IberianVoxel: Automatic Completion of Iberian Ceramics for Cultural Heritage Studies

Pablo Navarro1*, Celia Cintas2+, Manuel Lucena3, José Manuel Fuertes3, Antonio Rueda3, Rafael Segura7, Carlos Ogayar-Anguita4, Rolando González-José1 and Claudio Delrieux5

1 Instituto Patagónico de Ciencias Sociales y Humanas (CONICET). Puerto Madryn, Argentina.
2 IBM Research Africa, Nairobi, Kenya.
3 Research University Institute for Iberian Archaeology, University of Jaén, Spain.
4 Center for Advanced Studies in Information and Communication Technologies, University of Jaén, Spain.
5 Universidad Nacional del Sur, and CONICET. Bahía Blanca, Argentina.
*pnavarro@cenpat-conicet.gob.ar, +celia.cintas@ibm.com

Abstract

Accurate completion of archaeological artifacts is a critical aspect in several archaeological studies, including documentation of variations in style, inference of chronological and ethnic groups, and trading routes trends, among many others. However, most available pottery is fragmented, leading to missing textural and morphological cues. Currently, the reassembly and completion of fragmented ceramics is a daunting and time-consuming task, done almost exclusively by hand, which requires the physical manipulation of the fragments. To overcome the challenges of manual reconstruction, reduce the materials’ exposure and deterioration, and improve the quality of reconstructed samples, we present IberianVoxel, a novel 3D Autoencoder Generative Adversarial Network (3D AE-GAN) framework tested on an extensive database with complete and fragmented references. We generated a collection of 1001 3D voxelized samples and their fragmented references from Iberian wheel-made pottery profiles. The fragments generated are stratified into different size groups and across multiple pottery classes. Lastly, we provide quantitative and qualitative assessments to measure the quality of the reconstructed voxelized samples by our proposed method and archaeologists’ evaluation.

1 Introduction

During archaeological excavations, it is common to find fractured or damaged artifacts. Among the many factors, ceramic potteries are one of the most frequently discovered archaeological artifacts. Since they are usually short-lived, researchers find these artifacts useful to analyze chronological and geographical features, given that shape and decoration are subject to significant changes over time and space [Es- lami et al., 2020]. This analysis gives a basis for dating the archaeological strata, and provides evidence from a large set of valuable data, such as local production, trade relations, and consumer behavior of the local population [Kampel and Sablatnig, 2003; Kashihara, 2017].

Several prior studies analyze various aspects of ceramics using 2D complete pottery profiles. Automatic profile classification [Lucena et al., 2017; Cintas et al., 2020; Llamas et al., 2016; Di Angelo et al., 2017] and feature extraction [Shennan and Wilcock, 1975; Rice, 1987; Nautiyal et al., 2006; Mom, 2007; Saragusti et al., 2005; Karasik and Smilansky, 2011; Smith et al., 2014] have been widely studied, ranging from traditional image processing techniques to deep learning approaches. Unfortunately, ceramics are fragile, and therefore most of the actual ceramics recovered from archaeological sites are fractured, so the vast majority of the available material appears in fragments. The reassembly of the fragments is a daunting, delicate, and time-consuming task, done almost exclusively by hand, which requires the physical manipulation of the fragments, which ideally should be as short as possible for conservation purposes. Due to these factors and the rise of easy-to-use and affordable 3D scanners1, there is an increasing interest in automatic pottery reassembly and reconstruction [Rasheed and Nordin, 2014; Fragkos, Stergios et al., 2018; Kalasarinis and Koutsoudis, 2019] and fragment analysis [Wilcek et al., 2021]. Nonetheless, existing work resolves the fragments problem using comparisons between known pieces and mainly works on 2D representations [Navarro et al., 2022].

