

Uma Abordagem de Aprendizado Profundo para Estimar a Posição Relativa de Fragmentos Cerâmicos

Matheus F. C. Pinho¹, Guilherme L. A. Mota², Gilson A. O. P. Costa³
UERJ, Rio de Janeiro, RJ

Resumo. Cerâmica é um dos tipos mais comuns de vestígios humanos encontrados em contextos arqueológicos. A análise da cerâmica arqueológica tem grande potencial informativo, e sua reconstrução é uma tarefa demorada e repetitiva. Neste trabalho, investigamos uma abordagem baseada em aprendizado profundo para tornar esse processo mais eficiente, preciso e rápido. Dada uma nuvem de pontos de um fragmento em uma posição padronizada, o método proposto prevê a transformação geométrica que move o fragmento para sua posição relativa dentro do sistema de coordenadas do respectivo vaso. Duas redes neurais convolucionais profundas são treinadas para prever os parâmetros da transformação euclidiana 3D. A primeira rede se dedica a prever os momentos de translação, e a outra infere os parâmetros de rotação. Em aplicações práticas, uma vez que a forma de um vaso é identificada, as redes podem ser treinadas para prever os parâmetros da transformação alvo. Dado um modelo 3D do vaso de interesse, ele é virtualmente quebrado inúmeras vezes para a produção de dados sintéticos de treinamento, que consistem em um grande conjunto de fragmentos virtuais.

Palavras-chave. Patrimônio Cultural; Cerâmica; Inteligência Artificial; Redes Neurais Artificiais

1 Introdução

A importância da cerâmica na Arqueologia se dá pelo potencial informativo que ela agrega ao contexto arqueológico. A análise de cerâmica arqueológica é capaz de localizar o sítio arqueológico no tempo, bem como determinar sua distribuição espacial, evidenciando a tradição cultural daquela sociedade, os aspectos socioculturais envolvidos na utilização da cerâmica e as tecnologias empregadas em sua confecção.

Na restauração da cerâmica arqueológica deve-se, primeiramente, agrupar os fragmentos que possivelmente pertençam a um mesmo vaso. Em seguida, compara-se os fragmentos um a um, visando seu encaixe até que todo o vaso seja reconstruído. Todo esse processo é lento, repetitivo e dependente da expertise do profissional encarregado.

O presente trabalho propõe automatizar e acelerar o processo de reconstrução de cerâmica arqueológica a partir de seus fragmentos utilizando modelos 3D e redes neurais convolucionais. A abordagem envolve o treinamento de redes neurais artificiais para qualquer modelo específico de vaso. Mais especificamente, a nuvem de pontos 3D de um fragmento cerâmico é fornecida como entrada para o método. O resultado é uma matriz de transformação que posiciona o fragmento em sua posição esperada.

A arquitetura de rede neural aqui proposta, chamada PotNet [7], é inspirada na arquitetura PointNet [8]. Os dados de treinamento consistem em um conjunto de nuvens de pontos de fragmentos virtuais gerados pela quebra consecutiva de modelos 3D de vasos cerâmicos em ambiente virtual e são chamados de conjunto de dados sintéticos.

¹matheus.coelho@pos.ime.uerj.br

²guimota@ime.uerj.br

³gilson.costa@ime.uerj.br

2 Materiais

A Figura 1 mostra os três vasos cerâmicos reais usados nesse trabalho, denominados Vaso Grande (VG), Vaso Médio (VM) e Vaso Pequeno (VP). Os vasos foram largados de uma altura de aproximadamente 1,5 m, totalizando 57, 20 e 21 fragmentos para VG, VM e VP, respectivamente.

Os fragmentos resultantes foram digitalizados por um escâner 3D estereoscópico de luz estruturada, o escâner intraoral Virtuo VivoTM. Apesar de ter sido projetado para uso odontológico, ele se mostrou preciso o suficiente para gerar modelos 3D dos fragmentos.



