RESEARCH ARTICLE

Towards a Method for Discerning Sources of Supply within the Human Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

Shawn Graham[*], Alex Lane[*], Damien Huffer[*] and Andreas Angourakis[†]

While traders of human remains on Instagram will give some indication, their best estimate, or repeat hearsay, regarding the geographic origin or provenance of the remains, how can we assess the veracity of these claims when we cannot physically examine the remains? A novel image analysis using convolutional neural networks in a one-shot learning architecture with a triplet loss function is used to develop a range of 'distances' to known 'reference' images for a group of skulls with known provenances and a group of images of skulls from social media posts. Comparing the two groups enables us to predict a broad geographic 'ancestry' for any given skull depicted, using a mixture discriminant analysis, as well as a machine-learning model, on the image dissimilarity scores. It thus seems possible to assign, in broad strokes, that a particular skull has a particular geographic ancestry.

1.1 Introduction

The forensic determination of 'ancestry' and estimation of sex of unprovenienced human remains (i.e. remains for which the archaeological origin is unknown) relies on the careful measurement of 'landmarks' on the elements at hand (ideally crania and/or pelves) that have been found to be diagnostic when suitably calibrated, standardized for population, and analyzed. There are software packages available (e.g. FORDISC 3.1, CRANID) that can be used by researchers, who merely have to measure and input the values, and the program (with its reference data) will perform the calculations and provide the result within a certain probability range. In recent years, researchers have begun to apply machine learning techniques to this data, with very good results (e.g. Navega et al. 2015a, 2015b; Ousley 2016; Maier et al. 2015) suggesting the potential to improve the accuracy of identifying unprovenienced remains at the population and individual levels (e.g. Nagpal et al. 2017).

But what if one does not actually have the bones physically present to measure? Can machine *vision* extract anything useful about potential ancestry? This is a critical question to investigate because traders buy and sell human remains online; there is a very active market for human bone, and to date it is nearly impossible to say anything about which people(s) are being bought and sold (see section 1.2). Any given skeletal element might

appear in a handful of images; once purchased, it disappears again into someone's collection. How many people's remains are being bought and sold? From what areas of the world (or populations) have their remains been sourced? We believe we can now begin to answer such previously imponderable questions.

In this paper we explore the potential for machine vision on simple photographs of human skulls as an initial experiment that uses a particular kind of machine vision (neural network) architecture to develop a suite of 'distances' from known reference images, and then perform a mixture discriminant analysis (mda) comparing a dataset with known, grounded provenience against a dataset sourced from social media posts. We outline a method that may be able to in the broadest strokes say something of the 'ancestry' of a skull, based on a computer-vision approach to measuring visual dissimilarity using a convolutional neural network, a triplet loss function, and comparison to a group of reference images. The visual dissimilarities that the neural network picks up on seem to be the same diagnostic areas that are used to osteometrically estimate ancestry, such as orbital shape and dimensions, nasal width and breadth, average degree of alveolar and maxillary prognathism, cranial vault shape, etc. The mda results seem to indicate that at the present moment, the method is sufficient to confirm or deny the story about a skull told by the vendor. With refinement and better-grounded provenience data we believe that this machine vision approach holds enormous potential for developing useful insights from photographic evidence. The neural network, our resulting measurements, and our analytical R code are available in our project repository at (Graham et al. 2020).

* Carleton University, Ottawa, CA

† University of Cambridge, Cambridge, UK

Corresponding author: Shawn Graham (shawn.graham@carleton.ca)

We use the term 'ancestry' or 'origin' in their forensic anthropological senses; we are not thereby implying that these categories are how these individuals experienced 'race'. Dunn et al. 2020 provide an overview of the paradox of 'estimating a culturally constructed, peer-perceived category (social race) from biological tissues. These estimations are only possible because of the nonzero correlation between social race, skeletal morphology, and geographic origin which has been maintained (at least in the United States) through assortative mating and institutional racism' (Dunn et al. 2020: 2). Our research, and its use of social media artefacts, was declared 'research ethics exempt' as per the Canadian Tri-council Policy Statement on the Ethical Conduct for Research Involving Humans, by the Carleton University Ethics Research Board.

1.2 The Trade in Human Remains on Social Media and E-Commerce Platforms

'This little beauty of a skull was brought back from Vietnam by an American soldier. Uncut and in amazing condition. Message me for more information and if you want to buy it. […] Worldwide shipping available […]' [link withheld; see **Figure 1**].

How can we know if this purported origin is true? This is the only data on this skull – this person's – existence. After purchase, it will disappear into someone's collection.

We have been documenting the existence of trafficking in a diverse range of human remains on frequently used social media platforms, such as Instagram, Facebook and various e-commerce platforms (Huffer & Graham 2017, 2018). As part of the sales process for human remains (or any other trafficked item), claims are often made about provenience (archaeological origin) and provenance (ownership history) that cannot be verified without having the item in front of experts after being seized by law or border agents (leaving aside rare examples where seizures lead to identifications, e.g. ICE 2011; Weisberger 2019; Yates 2019).

The trade in human remains conducted over Instagram is extensive. It cross-feeds into other platforms and out onto the 'regular' web via professional online storefronts. Posts like the one above capture many of its typical features–a brief 'backstory' that turns the skull into something more than 'mere' bone, notes on its condition, and directions on

how to initiate the purchase. A sequence of hashtags makes sure that the post will be found in various overlapping circles of interest. In this particular post, if the backstory is true, then we also have evidence of at least one crime since the law of war in the United States does not envision human remains as war trophies (the directives on 'war trophies' in operation in 1969, Army Regulation 608–4 Departments of the Army 1969 at least required that a soldier obtain permission to take a war trophy or souvenir, in writing. Thus, a legitimate war trophy would also have associated paperwork. While the directive of 1969 does not specifically exclude human remains, it does describe a wide variety of prohibited items in Section II.5, including objects 'of a household nature, objects of art or historical value… of scientific value' which presumably would cover this situation). Buying and selling of human remains is not prohibited in all jurisdictions, and exists in a legal grey area.

Since 2015 we have screen-captured numerous examples of commentary on images of human remains for sale in which buyers/sellers also ask for help estimating the age, sex, or probable ancestry of the human remains they recently obtained, or else offer competing interpretations. These requests for help demonstrate that at least a proportion of those engaged in this trade are not well-versed in the osteological methods and techniques necessary to have at least a general concept of *who* they are collecting, and therefore cannot verify the claims of sellers before buying. In any event, testimony from collectors themselves in media outlets such as Wired UK demonstrate that an important concern is that the item should be *real bone*; provenance, provenience and accurate demographic information are usually less important (Schwartz 2019).

There seem to be two main story-tropes that are told by collectors and dealers in relation to the 'ancestries' of the remains they acquire (Pokines 2015a; Pokines et al. 2017; Hefner et al. 2016; see also Huffer and Graham 2017). The first is that many remains, especially whole skeletons or crania, were stolen or sourced from British-controlled and post-Independence India to supply medical students during the 1800s to as recently as the 1980s. The second, told by niche collectors interested in 'tribal art' tends to argue that their authentic 'ethnographic' specimens were somehow *meant* to be collected by the Western explorers, missionaries or 'natural historians' who first acquired them. These 'ethnographic' materials tend to be crania or infracranial elements modified by Indigenous people for ritual use in the past or present, or as part of early 'curio' markets *ca.* 1800–1950. Twenty-first century osteoarchaeology can more readily acknowledge that many such collections now held by museums, but especially by private individuals, first came into global circulation for the purposes of 'scientific racism' during the emergence of physical anthropology as a discipline (Redman 2016). These two possible origin stories for specific categories of human remains circulating on today's market do not include the myriad examples of remains being actively looted from known or unknown prehistoric and historic archaeological sites, more recent open-air cemeteries, recovered from clandestine burials or found by chance (Pokines 2015b; Huxley & Finnegan 2004; Halling & Seidemann 2016; Seidemann, Stojanowski & Rich 2009).

Figure 1: A post on Instagram advertising a human skull for sale. Thousands of similar posts and their associated metadata can easily be found following the hashtags or usernames of individuals who comment or like a post.

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

255

Neither of these stories ethically absolve participants in this trade, but the former is often presented as 'ok' while the latter, since it involves the stealing of remains of First Nations and other Indigenous groups, is seen by traders as less morally sound, due to the existence of explicit legislation – see for instance the interviews with traders in (Schwartz 2019) and (Troian 2019a,b). It is also worth noting that the pre-1985 trade in skeletal remains by and for medical students in India and Bangladesh never disappeared, and we have observed that it continues largely within public and closed Facebook groups.

In other research we showed that there were ethical and technological problems with using neural networks to classify these images of human remains, or in using transfer-learning techniques which require thousands of images of a particular classification in order to work (Huffer & Graham 2018; Huffer, Wood & Graham 2019). It is unfeasible and impractical to try to create such training data in the domain of human remains. For instance, in the case of the latter technique, it would be easy to *mis-classify* a particular cultural grouping, and in any event, culture and ethnicity does not map to osteology. What is worse, such a tool could be used by unethical actors to give the authority of algorithms to a selling point: 'the computer identifies this skull as 100% Tibetan!' In those papers, we were classifying the *whole* image, including backgrounds, with the ambition of identifying visual tropes in the composition of the image. In this paper, we mask the backgrounds out and focus on understanding the patterns of *difference* in the images of the skulls. While it is easy to slip into a habit of thinking of the approach, and the results, as saying something about the skulls themselves, we must always remember that the approach is exploring the web of *differences* of the *images.*

1.3 Current approaches to ancestry estimation of human remains

Traditional forensic anthropological approaches to estimating ancestry, especially of remains recovered from crime scenes, clandestine graves, or sites of mass disaster, as well as unprovenanced remains recovered from the market, seek to quantify and qualify the complex interrelationship between skeletal morphology, genetics, geographic origin, and socio-cultural constructs (Pilloud & Hefner 2016; Dunn et al. 2020). While various researchers have and continue to attempt to develop regression equations to estimate ancestry from various infracranial elements (e.g. Liebenberg et al. 2015; Meeusen et al. 2015; Wescott 2005; Tallman & Winburn 2015; Ünlütürk 2017; Swenson 2013), it is the collection of a battery of metric measurements and non-metric/macromorphoscopic traits from ideally intact crania that are considered the most reliable. Ideally ancestry estimation would occur as part of a suite of interdisciplinary research performed in collaboration with anthropologists and/or law enforcement to fully establish the biological profile and (as much as possible) the life history of the individual whose cranium was recovered from the market (e.g. Watkins et al. 2017; Dodrill et al. 2016). Given the ephemeral nature of what appears and disappears online, the preferred situation of being able to assess the remains in

person in controlled laboratory conditions is very rarely realized.