In this work, we propose a 3D Autoencoder Generative Adversarial Network (3D AE-GAN) framework to augment and reconstruct ceramics, in which the “best-matching pottery” is artificially generated based on a set of known 3D fragments in the model, thus providing reconstructed virtual pottery with the same geometric features as the real ones. The resulting augmented dataset will be very helpful for exploring and designing automatic procedures to aid experts with the pottery completion task.

13D scanners. (Last Accessed in January 2023)
The main contributions of this paper are the following:

- A 3D Autoencoder Generative Adversarial Network (3D AE-GAN), that aims to complete pottery automatically. Our method generates new ceramics from 156546 fragments (see Fig. 1A).
- A procedure for generating artificial 3D fragment samples from voxelized pottery models by adapting the Discrete Voronoi Chain algorithm (see Fig. 1B).
- We release a new database of 3D ceramics fragments generated from 1001 voxelized pottery models fragmented 16 times (in total 16016 models) with matching correspondence of fragments and full pottery.
- We evaluate the proposed approach with standard metrics (MSE, DSC, and Jaccard) under different fragment sizes (see Fig. 1C for results). Additionally, we validated our framework with a binary shape classifier. Lastly, we conduct a validation study with a group of archaeologist evaluators to qualitatively assess the quality of the completed samples with respect to domain experts.

2 Related Work

Artifact completion, reassembly, and restoration are key research topics in Computational Archaeology. There are several recent examples of GANs applied to cultural heritage do-
mains in the literature [Nguyen et al., 2019; Hermoza and Sipiran, 2018]. For instance, techniques such as automated image style transfer [Chen et al., 2020] were used to develop a model aimed at generating Cantonese porcelain-styled images departing from user-defined masks. Similar techniques were also applied to material degradation [Papadopoulos et al., 2021; Liu et al., 2017a; Isola et al., 2017; Zhu et al., 2017; Wang et al., 2017; Zachariou et al., 2020].

In a similar approach to our proposal, in [Hermoza and Sipiran, 2018] a 3D reconstruction GAN for archaeological restoration is proposed, based on an encoder-decoder 3D CNN on top of a GAN based on eGANs [Mirza and Osindero, 2014]. In [Kniaz et al., 2019], a translation network is presented. Both these studies work on the problem of prediction of missing geometry on damaged voxelized objects. Another example of cultural heritage preservation can be found in [Jboor et al., 2019], where an image completion approach is adapted [Liu et al., 2017b; Yeh et al., 2017; de Lima-Hernandez and Vergauwen, 2021] for the curation of damaged artwork by predicting the continuing decoration traces of broken heritage fragments in 2D. Finally, [Navarro et al., 2022] presents a method based on a GAN for creating 2D complete profile potteries from binary images of fragments. Even though these works show promising results in generating Iberian potteries, the previous proposed processes for managing fragments and fragmentation are unrealistic.

3 Proposed Approach

A typical GAN [Goodfellow et al., 2014] framework contains a generative (G) and a discriminative (D) neural network such that G(z) aims (in our context) to generate realistic artifacts, while D(x) learns to discriminate if a sample is from the real data distribution (px(x)) or not. D(x) should be high when x comes from training data and low when x comes from the generator. In a classic GAN, z is a latent space vector sampled from a statistical distribution (usually normal distribution).

In our case, as opposed from [Wu et al., 2016], z is a voxelized input fragment (matrix of 32 × 32 × 32), and G(z) represents the generator function which maps the fragment z to the data space of a complete Iberian voxelized pottery. This type of network is named Autoencoding GAN (AE-GAN), in which G added a network of encoders that are trained to learn an E : Z → W function, mapping each fragment sample to a point (w) in latent space [Lazarou, 2020] and the decoder learn to map each point (w) to a complete pottery.

Multiple iterations will inform G on how to adjust the generation process to generate a misclassification in D. In our particular case, the data element x, corresponds to a three-dimensional binary array containing the voxelized pottery geometry. D(G(z)) is the probability that the output of the generator G is a real artifact from the Iberian pottery 3D dataset. D tries to maximize (log(D(x))), which is the probability of having a correct classification of actual voxelized real ceramics, while G tries to minimize (log(1 − D(G(z)))) which is the probability of D recognizing any of the generated outputs by G (see Eq. 1).