Figura 1: Vasos VG, VM e VP antes da fragmentação. Fonte: Matheus Pinho (2022)



Figura 2: Modelos 3D dos vasos reconstruídos. Fonte: Matheus Pinho (2023)

Após a digitalização, os fragmentos correspondentes foram virtualmente encaixados usando o software de modelagem 3D Blender, até que todo o vaso fosse remontado. Após decimação, as malhas 3D dos fragmentos encaixados precisam ter seus vértices conectados, reproduzindo o procedimento de colagem de fragmentos no mundo real. Isso faz com que cada vaso seja restaurado em um único objeto sólido. Os vasos restaurados podem ser vistos na Figura 2 e são a entrada para a geração dos dados sintéticos utilizados durante o treinamento.

3 Modelagem do Problema

Para modelar o problema, assumimos que os vasos são simétricos em relação a y , permitindo infinitas posições ao redor desse eixo. Para simplificar, eliminamos o ângulo de rotação relativo a

y , definindo o chamado “sistema normalizado”⁴.

As nuvens normalizadas são produzidas rotacionando-se a nuvem de pontos de cada fragmento em torno do eixo y em sentido horário, movendo os centroides dos fragmentos para o plano yz — o plano que passa pelos eixos y e z . Os fragmentos em espaço normalizado representam as posições de referência (*ground truth*), que devem ser comparadas com as predições das redes.

Além disso, cada fragmento possui um “sistema canônico” absoluto. A nuvem canônica é a forma padronizada dos fragmentos na entrada do método, garantindo um modelo determinístico e único para cada fragmento. O centroide do fragmento coincide com a origem do sistema canônico, e o maior comprimento do paralelepípedo envoltório mínimo do fragmento alinha-se ao eixo x .

A conversão de uma nuvem de pontos para sua forma canônica ocorre em três etapas: cálculo do paralelepípedo envoltório mínimo, centralização do fragmento na origem e uso da Decomposição em Valores Singulares (SVD) [5] para projetá-lo no sistema de coordenadas canônico. A SVD consiste na fatoração de uma matriz retangular \mathbf{A} pelo produto de três outras matrizes: uma matriz ortogonal $m \times m$, \mathbf{U} ; uma matriz diagonal retangular $m \times n$, Σ ; e a transposta de uma matriz ortogonal $n \times n$, \mathbf{V} , conforme a forma:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \quad (1)$$

Neste trabalho, \mathbf{A} é uma matriz de ordem 8 por 3 contendo as coordenadas dos vértices do paralelepípedo envoltório mínimo com centroide na origem. A multiplicação \mathbf{AV} projeta os vértices do paralelepípedo na posição canônica.

A Figura 3 apresenta um fragmento arbitrário em suas três diferentes posições: posição original em relação ao vaso (vermelho), na posição normalizada (azul) e na posição canônica (verde).

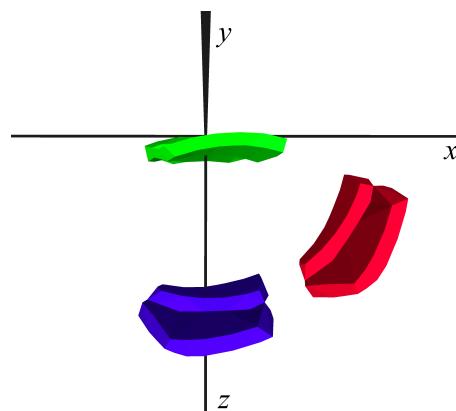


Figura 3: O mesmo fragmento em sua posição original, em relação ao sistema de coordenadas do vaso (em vermelho), na posição normalizada (em azul) e na posição canônica (em verde). Fonte: Matheus Pinho (2023)

Com as nuvens normalizada e canônica do fragmento, determina-se a transformação entre elas. A transformação da posição canônica para a normalizada define o *target* de treinamento \mathbf{y} , que será posteriormente comparado à predição $\hat{\mathbf{y}}$ da PotNet. O algoritmo de Kabsch [1, 6] calcula os componentes de rotação, minimizando o desvio quadrático médio entre os conjuntos de pontos.

⁴Adotamos um sistema de coordenadas de mão direita, com y apontando para cima.