Machine learning and neural networks have been employed by forensic anthropologists and archaeologists since at least 2001 (see for instance Bell & Jantz, 2002) for purposes of estimating ancestry, but the confounding factor here is that the algorithms are often trained solely on the metric data obtained from careful measuring a test population of crania of known age, sex and ancestry (Ousley, 2016). For the reasons discussed above and the nature of the specific category of unprovenienced remains we are concerned with in this pilot-level experiment, being able to employ machine learning by these 'traditional' methods is also not possible. We simply cannot obtain the data that forensic identification uses.

Our method is by default not as good as actually being able to analyze remains in person, but given the nature of the evidence of the human remains in question here, *it might be as good as it gets.*

1.4 Neural Networks and One-Shot Learning

A neural network for image classification consists of a sequence of layers of 'neurons' or computational functions that accept an input (text, pixel values, etc.) and performs a transformation which then gets passed onto the next layer. The initial weights connecting the neurons are randomized; the network can be trained on a known dataset by backpropagating increasing or decreasing weights until errors are minimized and the network correctly learns its training dataset. By comparing the pattern that lights up when exposed to a particular image, against the aggregate patterns for known classified images, a network can output the probability that a new image is a member of a particular class of images. The problem with this approach is that it requires extremely large training datasets. It also requires that the training dataset have example images of what one is trying to classify. Knowing whether or not something is a member of a class requires multiple examples so that the model can learn the extent of the variability.

So-called 'one-shot' learning on the other hand is predicated instead on the idea that we have only a few examples of the domain we are interested in – even only one. Then, the trained model is presented with two images that the model has not encountered before – a person of interest – and a second photograph which may or may not contain that person, for instance. The model is able to determine whether that second photograph contains the person depicted in the first image. This approach uses *two* neural networks that have the same pattern of weights and activations. The two images are presented to the two networks, which convert each image to a vectorized representation (**Figure 2**). The networks are joined together (which is why this architecture is sometimes known as 'siamese networks') by a final loss function that determines the *dissimilarity* between the two vectors ('siamese networks' were first introduced in 1993 for the purposes of signature verification, see Bromley et al. 1993).

We are interested in this approach because as we wrote in our earlier experiments (Huffer & Graham 2018; Huffer, Wood & Graham 2019), the 'authority' of the algorithms of
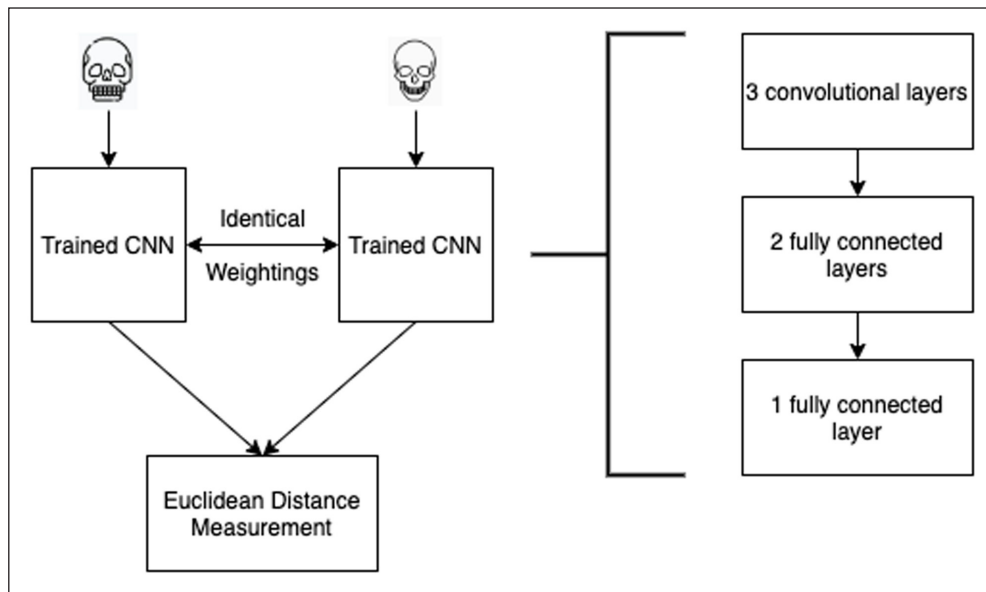
Figure 2: A schematic representation of the one-shot learning architecture, where two networks are used at once and a pair of images are evaluated. In practice, a single network where the images are evaluated in turn achieves the same effect at a great savings in memory. The neural network is implemented in the Pytorch framework, and may be explored in the Jupyter notebook file in our repository, 'Human_Remains_OSLNN_Triplet_Loss.ipynb'. Skull icons by user 'freepik' on flaticons.com, and user Jake Dunham, thenounproject.com.

classification could too easily be used in unethical ways, especially in the context of buying and selling human remains or antiquities without known provenance. For our purposes, dissimilarity is a better approach, because instead of saying what something *is*, we are saying what it is *not*. And we are interested in a series of 'is nots'.

For each image we study, we end up with a series of distance measures, alongside metadata describing whether or not the image has a secure provenance, and its ancestry estimation using the 3-group model (but see Dunn et al. 2020 for criticism of that model). We also have included in our dataset images of Indigenous skulls published in the 1940s from the United States that enable us to include a fourth category, 'Indigenous North America'. These distance measures can then be used to test whether or not the purported ancestry of the skull can be predicted.

2.1 Method

1. We use a convolutional neural network set up with a one-shot learning architecture using a triplet loss function. When pairs of images are dropped through the network one after the other, the difference between the two images perturbs the network, which we can measure. (NB: this is the same effect as having two identical networks and dropping a pair of images through once). This perturbation, or dissimilarity score, is expressed as Euclidean distance between the two image vectors.

2. Each pair, in the first iteration of the experiment, always contains one of six reference images (osteological study skulls), and one image from a corpus of photographs where 70 of the skulls depicted have grounded provenance, and 28 of the skulls depicted are from social media posts where we take the vendor at their word regarding what they say about the 'ancestry' of the skull. N = 98 study images.

3. Each photo in the corpus thus ends up with six dissimilarity scores, expressed as Euclidian distance from the reference images.

4. We then perform a mixture discriminant analysis on the scores for the grounded images, to see if the predicted 'ancestries' match with the observed 'ancestries'. We do the same again with a machine learning model.

5. Then, we take the scores for the grounded materials as our 'training' corpus, and use that to predict the 'ancestries' of our 'testing' materials, the images of skulls from social media. This enables us to suggest not only the 'ancestries' for these materials, but to explore the contrast with what the vendors say about the materials.

2.2 Data preparation for training the neural network model

Because neural networks are potentially sensitive to other elements of the photograph aside from the human remains themselves such as boxes in the background, the edges of windows, labels and so on, we removed the backgrounds from all images using the https://www.remove.bg service from (Kaleido 2018–2020), which itself is built on a neural network trained to recognize foreground versus background objects. The images used for training the neural network in the first place are not part of the N = 98 that we subsequently explored.

We augmented the training data set by adjusting the orientation, cropping, flipping the axes, and so on of that initial image (see for instance Shorten & Khoshgoftaar 2019). We automatically rotated, translated, and adjusted lighting so that we could account for the variability in the quality of target image, such that we build into the network knowledge of how skulls look under different conditions, both photographic and in terms of taphonomic condition

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

257

that themselves are indicative of primary burial/deposition and/or secondary storage conditions or use (e.g. Pokines 2015a,b; Yucha et al. 2017). We end up with 33 different views for each of the neural network training images, thus 363 images generated using standard data augmentation techniques (Tanner & Wong 1987; Van Dyke & Meng 2012). In this way, we can build a neural network representative of the very diverse category of 'the human skull'.

The images for training the neural network were sourced by DH from his research within several ethnographic and osteological collections over the last three years, including at: The Gustavianum Museum (Uppsala, Sweden), the Ethnographic Museum (Stockholm, Sweden), The Musée de L'Homme (Paris, France), the Sarawak Museum (Kuching, Malaysia), the Smithsonian National Museum of Natural History (Washington, DC, USA), the Tropenmuseum (Amsterdam, Netherlands), the Volkenkunde Museum (Leiden, Netherlands), The Pitt-Rivers Museum (Oxford, UK), the Oxford Museum of Natural History (Oxford, UK), the Duckworth Laboratory (Cambridge, UK).

When training the network, we use a 'triplet loss' function (**Figure 3**). We train the neural network on triplets where each triplet of images contains an anchor, a positive, and a negative. We select an 'anchor' image and then a 'positive' image, or the same object depicted from a different view; a 'negative' image is then selected from a different class as compared to the anchor, i.e. an image depicting a view of a different object. During training, we only select the 'hardest' triplets to train on which allows us to avoid spending valuable resources on evaluating 'easy' and 'semi-hard' triplets; the results of evaluating easy and semi-hard triplets do not meaningfully change the network weights to be worth spending computation time.

We find our 'hard' triplets by sampling a batch of images and for each valid anchor image selecting the hardest available positive (largest distance from anchor to positive) and hardest available negative (smallest distance from anchor to negative). The hard triplets are then used to update the neural network's weights. The advantage of this approach is that it teaches the network to detect subtle differences in the target domains (see Gómez 2019 for instance). More precisely put, we only select triplets where the negative is as visually similar as possible to the anchor while depicting a different class and where the positive is as visually distinct from the anchor as possible while depicting the same class; generally, a modified view of the anchor.

