

須恵器3次元計測データによる分類解析モデルの検討

市川 健太
Kenta Ichikawa

株式会社 BSN アイネット
ichikawa@bsnnet.co.jp, <https://www.bsnnet.co.jp>

板垣 正敏
Masatoshi Itagaki

板垣正敏中小企業診断士事務所
itagakim@gmail.com

山本 亮
Ryo Yamamoto

東京国立博物館
yamamoto-r62@nich.go.jp, <http://www.tnm.jp/>

藤田 晴啓
Haruhiro Fujita

新潟国際情報大学
fujita@nuis.ac.jp, <https://www.nuis.ac.jp/>

keywords: deep learning, 3d convolutional neural network (3D-CNN)

1. 序 論

本研究ではこれまで一部の専門家のみが行ってきた考古土器資料の分類を深層学習を用いた手法に転換し、専門家に頼らない土器分類が可能となるような手法の開発を目的とする。使用する土器は古墳時代・6世紀の須恵器であり、従来の形態・色・質感を表すカラー画像(2Dデータ)に加え、形状(3Dデータ)および資料複数箇所の法量(1Dデータ)を同時に入力、年代推定および型式分類を同時に多重出力できる革新的な深層学習モデルの開発を目的とする。

高精度光学スキャナにより600点以上の資料を蓋と身の個別にスキャン・3次元計測する。型式および年代分類に重要な法量を1Dベクトル値として入力、3次元計測データを多方向からキャプチャーしたRGB画像2Dデータは、それぞれ独立した画像分類器にて処理を行う。以上1D, 2D, 3D並列処理後に出力された数値データは、全結合層における統合処理を経て、型式分類器および年代推定器にて分類を行う。機械学習を援用した考古遺物の分析方法の構築は、須恵器以外の土器にも拡張でき、個人の経験に頼ってきた状況を打開するのみならず、考古学をより実証的な学問へと底上げし、広く埋蔵文化財行政の現場にも普及することで今後の考古学データの質の向上へとつながる成果となることが期待される。

2. 先 行 事 例

考古学では近年基礎資料である遺跡の調査報告書を中心にデジタル化が急激に進展し、きわめて膨大な量のビッグデータが存在するが、現状では十分な活用がなされて

いるとは言い難い。もともと紙ベースでの活用が想定されてきたこともあり、依然デジタル化に即応した資料の提示はおろか、その検討すらなされていないことが要因である。

深層学習は機械学習の一つで、これまで人間がさまざまな情報から行ってきた判断・分類等をコンピュータに学習させ実行させる、人工知能の急速な発展を支える基幹技術である。昨今、飛躍的に開発が進化した自動運転など、さまざまな分野での実用化が進められているが、考古関連分野への応用はごく最近みられはじめたばかりであり、研究の萌芽的な段階にある。

近年、日本情報考古学会などで試みられているのは調査報告書に多く掲載される2次元の実測図、あるいは断面図を利用してフーリエ解析を行う方法である[館内20]。実測図は作成にあたり実測者である人間の主観的認識が反映されるものであり、バイアスがかかるとともに深層学習利用時でも誤差が出てしまう。一方で本研究では純粋に対象資料の計測データのみを用い、主観を排した方法をとる。また当該文献では3Dデータを2Dデータに置き換えている点で、3Dデータをそのまま利用する本研究とは異なる。

また、なかには3Dデータを取得したのち断面図を作成して同様の分析を行うものもみられるが[井上20]、教師なしでクラスタリング分析を行いどのような分類結果が導かれるか検証するものであって、本研究のように正解データに基づく年代や生産地の推定を目指す目的とは異なる。

深層学習を用いた3次元物体認識では、入力データの形式により、RGBDベース[Lai 11]、点群データを利用する方法[Qi 17]、voxelデータを利用する手法[Wu 15]、