4 Método

A Figura 4 apresenta uma visão geral do método proposto. Dada uma nuvem de pontos \mathbf{X} de um fragmento em seu sistema de coordenadas canônico (*Canonical Cloud*), o método produz \mathbf{X}' , uma versão dessa nuvem de pontos expressa em termos do sistema de coordenadas (normalizado) do vaso. Matematicamente, esse processo consiste na aplicação de uma transformação euclidiana 3D no espaço afim, na forma de um operador linear \mathbf{T} , onde $\mathbf{X}' = \mathbf{X} \cdot \mathbf{T}$. Aqui, \mathbf{X} e \mathbf{X}' são matrizes $p \times 4$, onde p representa o número de pontos em cada nuvem. A matriz \mathbf{T} é obtida ao combinar a matriz de rotação 3×3 e o vetor de translação 3×1 em uma única matriz de transformação 4×4 . Assim, cada coluna dessas matrizes denota um ponto 3D da nuvem, expresso em coordenadas homogêneas (obtidas pela concatenação do valor 1 após as respectivas coordenadas 3D).

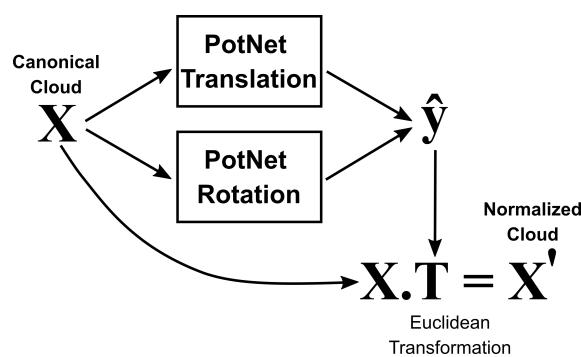


Figura 4: Visão geral do método proposto. Fonte: Guilherme Mota (2024)

A parte fundamental do método consiste em uma arquitetura de rede neural com duas ramiﬁcações, ambas prevendo um resultado com base exatamente na mesma entrada \mathbf{X} . Rotação e translação são treinadas de maneira independente. A rede de translação produz na saída os valores de translação nos eixos y e z , e a rede de rotação produz na saída um vetor 6D correspondente às duas primeiras colunas da matriz de rotação que leva o fragmento ao espaço de referência. A matriz de rotação 3×3 completa pode ser recuperada usando-se o processo de Gram-Schmidt [10]. Com o vetor de translação e a matriz de rotação, a matriz de transformação \mathbf{T} pode ser obtida.

As partes de rotação e translação da rede neural podem ser de qualquer arquitetura de rede neural que receba nuvens de pontos como entrada. No entanto, propomos que seja usado como *backbone* do método a arquitetura PotNet desenvolvida e apresentada neste trabalho. A PotNet é uma arquitetura de rede neural dedicada à regressão não linear, cujo design foi inspirado na PointNet [8]. De maneira geral, a PotNet é composta por várias camadas convolucionais unidimensionais, seguidas por uma camada de *max pooling* unidimensional global, com o objetivo de gerar uma série de características para a nuvem de pontos de entrada. Essas características passam por uma rede *multi-layer perceptron* (MLP) construída sobre as camadas convolucionais. A maioria das camadas da PotNet tem funções de ativação ReLU, sendo a única exceção a camada de saída, que tem ativação linear. Neste trabalho são usadas cinco camadas convolucionais, contendo respectivamente 64, 64, 64, 128 e 1024 filtros. As camadas MLP ocultas, que utilizam *batch normalization*, tem 512 e 256 neurônios, enquanto a camada de saída contém N neurônios. O número de neurônios na camada de saída pode ser 2, caso seja a rede de translação, ou 6, se for de rotação.

5 Procedimento Experimental

Desenvolvemos um método para produzir automaticamente fragmentos sintéticos usando a ferramenta *Cell Fracture*, um recurso nativo do software Blender. Essa ferramenta é baseada em uma implementação de Diagrama de Voronoi 3D para fragmentação de poliedros [2, 9]. Os modelos 3D dos vasos restaurados VG (ver Figura 2a), VM (Figura 2b) e VP (Figura 2c) são fragmentados milhares de vezes, e os fragmentos produzidos por esse procedimento compõem o conjunto de dados sintéticos de treinamento.