2.3 Creating the study dataset for testing

We created an initial dataset of 98 potential images, where their orientation faced the camera as square as possible. Spradley 2016 notes the difficulty of creating reference data for metric studies. Hefner (2018) is a new database of morphometric data, but for our purposes we need photographs of skulls rather than measurements of landmarks. 70 of the collected images are grounded in osteological study and so we know their provenance. They were sourced largely from forensic literature and from a further selection of photographs taken by DH in the institutions mentioned above (see 'raw-data/table-of-sourced-images.csv' in our data repository). The remaining 28 images came from our collection of materials from Instagram where the picture was a clear frontal view of an intact skull *and* the vendor provided a clear story regarding the provenience. While there are thousands of posts available, satisfying both of these requirements was more difficult and required visual examination of hundreds of images.
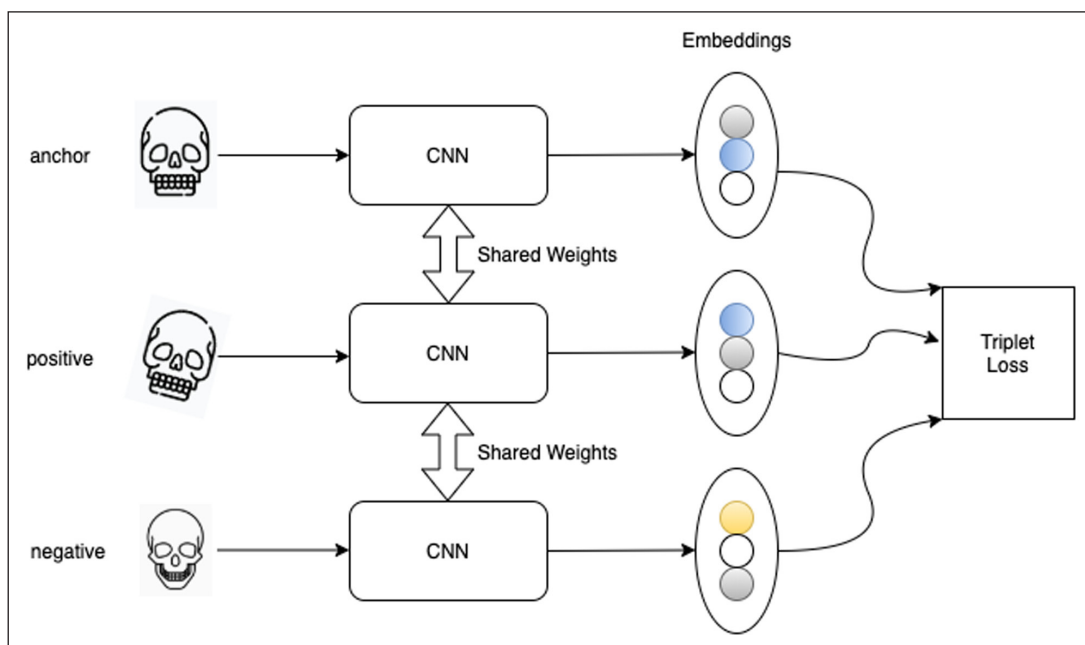
Figure 3: A schematic representation of triplet loss function on two images of skull A and one image of skull B (after Moindrot, 2018). While the 'anchor' and 'positive' images of skull A are different from each other, the resulting embeddings are closer to each other than they are to skull B's embeddings. The resulting network and its weights are then used in the one-shot learning architecture. Skull icons by user 'freepik' on flaticons.com, and user Jake Dunham, thenounproject.com.

2.4 Creating the reference image dataset against which we measure dissimilarity

We reasoned that patterns of *similar* dissimilarities in comparison to what we are calling 'references' might find useful groupings in the data that could shed light on the origins of the skulls depicted. That these are images of high-quality duplicates did not matter we reasoned as they were all created the same way, thus the dissimilarities being calculated should all measure from the same starting points (but see 'results and discussion below').

The 'reference' pictures chosen are all high-quality resin copies of crania with associated osteological reports, square to the camera, posted on boneclones.com:

African American male BC-110
European male BC-107
Asian male BC-253
Asian female BC-211
African American female BC-178
European female BC-891

In our experiments we found that the aspects of the skull that the neural network responds to seem to be the *same* things that anthropologists pay attention to, such as orbital shape (Gore et al. 2011). For some skulls, it attended to the nasal margin and the media orbital margin; on others it was the superior nasal margin, and sections of the left or right orbit. Sometimes for instance it was the interior nasal concha and right ethmoid and lacrimal bones. These aspects of maxilo-facial morphology such as orbital and nasal shape, zygomatic projection, alveolar projection, etc. are among those that forensic anthropologists pay attention to in order to determine ancestry estimates for unknown individuals. Markings on the skull, such as reference labels do attract attention, but in the context of the entire skull seems to make for a weaker signal that may or may not play a meaningful role.

3.1 Results and Discussion

Figure 4 depicts an initial visualization of the distance scores to each reference image. It shows that the distance scores are all highly positively correlated for the most part,
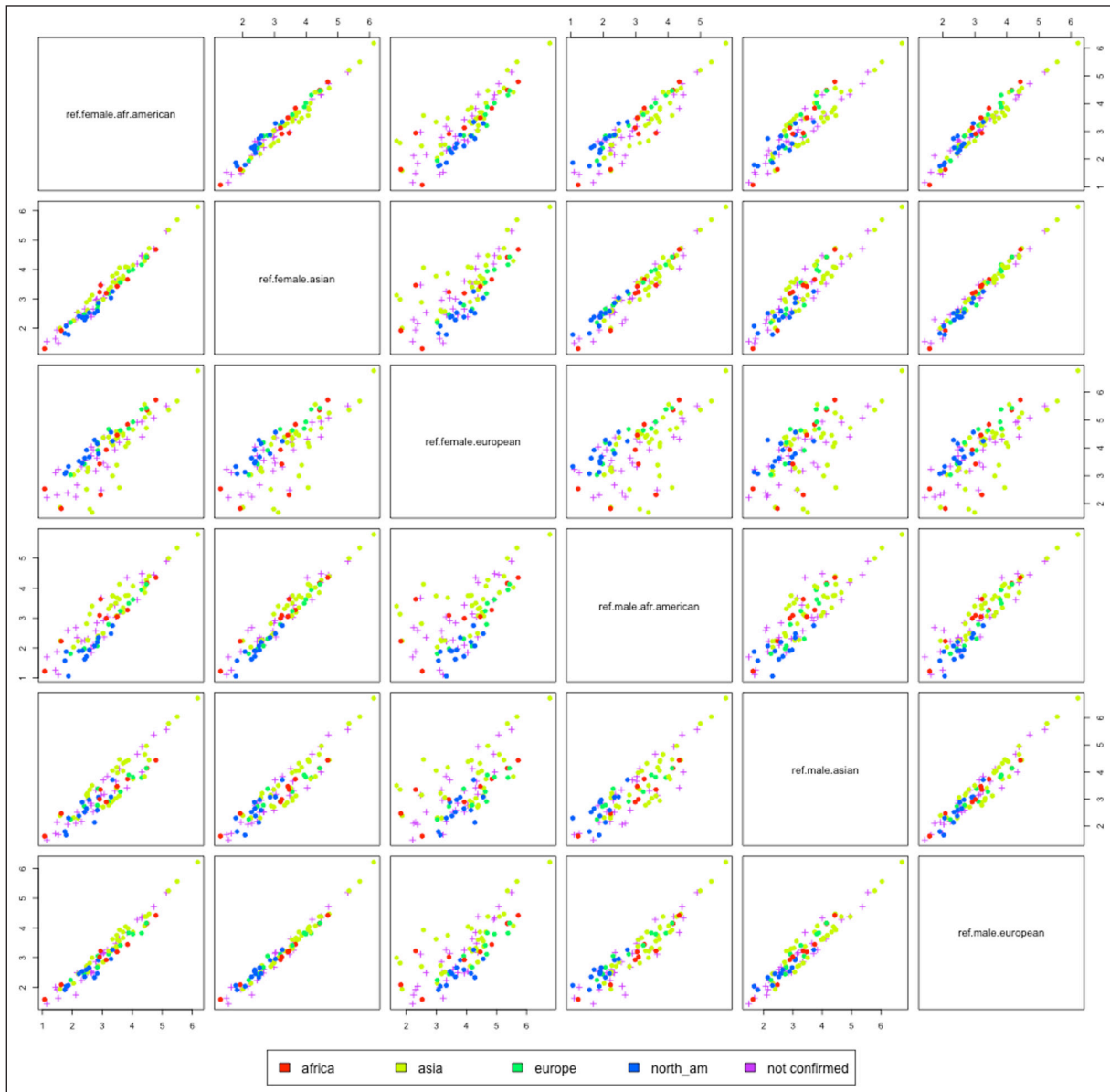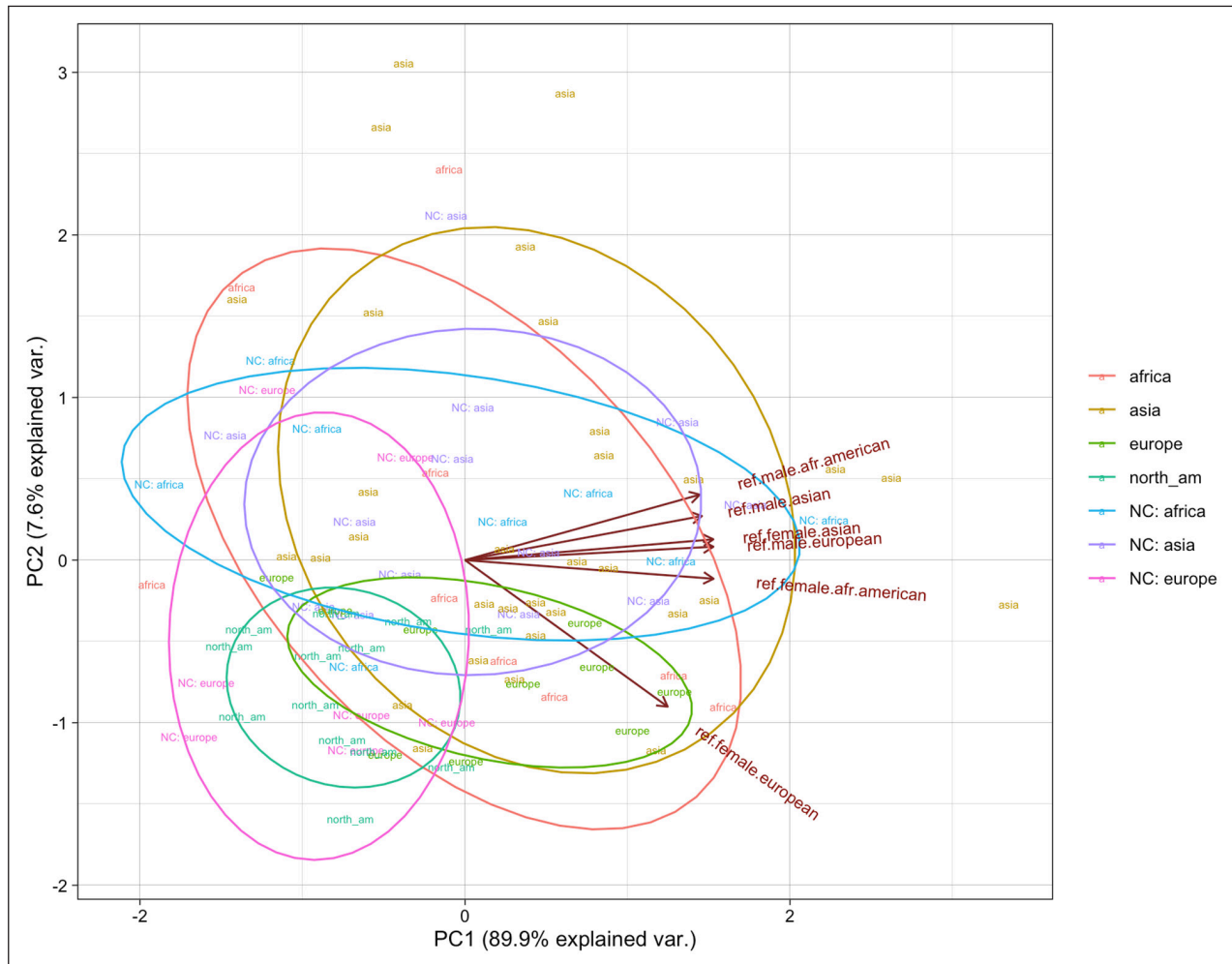