Multi View ベース [Su 15] などの手法が知られている。2次元データによる画像分類を行う場合、深層学習で利用するデータ数が少ない場合、ImageNet [Deng 09], COCO [Lin 14] などの大規模なデータセットで事前学習されたモデルを利用して、転移学習やファインチューニングを行い精度を確保する手法が考えられるが、3次元ではそのような大規模データセットは存在せず、転移学習やファインチューニングが利用できない。

2.1 voxel ベース

voxel ベースとは、3次元データを voxel と呼ばれる格子上のデータに変換したデータを入力として、3次元での畳み込みとプーリングを行う畳み込みニューラルネットワーク (CNN) を利用する手法である。CNN は2次元の画像分類などのタスクで広く利用されるメジャーな手法であり、3次元畳み込みニューラルネットワーク (3D-CNN) はこれを3次元データに拡張したものである。

3次元の voxel データでは、一辺の解像度を大きくすると3乗のオーダーでデータサイズが大きくなり、深層学習時の計算量が膨大になってしまうために大きな解像度が利用できないという問題がある。[Wu 15] では色情報なしの一辺が30の大きさの voxel を利用している。

また、3次元データの回転にどのように対応するかという課題もある。

2.2 Multi View ベース

Multi View ベースは、3次元データを元に複数の方向から画像を作成し、それぞれの画像を2次元のCNNを利用して抽出された特徴量から物体を分類する手法である。

[Su 15] では3次元データを30度ずつ向きを変えながら12枚の画像を生成。12枚の画像をそれぞれCNNを利用し特徴量を抽出し、回転普遍性の保持のための View Pooling 層を利用し、3次元物体の分類を行っている。

2.3 点群ベース

点群ベースとは、3次元のデータからサンプリングされた大量の点 (点群) を入力として、1次元の畳み込みを繰り返し行い局所的な特徴を抽出し、後続の層でプーリングを用いて大域的な特徴を捉える手法である。

3. 手法

本研究では、Creaform 社製の3Dスキャナーを用いて、東京国立博物館に収蔵されている須恵器杯の3次元ベクトルデータとテキストチャータータを取得した。voxelデータの作成と画像データの作成にはそれぞれ、Pythonの3次元データを扱うためのライブラリである trimesh (ver 3.9.26) と pyvista (ver 0.31.3) を利用した。

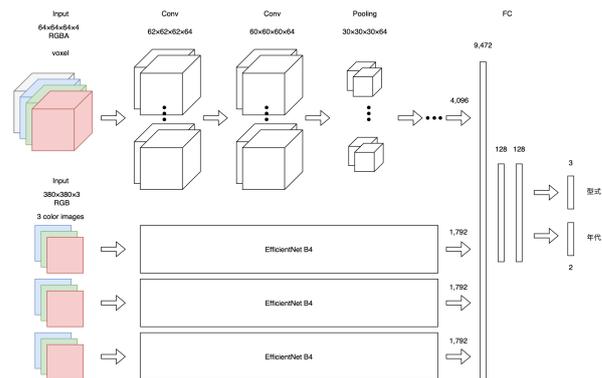


図1 3次元畳み込みニューラルネットワーク (3D-CNN) を利用したマルチヘッド・マルチタスクモデル

3.1 入力層

我々が提案するモデルは、voxel ベースと Multi View ベースを組み合わせた、複数の入力を持つマルチヘッドモデルである。(図1) voxel ベースの手法を利用することで、3次元の独特な形状と色の特徴をとらえ、データ数の不足を2次元画像を取得し、事前学習済みの EfficientNet で特徴量を抽出することでカバーする。

1つ目の入力データとしては色情報 (RGBA チャンネル) を持つ voxel データを利用する。今回利用する voxel データのサイズは $64 \times 64 \times 64$ とした。(図2) voxel の中でデータの存在しない部分については単純に白 (255,255,255) や黒 (0,0,0) で表現する案も考えられるが、今回はアルファチャンネルを利用してデータの有無を表現することとした。

