafaunal locality with cut-marked bones. *Proceedings of the Royal Society B* 281 (1774): 20132211.
8. Fariña RA. 2015. Bone surface modifications, reasonable certainty and human antiquity in the Americas: the case of the arroyo del Vizcaíno site. *American Antiquity* 80 (1): 193-200.
9. Domínguez-Rodrigo M, Pickering TR. The meat of the matter: an evolutionary perspective on human carnivory. Azania: Archaeological Research in Africa. 0: 1–29.

10. McPherron SP, Alemseged Z, Marean CW, Wynn JG, Reed D, Geraads D, et al. Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. Nature. 2010;466: 857–860.
11. Harmand S, Lewis JE, Feibel CS, Lepre CJ, Prat S, Lenoble A, et al. 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. Nature. 2015;521: 310–315.
12. Domínguez-Rodrigo M, Barba R, Egeland CP. Deconstructing Olduvai: a taphonomic study of the Bed I sites. Springer Science & Business Media; 2007.
13. Dambricourt Malassé A, Moigne A-M, Singh M, Calligaro T, Karir B, Gaillard C, et al. Intentional cut marks on bovid from the Quranwala zone, 2.6 Ma, Siwalik Frontal Range, northwestern India. C R Palevol. 2016;15: 317–339.
14. Domínguez-Rodrigo M, Pickering TR, Bunn HT. Configurational approach to identifying the earliest hominin butchers. Proc Natl Acad Sci U S A. 2010;107: 20929–20934.
15. Domínguez-Rodrigo M, Alcalá L. 3.3-Million-year-old stone tools and butchery traces? More evidence needed. PaleoAnthropology. 2016;2016: 46–53.
16. Fisher JW Jr. Bone surface modifications in zooarchaeology. J Archaeol Method Theory. Kluwer Academic Publishers; 1995;2: 7–68.
17. Domínguez-Rodrigo M, de Juana S, Galán AB, Rodríguez M. A new protocol to differentiate trampling marks from butchery cut marks. J Archaeol Sci. Elsevier; 2009;36: 2643–2654.
18. de Juana S, Galán AB, Domínguez-Rodrigo M. Taphonomic identification of cut marks made with lithic handaxes: an experimental study. J Archaeol Sci. Elsevier; 2010;37: 1841–1850.
19. Monnier, GF, Bischoff, E. Size matters. An evaluation of descriptive criteria for identifying cut marks made by unmodified rocks during butchery. J Archaeol Sci. Elsevier; 2014; 50: 305-317.
20. Hijazi S, Kumar R, Rowen C. Using Convolutional Neural Networks for Image Recognition [Internet]. 2015. Available: http://www.multimediadocs.com/assets/cadence_emea/documents/using_convolutional_neural_networks_for_image_recognition.pdf
21. Yalcin H, Razavi S. Plant classification using convolutional neural networks. 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics). IEEE; pp. 1–5.
22. Tivive FHC, Bouzerdoum A. A face detection system using shunting inhibitory convolutional neural networks. 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat No04CH37541). IEEE; pp. 2571–2575.
23. Fasel B. Facial expression analysis using shape and motion information extracted by convolutional neural networks. Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing. IEEE; pp. 607–616.
24. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems 25. Curran Associates, Inc.; 2012. pp. 1097–1105.
25. Ciresan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; pp. 3642–3649.
26. Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. Data Min Knowl Discov. Kluwer Academic Publishers; 1998;2: 121–167.
27. Phillips PJ. Support Vector Machines Applied to Face Recognition. In: Kearns MJ, Solla SA, Cohn DA, editors. Advances in Neural Information Processing Systems 11. MIT Press; 1999. pp. 803–809.