Figure 4: Raw dissimilarity scores comparing the study images to our initial group of 'reference' images.

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

259

Figure 5: Distribution of skull images according to reported origin (colors) along the two first principal components (97.5 per cent of variance) calculated through PCA using image-to-reference dissimilarity scores. The overlap of arrows indicates the overall strong correlation between dissimilarity scores. Note the variance range and the overlapping groups, and the mismatch between images with secured and non-confirmed origins.

which means that the more distant an image is to one reference, the more it is to any other reference. Stated differently, images tend to be equally similar to all reference images. Moreover, the groups formed by secured origin are spread over most of the variance range and overlap with each other greatly. Further multivariate analysis using Principal Components Analysis (PCA) summarises these trends (**Figure 5**).
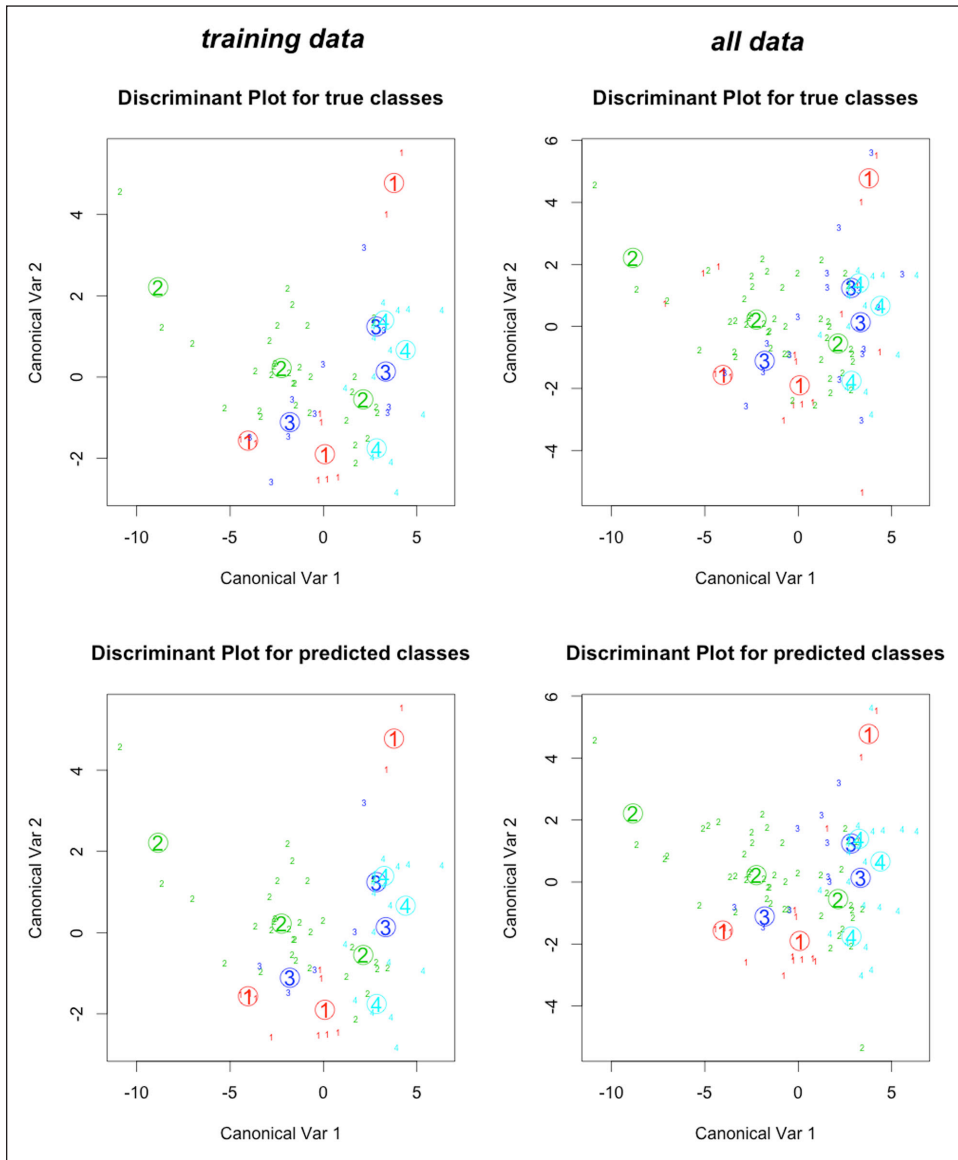
3.2 Mixture Discriminant Analysis

While linear discriminant functions have been used in the past to assess ancestry from craniometric data with some success (Giles & Elliott 1962), but also substantial critique upon further testing (e.g. Sauer, Wankmiller & Hefner 2009), we did not find that it gave results in our case any better than chance. We explored a variety of tests and found that MDA (Mixture Discriminant Analysis) was most suitable for our purposes – assessing whether a case belongs to a given category for our grounded materials. Since the subpopulations from which these materials are derived have different average metric dimensions and differing frequencies of macromorphoscopic trait expression, MDA is a good choice because MDA assumes each class is a mixture of subgroups following a Gaussian distribution, instead of a single Gaussian distribution per class as in LDA.

Table 1: Confusion matrix showing the performance of the MDA model predicting the origin of the skulls in the training dataset, where only the distances from the original six 'reference' images were considered. Here N = 64 secure examples rather than 70 because the six reference images are not being tested against themselves.

predicted.train	africa	asia	europe	north_am
africa	9	0	2	0
asia	0	28	3	0
europe	0	2	3	0
north_am	0	2	2	13

We first fit an MDA model to our dissimilarity distance scores for the grounded materials, and try to predict the appropriate group (**Figure 6**).

The diagonal values in the confusion matrix (**Table 1**) indicates where the observed group and the predicted group matched; thus, our grounded data was correctly discriminated 83 per cent of the time. In the second part of our experiment, we divide our dataset into two groups, training and testing. The training group is the grounded

260

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

Figure 6: Comparison of the distribution of skulls (small numbers), subgroup centroids (large numbers) and groups (colors) on the two first canonical coordinates calculated through MDA using dissimilarities to the six reference images, distinguishing the training (left column) and full (right column) dataset, and the given (top row) and predicted (bottom row) origin.