IberianVoxel is based on the AE-3DGAN, where the generator is an Autoencoding network Encode(z) → w ∈ R32, Decode(w) → x′, where x ∈ [0,1]32×32×32 is the input fragment, a three-dimensional binary array containing the fragment shape information by means of voxels, and x′ is a missing generated part. To train the discriminator network, we use D(y) where y = z + x′ for the generated examples. The complete codebase for the definition, implementation, training, and evaluation of IberianVoxel can be found here2.

4 Experimental Setup

In this section, we describe the process taken for the generation of the new 3D fracture dataset and pre-processing steps for the voxelization (see Fig. 2). Further, we define hyperparameters, architecture, and training steps implemented for IberianVoxel. We detailed the different assessments done to evaluate the performance of our proposed model using multiple initial fragment sizes. We evaluated the samples generated with standard completion and segmentation metrics. Additionally, we validated the output of IberianVoxel with a pottery shape classifier. Lastly, we show the protocol done to conduct a validation study with domain experts.

4.1 Iberian Pottery Data

The raw data belong to binary profile images, corresponding to Iberian wheel-made pottery from various archaeological sites of the upper valley of the Guadalquivir River (Spain). The ceramics are classified into eleven different classes based on their shape. These classes consider the forms of the lip, neck, body, base, and handles and the relative ratios between their sizes. Nine of these classes correspond to closed pottery shapes, and two others belong to open ones [Lucena et al., 2017]. The available images consist of a profile view of the pottery, where image resolutions (in pixels), corresponding to size scale, may vary according to the acquisition settings (see Fig. 2).

---

Table 1: Number of fragments generated by our proposed DVC adaptation per each ceramic type and size.

<table>
<thead>
<tr>
<th>Class</th>
<th>15-20</th>
<th>20-30</th>
<th>30-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>246 (2.76%)</td>
<td>194 (2.54%)</td>
<td>86 (2.31%)</td>
</tr>
<tr>
<td>Class 2</td>
<td>368 (4.12%)</td>
<td>313 (4.11%)</td>
<td>144 (3.86%)</td>
</tr>
<tr>
<td>Class 3</td>
<td>571 (6.4%)</td>
<td>511 (6.7%)</td>
<td>235 (6.31%)</td>
</tr>
<tr>
<td>Class 4</td>
<td>91 (1.02%)</td>
<td>68 (0.89%)</td>
<td>42 (1.13%)</td>
</tr>
<tr>
<td>Class 5</td>
<td>516 (5.78%)</td>
<td>430 (5.64%)</td>
<td>186 (4.99%)</td>
</tr>
<tr>
<td>Class 6</td>
<td>441 (4.94%)</td>
<td>404 (5.3%)</td>
<td>160 (4.29%)</td>
</tr>
<tr>
<td>Class 7</td>
<td>397 (4.5%)</td>
<td>380 (5.1%)</td>
<td>170 (4.56%)</td>
</tr>
<tr>
<td>Class 8</td>
<td>2419 (27.11%)</td>
<td>2063 (27.06%)</td>
<td>900 (24.15%)</td>
</tr>
<tr>
<td>Class 9</td>
<td>161 (1.8%)</td>
<td>156 (2.05%)</td>
<td>73 (1.96%)</td>
</tr>
<tr>
<td>Class 10</td>
<td>2865 (32.11%)</td>
<td>2352 (30.85%)</td>
<td>1317 (35.34%)</td>
</tr>
<tr>
<td>Class 11</td>
<td>848 (9.5%)</td>
<td>744 (9.76%)</td>
<td>414 (11.11%)</td>
</tr>
</tbody>
</table>

Total: 8923 | 7624 | 3727

---

2https://github.com/celialicantias/vasijas/tree/iberianVox

---

5835
show the voxelized mesh and the fragments generated by the adapted Discrete Voronoi Chain algorithm.