また、2つ目・3つ目・4つ目の入力データとしては、それぞれ3次元計測データから生成した、上面、側面、下面の画像データを利用する。EfficientNet B4 モデル [Tan 19] の入力層に合わせて、画像のサイズは 380×380 。色情報はアルファチャンネルを持たず、RGBの3チャンネルを持つ。データが存在しない部分は白 (255,255,255) で表現する。(図3)

3.2 中間層：3D-CNN

voxel データの特徴を抽出する中間層には、3D-CNN を利用する。3D-CNN は voxel データを入力とし、3次元の畳み込みと、プーリングを行う CNN である。畳み込み層では voxel データの3次元方向のパターンに対する各々のフィルタの反応が出力され、プーリング層では3次元方向の位置のずれが吸収される。

今回、畳み込み層では、カーネルサイズ (3,3,3)、ストライド (1,1,1) を利用し、64種類のフィルタを準備した。プーリング層では Max Pooling を利用した。3D-CNN 部では最終的に voxel から得られた特徴量を4,096次元のベクトルデータとして出力する。

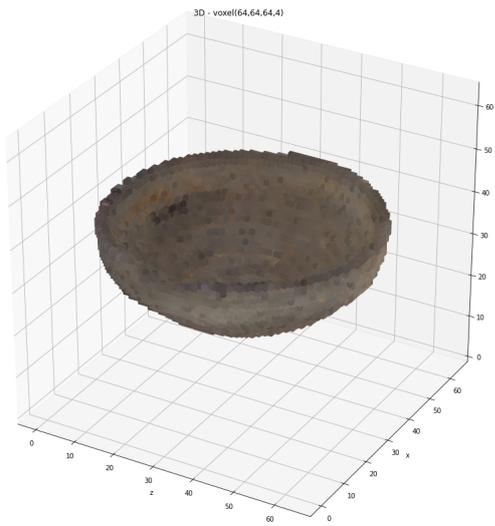


図2 3次元データの voxel 表現 (64 × 64 × 64)



図3 3D 計測データから生成した、上面、側面、下面のカラー画像データ (透視投影で生成)

3.3 中間層：EfficientNet

画像データの特徴を抽出する中間層には EfficientNet を利用する。EfficientNet は [Tan 19] で提案された画像分類の深層学習モデルである。モデルのハイパーパラメータの最適化により従来のモデルと比較してモデルのサイズが小さいにも関わらず、分類精度が非常に高いことで注目を浴びた。現在でも転移学習にも広く利用されている。今回は、モデルのサイズと精度のバランスのよい EfficientNet B4 を利用して、画像データからの特徴抽出を行う。1つの EfficientNet 部では最終的に 1 枚の画像から得られた特徴量を 1,792 次元のベクトルデータとして出力する。本モデルでは 3 つの EfficientNet 部を持つため、合計 5,376 次元のベクトルデータとして特徴が抽出される。

3.4 出力層

出力層では、中間層で voxel から特徴抽出した 4,096 次元のベクトルデータと、3 種類の画像から特徴抽出した 5,376 次元のベクトルデータの合計 9,472 次元のベクトルデータを利用して、須恵器杯の型式と年代を分類する。

出力層では全結合層と Dropout 層を繰り返し利用し、型式は 3 クラス、年代は 2 クラスの分類を行う。

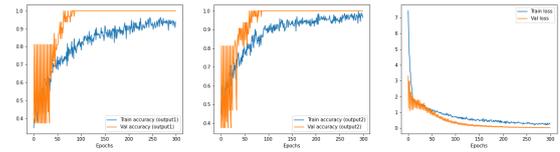


図4 学習時の学習用・検証用データの正解率と損失の遷移

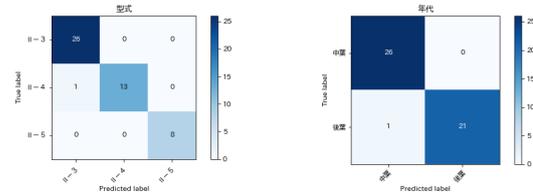


図5 テスト用データでの混同行列 (300 epoch)