Table 2: Confusion matrix showing the performance of the MDA model predicting the origin of the skulls in the unprovenanced dataset, where only the distances from the original six 'reference' images were considered.

predicted.test	africa	asia	europe	north_am
africa	1	2	1	0
asia	6	8	1	0
europe	0	3	1	0
north_am	1	0	4	0

materials; the testing group are the materials derived from social media.
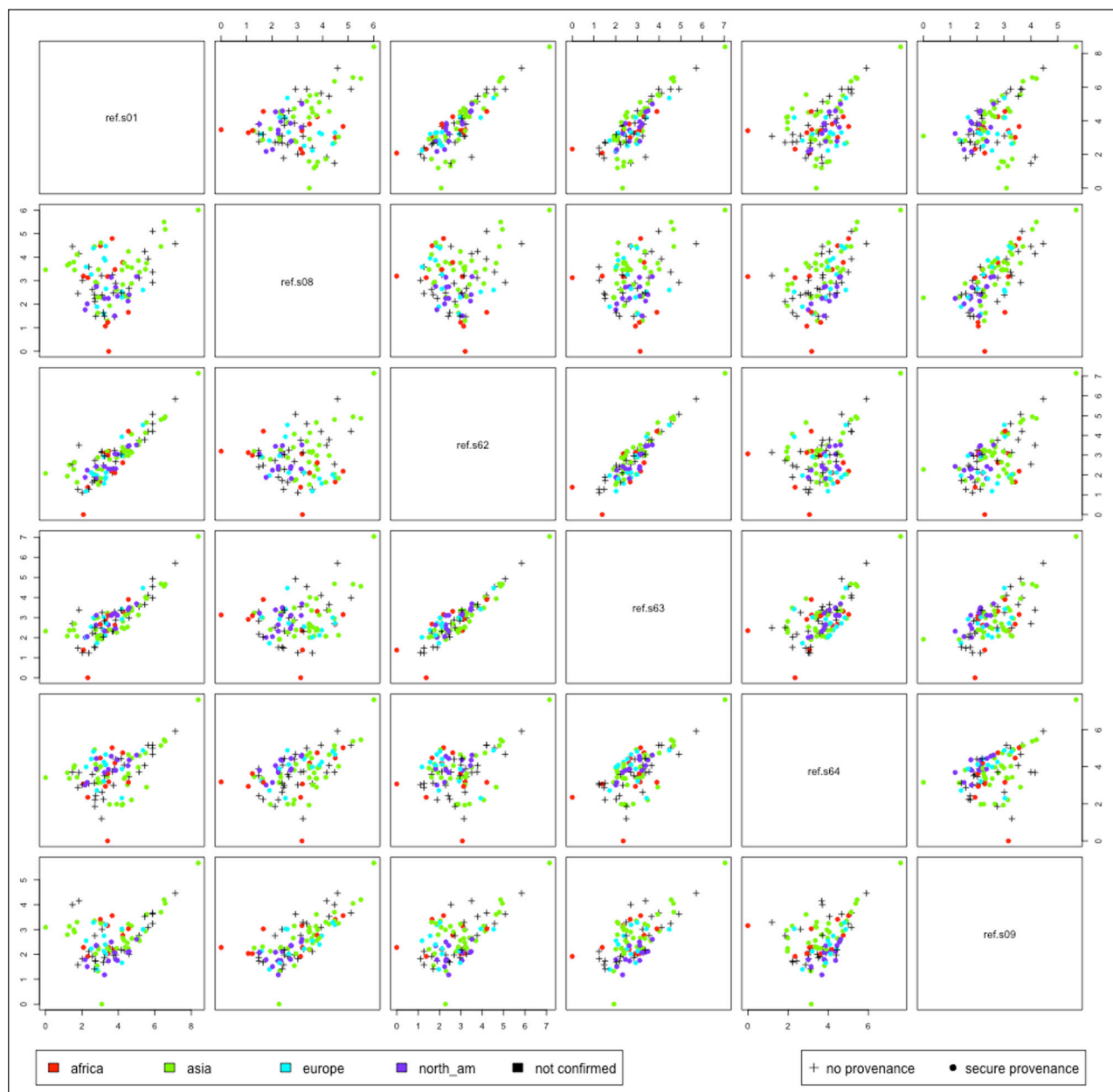
18 cases were mis-classified here (**Table 2**), meaning this model suggests over 2/3 of the ancestries claimed by vendors may be dubious.

3.3 Expanding the number of reference images and using non-metric multidimensional scaling

Given **Figures 4** and **5** it would appear that our choice to use 'reference images' which were images of high-quality resin copies was an error (our thinking had been they would function as a kind of neutral starting point from which to measure dissimilarities) – the neural network picked up on the differences (1) between standardized professional photos and photos taken under varying conditions, and (2) between resin models that were not aged or eroded by taphonomic processes and real osteological materials, hence the strong positive correlations.

We therefore re-ran the visual dissimilarity analysis by performing pair-wise comparisons for every image in our dataset, all 98 images, thus obtaining a matrix of 9,604 measurements (raw-data/square-materix-results. csv). A visual assessment of the covariances between

Graham et al: Towards a Method for Discerning Sources of Supply within the Human Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

261

Figure 7: Covariances of the dissimilarity scores between the first 6 elements of the square matrix of 70 images.

dissimilarities scores already shows promising results (**Figure 7**). Although correlations are generally lower, there is still indication of the influence of factors other than origin, as for example the sample batch.

In this iteration we created an MDA model using coordinates derived through non-metric multidimensional scaling (NMDS), which can process the dissimilarity matrix into an approximated projection of the points given a desired number of dimensions. We used the metaMDS function of the 'vegan' package in R, which performs NMDS trying to find a stable solution using random starts (Oksanen et al. 2019). Conceptually, these are like the components in a PCA, so each new dimension is the axis that represents most of the remaining variance (or dissimilarities between skulls). An advantage of NMDS is that we can preselect the number of dimensions to be calculated; the more dimensions, the less 'stress' the real distribution of points will suffer.

We explored how the number of NMDS dimensions affect the fitness of the MDA model with respect to the training and test data. We found that a two-dimensional NMDS projection is more than acceptable as a representation of the original dissimilarities. However, as expected, the MDA model predicts the training data better the more dimensions are included (**Figure 8**), with its best performance using 35 dimensions (**Figure 9**). This number is explained by the fact that the 70×70 matrix containing dissimilarities within the training data is symmetrical, and so half the number of rows or columns will suffice to capture 100 per cent of the variance. This progression in performance is, however, not the case for predicting social media claims which remains in the interval 25–50 per cent match with the MDA model predictions, suggesting that the origin stated in social media is often wrong.

We selected the MDA model created with 35 NMDS dimensions as the consolidated option, given it correctly
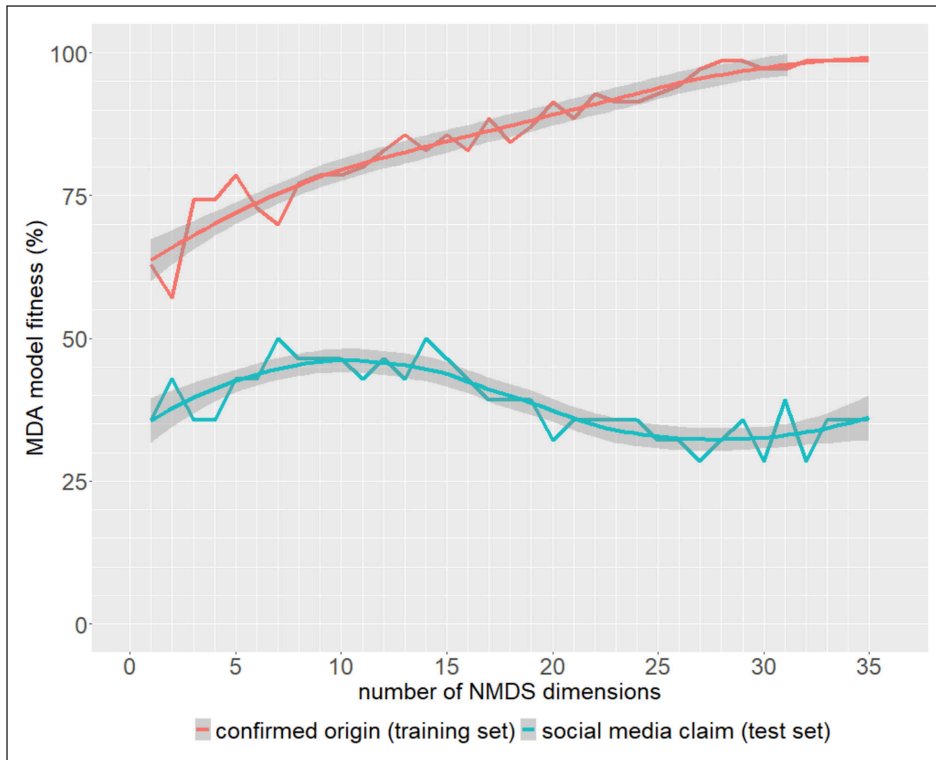
Figure 8: Showing how the number of NMDS dimensions affect the fitness of the MDA model with respect to the training and the test data.
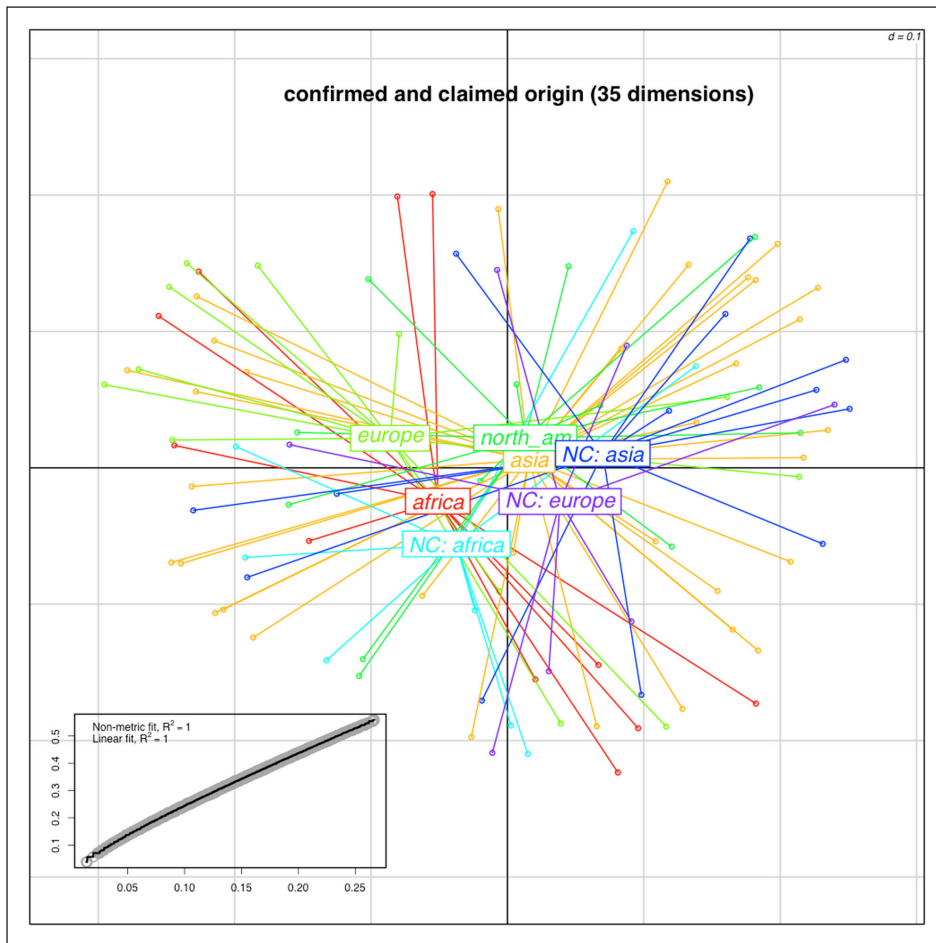
Figure 9: Distribution of skull images according to confirmed (training data) and non-confirmed (stated in social media) origin, along the two first dimensions of a 35-dimension NMDS projection. The stress plot (bottom-left) shows that the entire projection cover 100 per cent of the variance in the original data.