28. Chapelle O. Support Vector Machines et Classification d' Images. Available: https://pdfs.semanticscholar.org/fb6b/a3944cf1e534f665bc86075e0af2d2337eb9.pdf
29. Lee S-W, Verri A. Pattern Recognition with Support Vector Machines: First International Workshop, SVM 2002, Niagara Falls, Canada, August 10, 2002. Proceedings. Springer; 2003.
30. Egmont-Petersen M, de Ridder D, Handels H. Image processing with neural networks—a review. Pattern Recognit. 2002;35: 2279–2301.
31. Nielsen M. Using neural nets to recognize handwritten digits. Determination Press; 2015.
32. Ueli Meier JS. Multi-column deep neural networks for image classification. 2012; Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.367.484
33. Fernández López SR. Temas de tafonomía. eprints.ucm.es; 2000; Available: http://eprints.ucm.es/22003/1/087_00_Temas_Tafonomia.pdf
34. Fernández López SR. Tafonomía y fosilización. eprints.ucm.es; 1999; Available: http://eprints.ucm.es/21802/1/078_99_Tafonomia_y_Fosilizacion.pdf
35. Fernández López SR. Taphonomic alteration and evolutionary taphonomy. Journal of taphonomy. Palaeontological Network Foundation.; 2006;4: 111–142.
36. Domínguez-Rodrigo M, Pickering TR, Bunn HT. Experimental study of cut marks made with rocks unmodified by human flaking and its bearing on claims of 3.4-million-year-old butchery evidence from Dikika, Ethiopia. J Archaeol Sci. Elsevier; 2012;39: 205–214.
37. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86: 2278–2324.
38. Steinwart I, Christmann A. Support Vector Machines. Springer Science & Business Media; 2008.
39. Csurka G, Dance C, Fan L, Willamowski J, Bray C. Visual categorization with bags of keypoints. Workshop on statistical learning in computer vision, ECCV. Prague; 2004. pp. 1–2.
40. Clevert D-A, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) [Internet]. arXiv [cs.LG]. 2015. Available: http://arxiv.org/abs/1511.07289
41. Srivastava N, Hinton GR, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Over tting. Available: http://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf
42. Rampasek L, Goldenberg A. TensorFlow: Biology's Gateway to Deep Learning? Cell Syst. 2016;2: 12–14.
43. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-Up Robust Features (SURF). Comput Vis Image Underst. 2008;110: 346–359.
44. Pelikan M, Goldberg DE, Cantú-Paz E. BOA: The Bayesian optimization algorithm. of the 1st Annual Conference on …. dl.acm.org; 1999; Available: http://dl.acm.org/citation.cfm?id=2933973
45. Hansen N, Müller SD, Koumoutsakos P. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evol Comput. MIT Press; 2003; Available: http://www.mitpressjournals.org/doi/abs/10.1162/106365603321828970
46. Yravedra J, Maté-González MÁ, Palomeque-González JF, Aramendi J, Estaca-Gómez V, San Juan Blazquez M, et al. A new approach to raw material use in the exploitation of animal carcasses at BK (Upper Bed II, Olduvai Gorge, Tanzania): a micro-photogrammetric and geometric morphometric analysis of fossil cut marks. Boreas. Wiley Online Library; 2017; Available: http://onlinelibrary.wiley.com/doi/10.1111/bor.12224/full

47. Maté-González MÁ, Palomeque-González JF, Yravedra J, González-Aguilera D, Domínguez-Rodrigo M. Micro-photogrammetric and morphometric differentiation of cut marks on bones using metal knives, quartzite, and flint flakes. Archaeol Anthropol Sci. Springer Berlin Heidelberg; 2016; 1–12.
48. Bourgeon, L., Burke, A., Higham, T. Earliest human presence in North America dated tio the Last Glacial Maximum: new readiocarbon dates from Bluefish Caves, Canada. PLOS ONE, 2017, 12(1): e0169486 DOI:10.1371/journal.pone.0169486.
49. Fariña RA, Vizcaíno SF, De Iuliis G. 2013. Megafauna. Giant Beasts of Pleistocene South America. Bloomington, Indiana University Press, 416 pp.