Os fragmentos sintéticos são normalizados e posteriormente colocados na posição canônica. Calcula-se as respectivas matrizes de transformação alvo e, em seguida, aplica-se o filtro de *Poisson-disk sampling* [4] do software Meshlab [3] a cada fragmento virtual (já em posição canônica), amostrando o fragmento em 1.024 pontos. As redes são então treinadas com as nuvens 1024×3 em posição canônica e as matrizes de transformação alvo.

6 Resultados e Discussão

A análise desta seção diz respeito à qualidade das previsões ao gerar as matrizes de transformação \mathbf{T} que levam as nuvens canônicas dos fragmentos para o espaço normalizado — a posição esperada.

Após a predição, cada ponto \mathbf{x}_i na nuvem canônica é multiplicado pela matriz \mathbf{T} correspondente, resultando em $\hat{\mathbf{x}}'_i$, um ponto previsto que deveria coincidir com \mathbf{x}'_i , seu homólogo na nuvem normalizada de referência. As posições preditas dos pontos são comparadas às respectivas posições normalizadas de referência para calcular os erros relativos. Como as nuvens de pontos predita e normalizada de um fragmento possuem exatamente os mesmos pontos, o erro exato entre essas duas nuvens pode ser determinado. Os erros são calculados considerando cada ponto e cada eixo de coordenadas. A partir das diferenças nas posições de cada ponto, foi computado o Erro Quadrático Médio (RMSE).

6.1 Resultados Para o Conjunto de Testes Sintético

A Tabela 1 apresenta os valores de RMSE e desvio padrão obtidos para o conjunto de testes sintético. O erro médio de posicionamento entre as nuvens predita e normalizada dos fragmentos sintéticos é inferior a 1,8 cm para todos os vasos, o que consideramos pequeno. O desvio padrão dos erros é inferior a 1 mm, indicando estabilidade dos erros nas predições.

Tabela 1: Métricas de avaliação da PotNet para todo o conjunto sintético de testes de VG, VM e VP (valores em metros).

	RMSE	STD
LV	0.025	0.001
MV	0.017	0.0004
SV	0.013	0.0002

6.2 Resultados Para o Conjunto de Testes Real

Aqui, o desempenho das redes treinadas com o conjunto sintético de fragmentos é avaliado com fragmentos reais, para demonstrar sua aplicabilidade à problemas do mundo real.

O conjunto de testes do mundo real é composto por 57, 20 e 21 fragmentos reais obtidos a partir da quebra física dos vasos reais VG, VM e VP, respectivamente. Cada fragmento real é

escaneado em 3D, decimado e amostrado em 1.024 pontos pelo filtro Poisson-disk sampling do Meshlab, resultando em uma representação em forma de nuvem de pontos. Cada nuvem de pontos foi posicionada em sua posição canônica e utilizada como entrada para as redes treinadas.

Tabela 2: Métricas de avaliação da PotNet para todo o conjunto de testes real de VG, VM e VP (valores em metros).

	RMSE	STD
LV	0.034	0.0006
MV	0.019	0.0002
SV	0.023	0.0003

A Tabela 2 apresenta os erros obtidos para os fragmentos reais de VG, VM e VP. Acreditamos que os resultados ligeiramente piores para os fragmentos reais ocorrem pois estes tem superfícies mais irregulares se comparados àqueles gerados por computador, resultando em uma distribuição de vértices mais complexa nos fragmentos reais, o que torna o aprendizado mais desafiador para a rede.

Para melhor visualização, fazemos a rotação das nuvens preditas ao redor do eixo y no sentido anti-horário, trazendo o fragmento do espaço normalizado para o espaço do vaso. Dessa forma, é possível observar que, de maneira geral, os resultados qualitativos apresentados na Figura 5 mostram que os modelos de rotação e translação aproximam-se dos resultados esperados em um processo de reconstrução de vasos cerâmicos a partir de seus fragmentos.