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

263

discriminates 100 per cent of our grounded materials (**Table 3**). Additionally, this approach achieves a greater spread of points over MDA canonical space which is helpful for delimiting subgroups (**Figure 10**). However, we are aware that there is a trade-off between fitting the training data and predicting new data, and that the model predictions on ungrounded materials must be put under a critical perspective.

With regard to the images of skulls from social media (**Table 4**), this model predicts 8 of 28 into the 'correct' group, while 20 are mis-classified.

The above analyses confirmed the importance of having a large enough dataset of securely grounded materials, where 'origin' has been determined by forensic anthropologists and osteologists, through direct observation or, even better, through morphological and genetic data. It would seem, however, that we can in fact use visual dissimilarity as determined by a neural network as a proxy measurement for predicting origin of materials on social media from a single photograph.

Table 3: Confusion Matrix for the provenanced materials, using all distances to provenanced materials for training.

predicted.train	africa	asia	europe	north_am
africa	11	0	0	0
asia	0	34	0	0
europe	0	0	12	0
north_am	0	0	0	13

Table 4: Confusion Matrix for the unprovenanced materials, using all distances to provenanced materials for training.

predicted.test	africa	asia	europe	north_am
africa	1	5	2	0
asia	4	6	1	0
europe	1	1	1	0
north_am	2	1	3	0
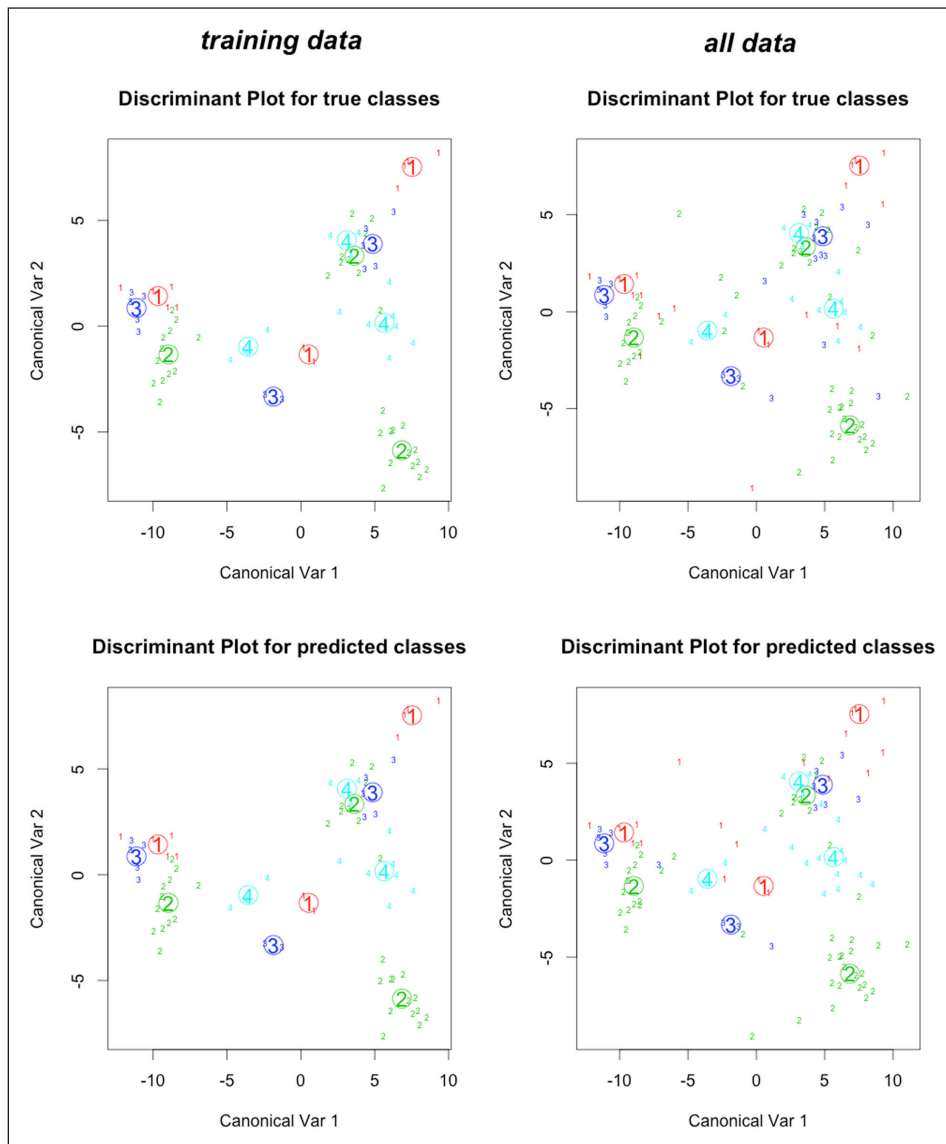
Figure 10: Comparison of the distribution of skulls (small numbers), subgroup centroids (large numbers) and groups (colors) on the two first canonical coordinates calculated through MDA using 35 NMDS dimensions, distinguishing the training (left column) and full (right column) dataset, and the given (top row) and predicted (bottom row) origin.

3.4 A deep learning model with TensorFlow

As a further experiment, we also considered the problem of predicting group membership with a deep learning approach, building a model with TensorFlow and the Ludwig deep learning toolbox by Molino, Dudin, and Miryala (2019). The idea here is to triangulate via a different method towards our same goal of prediction, and exploring how the predictions from the two methods coincide with each other and/or with vendor attributions of origin.

'Ludwig' allows one to build a model by specifying the training, testing, and validation 'hyperparameters' in a metadata text file, rather than having to write code. Hyperparameters are settings extrinsic to the data (training rate, optimizer settings, pre-processing procedures). Our model specification file is included in our code repository, as well as instructions for running Ludwig locally to train the model.

In this approach, the neural network attempts to learn a model from our square matrix of distances (every image compared to every other image), where it trains its model on our grounded materials. The grounded materials were divided at random into an 80–20 per cent split into 'training' and 'validation' sets, and then the trained model was used to predict (test) the accuracy of the social media posts. We explored a variety of settings by performing a 'grid search' (multiple training runs while manually modifying one hyperparameter at a time). We found that a model trained using the Adam optimizer and a learning rate of 0.00375, while preprocessing our distance measurements to turn them into z-scores, achieved the highest training accuracy of 80%.

3.5 Comparing vendors' attributions of origin to the model predictions

In the Venn diagram below (**Figure 11**, from **Table 5**), we can see where the stated ancestry (by the vendor) and the predicted ancestry (according to the two mod-

els) agreed or where they differed. Thus, it seems that skulls with an Indigenous North American ancestry are circulating in this market far more than vendors either know or let on: indeed, vendors are often quite careful to state that they would never knowingly trade in skulls from Indigenous groups. These results suggest in part that *knowingly* is the key word here; perhaps the correct synonym might be 'openly'. More skulls are claimed to be Asian than what the model predicts, as the historic trade in bodies from India and China perhaps provides 'moral cover'. According to this model, *none* of the purportedly 'europe' skulls can be so classified.

Table 5: Stated ancestry of each non-provenanced cranium vs. predicted category from MDA and Ludwig methods.

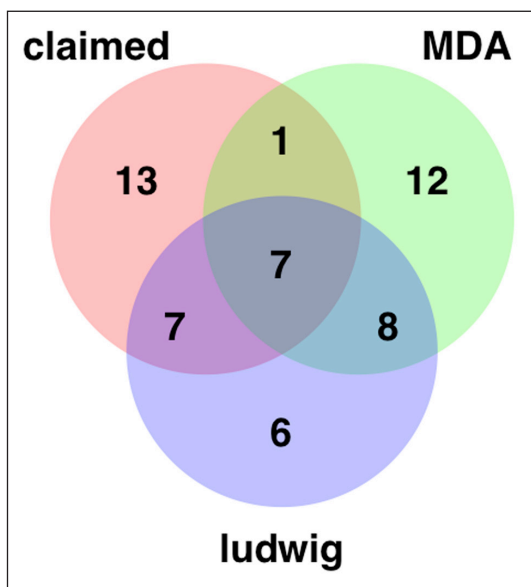| labels	Claimed origin	MDA model predicted origin	Ludwig model predicted origin
np01	europe	africa	north_am
np02	asia	africa	asia
np03	europe	asia	asia
np04	europe	north_am	north_am
np05	asia	asia	asia
np06	asia	africa	asia
np07	asia	europe	asia
np08	europe	north_am	asia
np09	asia	africa	asia
np10	europe	north_am	north_am
np11	asia	africa	asia
np12	africa	asia	asia
np13	africa	north_am	asia
np14	africa	asia	asia
np15	africa	asia	asia
np16	africa	asia	asia
np17	africa	north_am	north_am
np18	africa	africa	africa
np19	africa	europe	asia
np20	asia	north_am	asia
np21	asia	asia	asia
np22	asia	asia	asia
np23	asia	asia	asia
np24	asia	asia	asia
np25	asia	asia	asia
np26	europe	africa	north_am
np27	europe	europe	north_am
np28	asia	africa	asia

Figure 11: Venn diagram showing the overlap of stated ancestry of each non-provenanced cranium vs. predicted category from MDA and Ludwig methods.