4. 実験結果

本研究では東京国立博物館に収蔵される須恵器杯を対象としており、2021 年 10 月までに計測できた 24 個の須恵器杯の 3 次元計測データを用いてモデル開発とその分類精度の実験を実施した。今後、600 個程度の計測を行い改めて学習を行うことでモデルの精度を向上させる予定であるが、現時点では取得できた 3 次元計測データの数が少ないことを断っておく。

また、3D-CNN, EfficientNet はともに、物体の移動に対しての移動普遍性はあるが、回転に対しては弱いという特徴を持つ。3 次元計測データの回転を行うことで、学習用データの拡張を行っている。具体的には、須恵器杯の中心を通る垂直方向の軸を回転軸として 30 度ずつ回転させたデータを生成することで、須恵器杯 1 個体から 12 個のデータを生成した。

なお、法量データの取得はできていないため、法量データを利用する部分の実装を省いている。

須恵器杯の 3 次元計測データを回転させて生成した 1 個体辺り 12 個のデータを学習用 (train)・検証用 (validation)・テスト用 (test) のデータとして 8:2:2 の割合で分割して学習と評価を行った。

学習用に 192 個のデータ、検証用に 48 個のデータを利用し、300 epoch 学習を行った場合の epoch 毎の正解率と損失の遷移を図 4 に示す。今回利用した個体には型式・年代の数に偏りがあり、型式では「II-3」クラス、年代では「中葉」クラスが多くの割合を占めているが、テスト用データから作成した混同行列 (図 5) と表 1、表 2 より、高い精度での分類を行うことができていくことがわかる。

表1 型式分類の適合率と再現率

クラス	適合率	再現率
II - 3	0.9630	1.0000
II - 4	1.0000	0.9286
II - 5	1.0000	1.0000

表2 年代分類の適合率と再現率

クラス	適合率	再現率
中葉	0.9630	1.0000
後葉	1.0000	0.9545

5. 結 論

本研究では、須恵器杯の型式・年代推定のために、3次元計測データから作成される複数の入力データを利用した深層学習モデルの提案を行った。本論文では、回転させているとはいえ、学習用データとテスト用データに同一個体からのデータを利用しているため、今回の実験結果ではモデルの汎化性能について正確に議論することはできない。今後、須恵器杯の3次元計測を進め、個体の数を十分に増やし、学習用データとテスト用データに別々の個体を利用して再度、学習と評価を行う予定である。

◇ 参 考 文 献 ◇

- [Deng 09] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database (2009)
- [Lai 11] Lai, K., Bo, L., Ren, X., and Fox, D.: Sparse Distance Learning for Object Recognition Combining RGB and Depth Information. (2011)
- [Lin 14] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P.: Microsoft COCO: Common Objects in Context (2014)
- [Qi 17] Qi, C. R., Su, H., Mo, K., and Guibas, L. J.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. (2017)
- [Su 15] Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E.: Multi-view Convolutional Neural Networks for 3D Shape Recognition. (2015)
- [Tan 19] Tan, M. and Le, Q. V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. (2019)
- [Wu 15] Wu, Z., Song, S., Khosla, A., Fisher Yu, L. Z., Tang, X., and Xiao, J.: 3D ShapeNets: A Deep Representation for Volumetric Shape Modeling. (2015)
- [井上 20] 井上隼多, 堀涼, 川西康友, 村瀬洋, 梶原義実: 人工知能による機械学習を用いた須恵器資料の断面形状分析, 日本情報考古学会講演論文集, Vol. 23, (2020)
- [館内 20] 館内魁生: 実測図を利用した形態の定量的解析の有効性, 日本情報考古学会講演論文集, Vol. 23, (2020)