Figure captions

Figure 1. Arroyo del Vizcaíno site. (A) panoramic view of the bonebed in its westernmost area, the grid used for reference the elements measures 1 m. (B) close up of the fossils in situ. (C) close up of partial excavated area. (D) stratigraphic sequence of the site showing the provenance of the fossils and the obtained radiocarbon dates.

Figure 2. *Lestodon* ribs (A) CAV 451 and (B) CAV 452 with cut marks. The insets show the marks under a magnification of 40x. The scale bars measure 5 cm.

Figure 3. Convolutional Neural Network Architecture: The network include two convolutional with Rectified Linear Unit (ReLU) and Max-Pooling (MP) layers. The output of the last MP layer is connected to one fully-connected layer, then the network outputs the final classification.

Figure 4. Examples of the challenging images depicting trampling and cut marks, which show different contrast and lighting. Despite this, CNN identified >90% of marks.


A
B
C
D
Sand
Gravel
Organic sediment
0.4 kyBP
30 kyBP
1
2
3
4
(m)
c s vf m vc g p c b

Fig 1

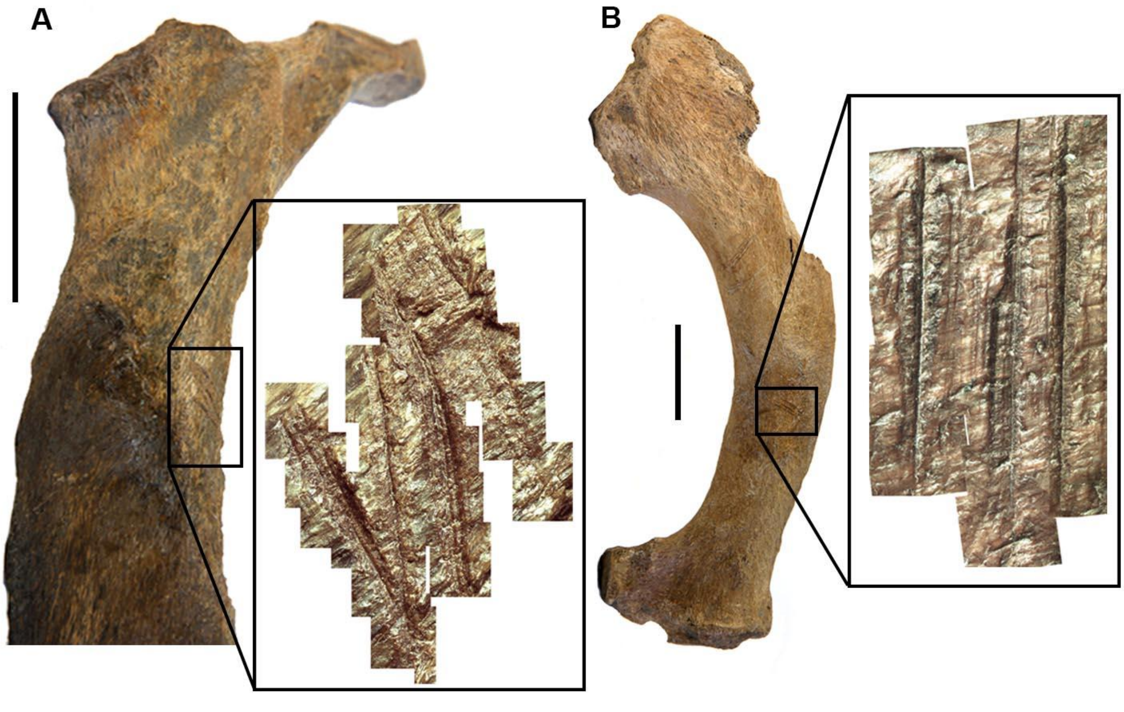
A
B

Fig 2

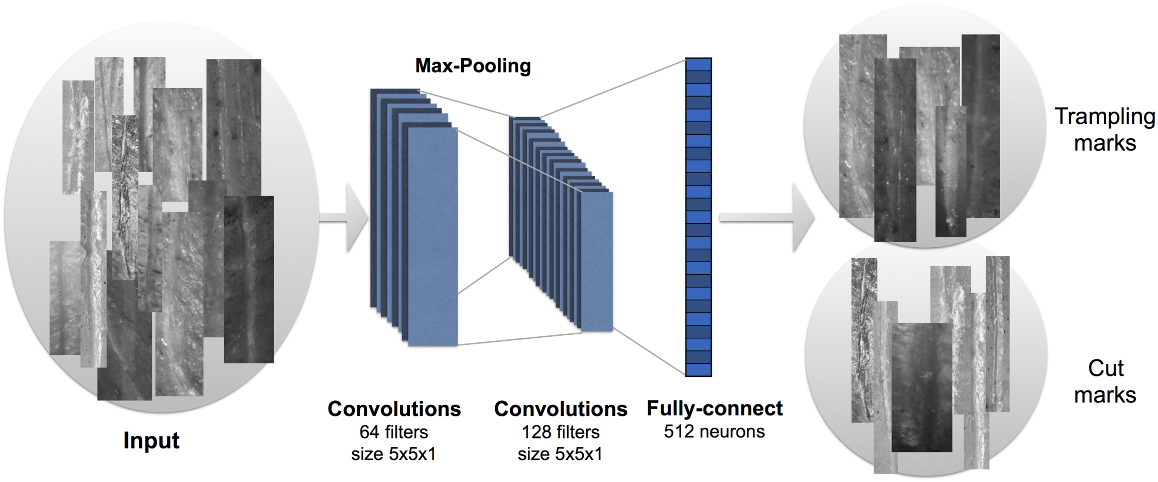
Max-Pooling
Trampling
marks
Cut
marks
Input
Convolutions
64 filters
size 5x5x1
Convolutions
128 filters
size 5x5x1
Fully-connect
512 neurons

Fig 3

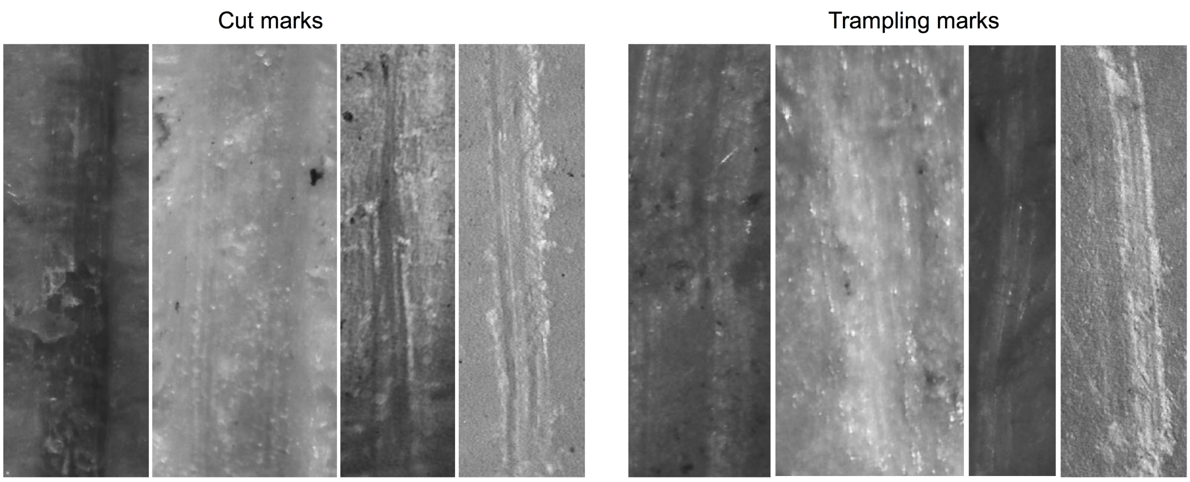
Cut marks
Trampling marks

Fig 4