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

265

Of the unprovenanced images of human skulls:

- 13 images have a purported 'ancestry' or origin that neither the MDA nor the Ludwig models predict.
- 7 images have a purported 'ancestry' that both the MDA and the Ludwig model also predict.
- 8 images have a purported 'ancestry' that both models predict, but is different from what the vendors claimed.
- There are 7 instances where the Ludwig model agreed with the claimed origin, but not with the MDA prediction.
- There is 1 instance where the MDA predicted origin agreed with the claimed origin, but not with the Ludwig origin.

Thus, the models lend support to the vendor's story in 15 cases overall, but only in 7 of those cases does it seem particularly strong, i.e. two different models predict the same origin (within all the usual caveats of this study). In another 18 cases, the models do not support the vendor, nor do they support each other, suggesting that those particular skulls might be worth further investigation, in that they have an origin not captured by our grounded examples. In 8 cases, the two models agree with each other, but not the vendor, which perhaps suggests cases where the vendor is either unknowledgeable or is misrepresenting the origin of the skulls. There are three cases where both models predict a North American origin, and a further four cases where one or the other model also points to a North American origin. This lends a certain amount of weight to the idea that North American Indigenous materials are a source of human remains not yet observed to be acknowledged on the social media platforms we track.

This has been an experiment; the mismatch between what the vendors *say* about the skulls and what the models *predict* lends weight to the idea that at least some collectors misrepresent the origins and life histories contained within the remains of the once-living individuals they have now reduced to commodities (either through a lack of concern for such details or ignorance of how to discern such details from the bones themselves). And the image of the skull purportedly a war trophy from Vietnam? This was np02 in our dataset. The MDA model predicts an 'african' ancestry, while the Ludwig model suggests 'asian'. The vendor's story is thus suspect.

To extrapolate further we need to understand the impact different training datasets can have, and to figure out how accumulating maps of a landscape of dissimilarities connects back to the 'real world'. Formal photographs of carefully curated reference collections from the major research centers are needed to test this potential method further. MDA with non-metric multidimensional scaling of the dissimilarity measurements seems to be a most productive avenue for further exploration. It is important to reiterate here that what we are exploring with our experiment and what the results show are measures of dissimilarity. The inferences we make on those grounds, of a similar ancestry (or not), are where one's archaeological knowledge intersects with algorithmic agency.

No predictive model will be as accurate as physically measuring a skull, and even there, ancestry estimation without DNA is always going to be broad. That is why, in a perfect world (and especially for seized alleged archaeological or ethnographic specimens), isotopes and DNA can, and where possible, should play a significant role in an analysis (as for instance with Watkins et al. 2017). Although the results of this initial experiment are promising, they are not without issues, and we invite reuse and critique of our code so that the method can be improved. We can identify several future directions that are worth exploring and areas of methodological improvement worth implementing in future iterations. These include:

- Obtaining more photographs from a wider variety of angles. While our one-shot neural network for measuring image dissimilarity was trained on images where the skull was not always quite square to the camera, we sought out photos that were square to the camera for all of our grounded photographs, and the tested photographs from social media. This may have contributed to the positive correlations seen in the initial PCA.
- Better, and more, reference images, preferably from forensic cases in which remains with secure ancestry estimations derived from in-person analysis and/or DNA are at hand.
- Exploration and capture of the variability between ancestral populations and subpopulations at a level more equivalent to that discernable by forensic anthropologists when analyzing unprovenienced remains in person.
- Using a explore-exploit strategy to neural network training (see Martin et al. 2018) to investigate whether or not this can improve accuracy and lower the number of required training epochs needed.

4. Conclusion

Photographs of human remains appear briefly on social media, and disappear again once the remains are sold. A single photograph might be the only evidence of a life lived. One-shot learning holds potential for us to be able to map a landscape of sourcing or broad-strokes geographic 'ancestry'. The resulting picture might be incorrect in terms of fine details, but taken at a more macroscopic level as a relative positioning vis-a-vis other remains, might be our best bet at understanding the broad patterns. Other similar approaches to this problem that rely on 3D scanning technology and photogrammetry are novel enough to require additional testing to refine, and also rely on physical access to the crania in question and obtaining tens to hundreds of photos from all angles of a single skull (e.g. Berezowski, Rogers & Liscio 2020).

Without ground-truthing, i.e. physically examining the skulls in question and ideally also performing stable and radiogenic isotope analysis and taking DNA (e.g. Watkins et al. 2017), we cannot source these skulls to a specific cultural group or localized region in the world with the certainty required to serve as evidence in a prosecution. But by situating what we do have, these one-off

photographs, through a neural network, we can begin to make a web of associations that allow us to discern more about this trade. That is, we could begin to assess the likelihood of accuracy in the professed 'professional' knowledge of certain collectors, or the degree to which terms like 'Dayak' and 'Asmat' and 'kapala' are used for marketing (rather than actual true descriptors of the remains in question).

With a refined model and method, what might we see if we could begin to quantify the numbers of individual humans bought and sold in this trade, and where they are coming from? Would we see patterns similar to what are observed in the larger antiquities trades, of source and destination countries, as multi-year systematic research has revealed (Mackenzie et al. 2020)? Or would it be similar to the patterns of opportunistic looting documented for Facebook (Al-Azm & Paul 2019)? Achieving these goals to the level of consistency and reliability needed to be useful to investigations is a long way off, but we suggest here that the outcome of this experiment warrants further development of this research direction.

Additional File
The additional file for this article can be found as follows:

- **59-1047-2-SP.zip.** Supporting code and data, mirroring release at github.com/bonetrade/visual-dissimilarity. DOI: https://doi.org/10.5334/jcaa.59.s1

Acknowledgements

Competing Interests
The authors have no competing interests to declare.

References
Bell, S and **Jantz, R.** 2002. Neural Network Classification of Skeletal Remains. In: Burenhult, G and Arvidsson, J (eds.), Archaeological Informatics: Pushing The Envelope. CAA2001. *Computer Applications and Quantitative Methods in Archaeology.* Proceedings of the 29th Conference, Gotland, April 2001 (BAR International Series 1016). Archaeopress: Oxford, 205–212.

Berezowski, V, Rogers, T and **Liscio, E.** 2020. Evaluating the morphological and metric sex of human crania using 3-dimensional (3D) technology. *International Journal of Legal Medicine.* DOI: https://doi.org/10.1007/s00414-020-02305-0

Bromley, J, Bentz, JW, Bottou, L, Guyon, I, LeCun, Y and **Moore, C.** 1993. Signature verification using a siamese time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence,* 7(4): 669–688. DOI: https://doi.org/10.1142/S0218001493000339

Departments of the Army, the Navy, and the Air Force. 1969. Personal affairs control and registration of war trophies and war trophy firearms. Available at https://www.marines.mil/Portals/1/Publications/MCO%205800.6A.pdf [Last accessed 07 July 2020].

Dodrill, TN, Nelson, GC, Stone, JH and **Fitzpatrick, SM.** 2016. Determining ancestry of unprovenienced human remains from the Grenadines, Southern Caribbean: Dental morphology and craniometric analyses. *Poster presented at: Undergraduate Research Symposium 2016,* University of Oregon. https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/19937/Dodrill_UGRS_2016.pdf?sequence=1.

Dunn, RR, Spiros, MC, Kamnikar, KR, Plemons, AM and **Hefner, JT.** 2020. Ancestry estimation in forensic anthropology: A review. *WIREs Forensic Science,* e1369. DOI: https://doi.org/10.1002/wfs2.1369

Giles, E and **Elliot, O.** 1962. Race identification from cranial measurements. *Journal of Forensic Sciences,* 7(2): 147–157.

Graham, S, Lane, A, Huffer, D and **Angourakis, A.** 2020. Visual Dissimilarity. *Github.com.* Available at https://github.com/bonetrade/visual-dissimilarity. DOI: https://doi.org/10.5281/zenodo.3979350 [Last accessed 07 July 2020].

Gómez, R. 2019. Understanding ranking loss, contrastive loss, margin loss, triplet loss, hinge loss and all those confusing names. *Raúl Gómez blog.* https://gombru.github.io/2019/04/03/ranking_loss/ [Last accessed 08 August 2019].

Gore, T, Nawrocki, SP, Langdon, J and **Bouzar, N.** 2011. The use of elliptical fourier analysis on orbit shape in human skeletal remains. In: Lestrel, PE (ed.), *Biological Shape Analysis.* Proceedings of the 1st International Symposium, Tsukuba, Japan, 3–6 June 2009, 242–265. Tokyo: World Scientific Publishing Co. DOI: https://doi.org/10.1142/8181

Halling, CL and **Seidemann, RM.** 2016. They sell skulls online?! A review of internet sales of human skulls on eBay and the laws in place to restrict sales. *Journal of Forensic Sciences,* 61(5): 1322–26. DOI: https://doi.org/10.1111/1556-4029.13147

Hefner, JT. 2018. The macromorphoscopic databank. *American Journal of Physical Anthropology,* 166(4): 994–1004. DOI: https://doi.org/10.1002/ajpa.23492

Hefner, JT, Spatola, BF, Passalacqua, NV and **Gocha, TP.** 2016. Beyond taphonomy: Exploring craniometric variation among anatomical material. *Journal of Forensic Sciences,* 61(6): 1440–1449. DOI: https://doi.org/10.1111/1556-4029.13177

Huffer, D and **Graham, S.** 2017. The Insta-Dead: The rhetoric of the human remains trade on Instagram. *Internet Archaeology,* 45(5). DOI: https://doi.org/10.11141/ia.45.5

Huffer, D and **Graham, S.** 2018. Fleshing out the bones: Studying the human remains trade with Tensorflow and Inception. *Journal of Computer Applications*

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

267

in Archaeology, 1(1): 55–63. DOI: https://doi.org/ 10.5334/jcaa.8

Huffer, D, Wood, C and **Graham, S.** 2019. What the machine saw: Some questions on the ethics of computer vision and machine learning to investigate human remains trafficking. *Internet Archaeology*, 52(1). DOI: https://doi.org/10.11141/ia.52.1

Huxley, A and **Finnegan, M.** 2004. Human remains sold to the highest bidder! A snapshot of the buying and selling of human skeletal remains on eBay®, an internet auction site. *Journal of Forensic Sciences*, 49(1): 1–4.