Semilandmarks capturing the 2D pottery shape information of voxels following [Oguyar-Anguita et al., 2009].

each 3D model as a solid of revolution (see Fig. 2 for an example for each type of open and closed shape). To the left, we show the original 2D profile [Lucena et al., 2017]. Second, we show the spin calculated over the y-axis. Third, we show the voxelized mesh and the fragments generated by the adapted Discrete Voronoi Chain algorithm.

<table>
<thead>
<tr>
<th>Classes</th>
<th>15-20</th>
<th>Fragments (%)</th>
<th>20-30</th>
<th>30-100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE((%))</td>
<td>DSC((%))</td>
<td>Jacc.((%))</td>
<td>MSE((%))</td>
</tr>
<tr>
<td>Class 1</td>
<td>0.04±0.01</td>
<td>0.70±0.08</td>
<td>0.31±0.09</td>
<td>0.74±0.09</td>
</tr>
<tr>
<td>Class 2</td>
<td>0.06±0.02</td>
<td>0.72±0.09</td>
<td>0.32±0.11</td>
<td>0.74±0.09</td>
</tr>
<tr>
<td>Class 3</td>
<td>0.05±0.02</td>
<td>0.60±0.12</td>
<td>0.55±0.12</td>
<td>0.62±0.13</td>
</tr>
<tr>
<td>Class 4</td>
<td>0.07±0.03</td>
<td>0.47±0.13</td>
<td>0.68±0.11</td>
<td>0.55±0.13</td>
</tr>
<tr>
<td>Class 5</td>
<td>0.09±0.02</td>
<td>0.62±0.11</td>
<td>0.54±0.12</td>
<td>0.67±0.10</td>
</tr>
<tr>
<td>Class 6</td>
<td>0.09±0.02</td>
<td>0.62±0.11</td>
<td>0.53±0.12</td>
<td>0.71±0.10</td>
</tr>
<tr>
<td>Class 7</td>
<td>0.13±0.11</td>
<td>0.55±0.24</td>
<td>0.58±0.21</td>
<td>0.59±0.24</td>
</tr>
<tr>
<td>Class 8</td>
<td>0.03±0.02</td>
<td>0.75±0.09</td>
<td>0.39±0.11</td>
<td>0.77±0.09</td>
</tr>
<tr>
<td>Class 9</td>
<td>0.05±0.02</td>
<td>0.75±0.11</td>
<td>0.38±0.13</td>
<td>0.78±0.10</td>
</tr>
<tr>
<td>Class 10</td>
<td>0.01±0.01</td>
<td>0.82±0.06</td>
<td>0.28±0.09</td>
<td>0.84±0.06</td>
</tr>
<tr>
<td>Class 11</td>
<td>0.02±0.00</td>
<td>0.75±0.08</td>
<td>0.39±0.10</td>
<td>0.75±0.09</td>
</tr>
<tr>
<td>Average</td>
<td>0.06±0.03</td>
<td>0.70±0.07</td>
<td>0.44±0.08</td>
<td>0.70±0.07</td>
</tr>
</tbody>
</table>

Table 2: Performance metrics across different fragment sizes and ceramic types. The percentage is relative to each pottery’s size and shape.

Fracture ceramic simulations The voxelized pottery models were fragmented, simulating real fractured ceramics. We implemented a geometric fragmentation method based on the Discrete Voronoi Chain (DVC) algorithm [Velci et al., 2009].

The DVC algorithm is composed of several steps. First, it generates a random list of Voronoi region centers in the model, each corresponding to a fragment. Second, it assigns the voxel to a section by following a region-growing approach taking each center as a seed until all the voxels have been traversed. However, this procedure can assign voxels in the border between two regions to an incorrect one. Therefore, an additional distance check to the centers of each region is required to guarantee the correct assignment. The resulting number of fragments depends on the type of vessel (closed or open, see Fig. 1-B).