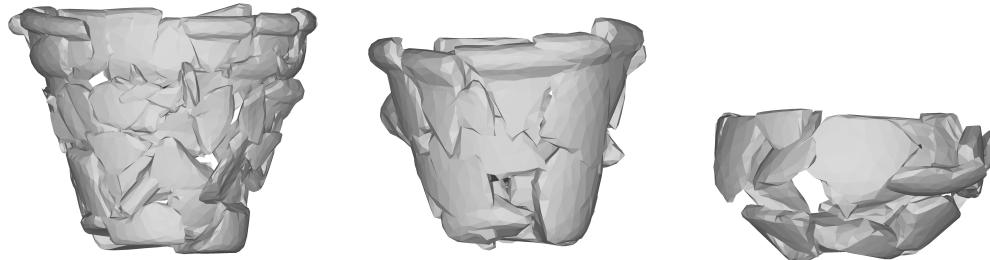


Figura 5: Fragmentos do conjunto de testes do mundo real colocados nas posições preditas e rotacionados em sentido anti-horário pelo inverso da matriz R_φ . Fonte: Matheus Pinho (2024)

7 Conclusão

No presente trabalho, propomos uma abordagem baseada em aprendizado profundo para auxiliar na reconstrução de vasos cerâmicos a partir de fragmentos. O método infere a transformação geométrica que move um modelo 3D de um fragmento da posição canônica para sua posição relativa no vaso.

Avaliamos o método com três vasos reais, quebrados fisicamente em fragmentos, escaneados digitalmente e convertidos em nuvens de pontos 3D. Modelos 3D reconstruídos foram quebrados digitalmente para gerar fragmentos virtuais, usados no treinamento da arquitetura PotNet. Dois modelos da PotNet foram treinados para cada vaso: um para prever a translação e outro para a rotação dos fragmentos.

Nos experimentos, as redes previram corretamente as transformações, com erros médios de poucos centímetros e desvios padrão milimétricos. A avaliação qualitativa também foi satisfatória.

Para trabalhos futuros, buscamos aprimorar a precisão da PotNet, estender o método para classificar fragmentos por tipo de vaso e explorar texturas e padrões decorativos para auxiliar na reconstrução.

Agradecimentos

Gostaríamos de expressar nossa gratidão à Neodent e ao Grupo Straumann por fornecerem o scanner intraoral Virtuo Vivo™, que possibilitou o desenvolvimento deste trabalho. Este trabalho foi parcialmente apoiado pela Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pela Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Referências

- [1] K. S. Arun, T. S. Huang e S. D. Blostein. “Least-squares fitting of two 3-D point sets”. Em: **IEEE Transactions on pattern analysis and machine intelligence** PAMI-9.5 (1987), pp. 698–700.
- [2] M. de Berg, O. Cheong, M. van Kreveld e M. Overmars. **Computational geometry algorithms and applications**. Springer, 2008.
- [3] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli e G. Ranzuglia. “MeshLab: an Open-Source Mesh Processing Tool”. Em: **Eurographics Italian Chapter Conference**. Ed. por Vittorio Scarano, Rosario De Chiara e Ugo Erra. The Eurographics Association, 2008. ISBN: 978-3-905673-68-5. DOI: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129–136.
- [4] M. Corsini, P. Cignoni e R. Scopigno. “Efficient and flexible sampling with blue noise properties of triangular meshes”. Em: **IEEE transactions on visualization and computer graphics** 18.6 (2012), pp. 914–924.
- [5] C. F. van Loan. “Generalizing the singular value decomposition”. Em: **SIAM Journal on numerical Analysis** 13.1 (1976), pp. 76–83.
- [6] S. Malischewski, H. Schumann e D. Hoffmann. **Kabsch Algorithm**. 2016. URL: <https://biomolecularstructures.readthedocs.io/en/latest/kabsch/>.
- [7] M. F. C. Pinho. “Aprendizagem Profunda Aplicada à Reconstrução de Artefatos Cerâmicos”. Mestrado Acadêmico em Ciências Computacionais. Programa de Pós-Graduação em Ciências Computacionais, Universidade do Estado do Rio de Janeiro, 2023, p. 104.
- [8] C. R. Qi, H. Su, K. Mo e L. J. Guibas. “Pointnet: Deep learning on point sets for 3d classification and segmentation”. Em: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2017, pp. 652–660.
- [9] J. Ronnegren. **Real Time Mesh Fracturing Using 2D Voronoi Diagrams**. 2020.
- [10] Y. Zhou, C. Barnes, J. Lu, J. Yang e H. Li. “On the continuity of rotation representations in neural networks”. Em: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. 2019, pp. 5745–5753.