ICE (Immigration and Customs Enforcement). 2011. ICE returns tribal artifacts to Indonesian authorities. *ICE.gov/news.* https://www.ice.gov/news/releases/ ice-returns-tribal-artifacts-indonesian-authorities [Last accessed 21 July 2019].

Kaleido. 2018–2020. Remove Image Background. Available at http://remove.bg [Last accessed 15 July 2020].

Liebenberg, L, L'Abbé, EN and **Stull, KE.** 2015. Population differences in the postcrania of modern South Africans and the implications for ancestry estimation. *Forensic Science International*, 257: 522–529. DOI: https://doi.org/10.1016/j.forsciint. 2015.10.015

Mackenzie, S, Brodie, N, Yates, D and **Tsirogiannis, C.** 2020. Trafficking Culture: New Directions in Researching the Global Market in Antiquities. New York: Routledge. DOI: https://doi.org/10.4324/97 81315532219

Maier, CA, Zhang, K, Manhein, MH and **Li, X.** 2015. Palate shape and depth: A shape-matching and machine learning method for estimating ancestry from human skeletal remains. *Journal of Forensic Sciences*, 60(5): 1129–1134. DOI: https://doi.org/ 10.1111/1556-4029.12812

Martin, K, Wiratunga, N, Massie, S and **Clos, J.** 2018. Informed pair selection for self-paced metric learning in siamese neural networks. In: Bramer, M and Petridis, M (eds), *Artificial Intelligence XXXV. SGAI 2018. Lecture Notes in Computer Science*, 11311: 34–49. Cham: Springer. DOI: https://doi. org/10.1007/978-3-030-04191-5_3

Meeusen, RA, Christensen, AM and **Hefner, JT.** 2015. The use of femoral neck axis length to estimate sex and ancestry. *Journal of Forensic Sciences*, 60(5): 1300–1304. DOI: https://doi.org/10.1111/ 1556-4029.12820

Moindrot, O. 2018. Triplet Loss and Online Triplet Mining in TensorFlow. *Olivier Moindrot blog* [Last accessed 15 June 2020].

Molino, P, Dudin, Y and **Miryala, SS.** 2019. Ludwig: a type-based declarative deep learning toolbox. *arXiv:1909.07930.*

Nagpal, S, Maneet, S, Jain, A, Singh, R, Vatsa, M and **Noore, A.** 2017. On matching skulls to digital face images: A preliminary approach. *IEEE International Joint Conference on Biometrics (IJCB)*, 813–819. DOI: https://doi.org/10.1109/BTAS.2017.8272775

Navega, D, Coelho, C, Vicente, R, Ferreira, MT, Wasterlain, S and **Cunha, E.** 2015a. AncesTrees: ancestry estimation with randomized decision trees. *International Journal of Legal Medicine*, 129: 1145–1159. DOI: https://doi.org/10.1007/ s00414-014-1050-9

Navega, D, Vicente, R, Vieira, DN, Ross, AH and **Cunha, E.** 2015b. Sex estimation from the tarsal bones in a Portuguese sample: A machine learning approach. *International Journal of Legal Medicine*, 129: 651–659. DOI: https://doi.org/10.1007/ s00414-014-1070-5

Oksanen, J, Guillaume Blanchet, F, Friendly, M, Kindt, R, Legendre, P, McGlinn, D, Minchin, PR, O'Hara, RB, Simpson, GL, Solymos, P, Henry, M, Stevens, H, Szoecs, E and **Wagner, H.** 2019. vegan: Community Ecology Package. *R package version 2.5-6.* https://CRAN.R-project.org/package=vegan.

Ousley, SD. 2016. Forensic classification and biodistance in the 21st Century: The rise of learning machines. In: Pilloud, MA and Hefner, JT (eds.), *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*, 197–212. New York: Elsevier, Inc. DOI: https://doi.org/10.1016/ B978-0-12-801966-5.00010-X

Pilloud, MA and **Hefner, J.** (eds.) 2016. Biological Distance Analysis: Forensic and Bioarchaeological Perspectives. New York: Elsevier, Inc.

Pokines, JT. 2015a. Taphonomic characteristics of former anatomical teaching specimens received at a medical examiner's office. *Journal of Forensic Identification*, 65(2): 173–195.

Pokines, JT. 2015b. A Santería/Palo Mayombe ritual cauldron containing a human skull and multiple artifacts recovered in western Massachusetts, U.S.A. *Forensic Science International*, 248: e1–7. DOI: https://doi. org/10.1016/j.forsciint.2014.12.017

Pokines, JT, Appel, N, Pollock, C, Eck, CJ, Maki, AG, Joseph, AS, Cadwell, L and **Young, CD.** 2017. Anatomical taphonomy at the source: Alterations to a sample of 84 teaching skulls at a medical school. *Journal of Forensic Identification*, 67(4): 600–32.

Redman, SJ. 2016. Bone rooms: From scientific racism to human prehistory in museums. Harvard: Harvard University Press. DOI: https://doi.org/10. 4159/9780674969711

Sauer, NJ, Wankmiller, JC and **Hefner, JT.** 2009. The assessment of ancestry and the concept of race. In: Blau, S and Ubelaker, DH (eds.), *Handbook of Forensic Anthropology and Archaeology*, 243–261. New York: Routledge.

Schwartz, O. 2019. Instagram's grisly human skull trade is booming. *Wired UK.* https://www.wired.co.uk/ article/instagram-skull-trade [Last accessed 08 August 2019].

Seidemann, RM, Stojanowski, CM and **Rich, FJ.** 2009. The identification of a human skull recovered from an eBay sale. *Journal of Forensic Sciences*, 54(6): 1247–1253. DOI: https://doi.org/10.1111/ j.1556-4029.2009.01194.x

268

Graham et al: Towards a Method for Discerning Sources of Supply within the Human
Remains Trade via Patterns of Visual Dissimilarity and Computer Vision

Shorten, C and **Khoshgoftaar, TM.** 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(60). DOI: https://doi.org/10.1186/s40537-019-0197-0

Spradley, MK. 2016. Metric methods for the biological profile in forensic anthropology: Sex, ancestry, and stature. *Academic Forensic Pathology*, 6(3): 391–399. DOI: https://doi.org/10.23907/2016.040

Swenson, VM. 2013. Ancestral and Sex Estimation Using E.A. Marino's Analysis of the First Cervical Vertebra Applied to Three Modern Groups. Unpublished MA thesis, Department of Anthropology, University of Montana.

Tallman, SD and **Winburn, AP.** 2015. Forensic applicability of femur subtrochanteric shape to ancestry assessment in Thai and White American males. *Journal of Forensic Sciences*, 60(5): 1283–1289. DOI: https://doi.org/10.1111/1556-4029.12775

Tanner, MA and **Wong, WH.** 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398): 528–540. DOI: https://doi.org/10.1080/01621459.1987.10478458

Troian, M. 2019a. Federal Conservative candidate gives boyfriend human skull for birthday. *APTN News*. https://aptnnews.ca/2019/07/03/federal-conservative-candidate-gives-boyfriend-human-skull-for-birthday/ [Last accessed 08 August 2019].

Troian, M. 2019b. Human skull purchased from oddity shop by Conservative candidate, likely an orphaned skull says owner. *APTN News*. https://aptnnews.ca/2019/07/08/human-skull-purchased-from-oddity-shop-by-conservative-candidate-likely-an-orphan-skull-says-owner/ [Last accessed 08 August 2019].

Ünlütürk, Ö. 2017. Metric assessment of ancestry from the vertebrae in South Africans. *International Journal of Legal Medicine*, 131: 1123–1131. DOI: https://doi.org/10.1007/s00414-016-1483-4

Van Dyke, DA and **Meng, XL.** 2012. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1): 1–50. DOI: https://doi.org/10.1198/10618600152418584

Watkins, JK, Blatt, SH, Bradbury, CA, Alanko, GA, Kohn, MJ, Lytle, ML, Taylor, J, Lacroix, D, Nieves-Colón, MA, Stone, AC and **Butt, DP.** 2017. Determining the population affinity of an unprovenienced human skull for repatriation. *Journal of Archaeological Science: Reports*, 12: 384–394. DOI: https://doi.org/10.1016/j.jasrep.2017.02.006

Weisberger, M. 2019. Stolen mummy feet, arms and more found stashed in speakers at Cairo airport. *Livescience.com*. https://www.livescience.com/64851-mummy-parts-recovered-airport.html [Last accessed 21 July 2019].

Wescott, D. 2005. Population variation in femur subtrochanteric shape. *Journal of Forensic Sciences*, 50(2): 281–288. DOI: https://doi.org/10.1520/JFS2004281

Yates, D. 2019. Cultural heritage offences in Latin America: Textile traffickers, mummy mailers, silver smugglers, and virgin vandals. In: Hufnagel, S and Chappell, D (eds.), *The Palgrave Handbook on Art Crime*, 483–501. London, UK: Palgrave Macmillan. DOI: https://doi.org/10.1057/978-1-137-54405-6_23

Yucha, JM, Pokines, JT and **Bartelink, EJ.** 2017. A comparative taphonomic analysis of 24 trophy skulls from modern forensic cases. *Journal of Forensic Sciences*, 62(5): 1266–1278. DOI: https://doi.org/10.1111/1556-4029.13426