Our approach, like most geometric methods based on Voronoi [Raghavachary 2002; Domaradzki and Martyn, 2016], generates convex fragments. Thus, in order to generate more complex fragments, including concavities, we need to disable at runtime the distance check from the DVC algorithm (10% of the runs).

Lastly, in order to evaluate the performance under different relative sizes of fragments, we stratified our fragment generation into three main groups: 15-20%, 20-30%, and 30-100% of the initial ceramic model.

2D to 3D pottery Given that the existing dataset is based on wheel-made pottery techniques, we can assume small asymmetric perturbations in the ceramic. Thus, we can generate each 3D model as a solid of revolution (see Fig. 2 for an overview of the steps implemented in this work). First, we extract the shape information from the profile by means of 200 semilandmarks equally spaced in the contours. For the spin, we use 100 steps per revolution. This task converts the semilandmarks capturing the 2D pottery shape information into a mesh. Lastly, we converted the resulting mesh into a set of voxels following [Oguyar-Anguita et al., 2020] with sizes 32×32×32. The resulting dataset is composed of 1001 voxelized potteries, randomly divided into a training subset containing 704 ceramics (70%) and a test set of 297 artifacts (30% of the total dataset).
Figure 3: Qualitative examples for visual comparison of completed samples by our proposed method and the ground truth (real voxelized models). Each batch of examples shows the completion model started from different sizes of initial fragments (dark pink), and the completed ceramics (light pink) in different views.
4.2 Implementation Details
The first module of the Generator consists of an Encoder of 5 convolutional layers of kernel sizes $4 \times 4 \times 4$, stride 2, and padding $1 \times 1 \times 1$ (with $0 \times 0 \times 0$ for last one), with 3D batch normalization and LeakyReLU layers added in between the convolutional layers (except the last one). The intermediate result is a $128 \times \text{dimensional}$ vector. The last part of the Generator is a Decoder of 5 layers, all of them with a 3D transposed convolution with kernel sizes $4 \times 4 \times 4$, stride 2, and padding $1 \times 1 \times 1$. Unlike the encoder, the decoder has ReLU layers instead of LeakyRelu. The output is a binary array of $32 \times 32 \times 32$. The architecture of the $D$ is basically the same as that of the $G$ encoder except that it uses a sigmoid layer, previously reshaped the result of the last layer, for the binary classification (real/generated).

3D AE-GAN training IberianVoxel was trained for 100 epochs using Binary Cross Entropy as a loss function, at a learning rate $2 \times 10^{-3}$ for the generative 3D encoder-decoder network ($G$) and $2 \times 10^{-4}$ for the discriminator ($D$). To optimize the training process of all models, we scaled the artifacts to a uniform resolution space of $32 \times 32 \times 32$ voxels. We used ADAM optimization [Kingma and Ba, 2015] for both $G$ and $D$ with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Particularly, for the training of $D$, we used Label Smoothing [Salimans et al., 2016], the real set was represented with a random number between 0.7 and 1.1 and the generated set with 0.0 and 0.30.

4.3 Evaluation Metrics
The evaluation of the generated artifacts’ quality was done using standard completion and segmentation metrics such as the Dice similarity coefficient (DSC) [Sorensen, 1948], Jaccard distance, and Mean squared error (MSE). The DSC coefficient is used to calculate the similarity between the two examples [Feng et al., 2016; Navarro et al., 2022]. Given $A$ the generated set of voxels and $B$ the real ceramic, DSC is formulated:

$$DSC = \frac{2|A \cup B|}{|A| + |B|},$$

where $|A|$ and $|B|$ is the size in voxels of the pottery. The values of the formula are between 0 and 1, where 1 represents the maximum similarity, and 0 is the minimum.

Jaccard distance, like DSC, is a statistic used to estimate the dissimilarity of a dataset, but unlike DSC, Jaccard distance satisfies the triangle inequality. Jaccard distance is defined as:

$$JD(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|},$$

where $A$ is a three-dimensional binary array containing the generated pottery voxelized geometry, and $B$ is the real ceramic. The values of the formula are between 0 and 1 too, but unlike DSC, 1 represents the minimum similarity and 0 the maximum. Lastly, we evaluated our generated samples using MSE, which is a common quality estimator in 2D and 3D classification and reconstruction problems.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (d_i - f_i)^2,$$
fully Iberian Style. From this preliminary study, we can conclude that archaeologists judge that IberianVoxel generated a correct Iberian style from an initial fragment, and also consider that the reconstructed pottery is between Good and Very Good. At the end of the questionnaire, we included a comment section to enable unstructured feedback. The comments across evaluators agree on the need for better visualization tools, such as including a scale factor and improving the edge inspection of the models, as these are key factors while evaluating the Iberian Style.

5.1 Limitations

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Precision (↑)</th>
<th>Recall (↑)</th>
<th>F1-Score (↑)</th>
<th>Accuracy (↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.980</td>
<td>0.984</td>
<td>0.982</td>
<td>0.983</td>
</tr>
<tr>
<td>IberianVoxel (30%-100%)</td>
<td>0.984</td>
<td>0.987</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>IberianVoxel (20%-30%)</td>
<td><strong>0.988</strong></td>
<td><strong>0.990</strong></td>
<td><strong>0.989</strong></td>
<td><strong>0.989</strong></td>
</tr>
<tr>
<td>IberianVoxel (15%-20%)</td>
<td>0.984</td>
<td>0.987</td>
<td>0.986</td>
<td>0.986</td>
</tr>
</tbody>
</table>

Table 3: Performance metrics for open-close shapes classifier from test 3D data. All metrics were estimated with a test set of 121 open and 176 closed ceramics. We show results with original samples and ceramics reconstructed from initial fragments of different sizes (in %).

Irregularity in generated potteries In Fig. 4-B and Fig. 4-C we show two examples with irregular results in the completion. For example, incomplete reconstructions (see Fig. 4-B) or the mismatching alignment of the fragment with the proposed completion model (see Fig. 4-C). These problems could be reduced using post-processing methods, such as classic morphological operations.

6 Conclusion

Ceramics are one of the most common archaeological artifacts, usually used to investigate variations in style, materials employed, social activities, and manufacturing techniques. Nonetheless, most available pottery is fragmented. The completion and assembly of fragments is a time-consuming task, which requires the physical manipulation of the ceramic fragments, which exposes them to material decay. We present a 3D generative approach to process fragments automatically, and reconstruction analysis that can assist archaeologists in a more realistic 3D reassembly process. The IberianVoxel framework is generalizable to work with ceramic materials or other brittle objects (e.g., projectile points, ancient buildings, or anthropological remains).

We have evaluated the performance of IberianVoxel based on two complementary assessments. First, we used classical metrics to evaluate the generative process of voxelized samples across different fragment sizes (See Table 2 and Fig 3). Second, the results were validated via independent examination of archaeologists specialized in Iberian heritage. Results suggest that our approach can generate potteries that conform to the structure of Iberian ceramics and fulfill experts’ validation criteria for the three different sizes of initial fragments tested. Although the described approach has an encouraging performance, shown in this work for the completing fragments of Iberian wheel-made pottery, some limitations were found and discussed in the paper. Nonetheless, our proposed framework and the available voxelized dataset are the first steps toward the broader use of 3D generative networks for the completion of fragments.

Acknowledgments

This work was supported by the European Union through the Programa Operativo FEDER Andalucía 2014-2020 Research Project under Grant UJA-1381115, the Center for Advanced Studies in Information and Communication Technologies (CEATIC) and the Research University Institute for Iberian Archaeology of the University of Jaén.
Ethical Statement
The research reported in this paper does not incur in any ethical issues.

References


