

マルチヘッド・マルチタスク 3D-2D-CNN モデルの開発

藤田 晴啓
Haruhiro Fujita

新潟国際情報大学
fujita@nuis.ac.jp

市川 賢太
Kenta Ichikawa

株式会社 BSN アイネット
ichikawa@bsnnet.co.jp

板垣 正敏
Itagaki Masatoshi

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

keywords: multi-head, multi-task, 2D-3D-CNN parallel model, Sueki

1. 本研究の目的

機械学習によるデータ解析, 特に分類問題では多数の写真画像すなわち 2D データをサンプルの型式あるいは年代ラベルと合わせて 2D 畳込み学習ニューラルネットワーク (CNN) に学習させ, 未分類の画像データの型式や年代を推測する方法が多い。この場合分類精度を上げるには膨大な数量の学習データによる訓練が必要であるが, この初期状態からの多量の学習を回避するため, 多くの場合は, 2D データにて学習済みのモデルを利用する転移学習が一般的である。一方, 膨大な量の 3D データを既学習した 3D-CNN モデルは皆無であり, 最近利用が一般化してきた光学スキャナーにより取得される 3D-RGB データも含め 3D データを解析する上での大きなボトルネックとなっている。そこで初期状態からしかも少ない訓練データ数で学習を行う 3D-CNN の低い特徴抽出を補完する目的で複数の既学習 2D-CNN モデルを 3 つ並列に 3D-CNN モデル併せて配置する 3x2D-CNN+3D-CNN を構築し, 考古資料である古墳時代の須恵器の複数分類ラベルの出力を可能とするマルチヘッド・マルチタスクモデルによる学習・推論を試行したのでその結果を報告する。

2. 考古資料と分類ラベル

分析対象は 6 世紀の須恵器の坏とする (図 1)。6 世紀の須恵器の坏を取り上げる利点としては, 1) 形態の個体差が小さく, 変化を追いやすいこと, 2) 身と蓋が対になって出土することが多いため, それぞれの結果を併せて検討することで成果の検証がしやすいこと, 3) 古墳から良

好な状態で出土する事例が多く前後の時期に比して多量の良好な資料が存在すること, 4) 生産地がある程度限られ適度に地域別の分析を行い比較しやすいこと, 5) 法量が小さいため 3D データも容量が小さいこと, が挙げられる。今回のプロトタイプモデル実験で使用したのは東京国立博物館所蔵の須恵器蓋が 16 サンプル, 身が 8 サンプルである。これらの資料のうち, 陶邑古窯址群の多くと東京国立博物館の全ての資料については, 代表的な先行研究者である中村浩氏ひとりにより従来の人文学的方法により型式, 年代の評価がなされている。そのため, 正答となる分類ラベルは中村氏のものを採用する [中村 81]。

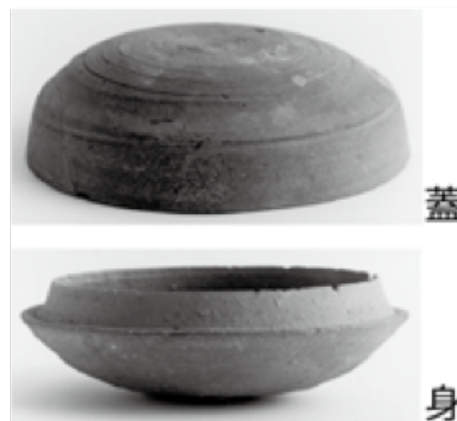


図 1 須恵器の蓋と身

3. 試供データ前処理とデータ拡張

モデル構築にあたり詳しい参考とした先行モデル、スキャナーデータを Voxel3D-RGBA に変換し、3 方向（真上、横、真下）から投影した 2D データを出力する詳細に関しては市川ら (2021)[市川健太 21] を参照いただきたい。

3D スキャンしたのは、須恵器蓋 16 資料および身 8 資料である。型式は II-3, II-4, II-5 の 3 クラス、年代は中葉か後葉の 2 クラスである。計測データファイナライズ時に指定した蓋および身のそれぞれ Y 軸 0°（垂直方向）を中心に obj から 30° ずつ回転させ Voxel データ (64x64x64x4) に変換、さらに obj から Z 軸 0°（正面）、90°（真上）、-90°（真下）方向からの透過投影画像 (380x380x3) を作成する（図 3）。すなわち 1 サンプルごとに 12 倍のデータ拡張を行い、24 サンプルから 288 データを作成する [Su 15]。このうち 192 データは分類学習の訓練データ、48 データは検証用、48 データは推論テスト用とした。

4. モデル構造

この 3 方向の 380x380,x3 チャンネルの 2D データはリスケール、正規化、ゼロパディング、2 次元畳込み、バッチ正規化、活性化関数出力、Depthwise2 次元畳込み、バッチ正規化、活性化関数出力を経た後、以下の一連の処理を 30 回繰り返す。2 次元数値を全て平均する GlobalAverage-Pooling2D、ベクトルを 3 次元にリシェイプ、2 次元畳込みを 2 回、行列乗算、2 次元畳込み、バッチ正規化、活性化関数出力。以上の処理の 31 回目に 3D-RGB データすなわち 64x64x64x4(RGB+透明度) チャンネルの Voxel データが入力され、3 次元畳込みを 2 回、3 次元マックスプーリング、バッチ正規化のセットを 2 回くりかえし、最後は 4x4x4x64 データを展開する。これと同時に先述 3 つの 2D ネットワークでは 2 次元数値を全て合計する処理を行い、これら 4 つのネットワークの数値を結合、全結合層とドロップアウトを 2 回くりかえし、最後は 2 つの異なる分類ラベルを出力する構造となる。この 2D ネットワークが 31 回目の処理を行うのと同時に 3D ネットワークが処理を開始する構造図を示した（図 4）。

5. 結果

市川ら (2021) で発表された適合率と再現率を参照いただきたい。図の混同行列の列方向は推測されたひとつのクラスに対する行方向の正解クラスごとの頻度である。型式分類の II-3, II-4, II-5 クラスの適合率はそれぞれ、0.96, 1.00, 1.00, 再現率は 1.00, 0.93, 1.00 となった。年代分類の中葉および後葉クラスの適合率はそれぞれ 0.96, 1.00, 再現率は 1.00, 0.95 と高い精度で分類された結果となった。

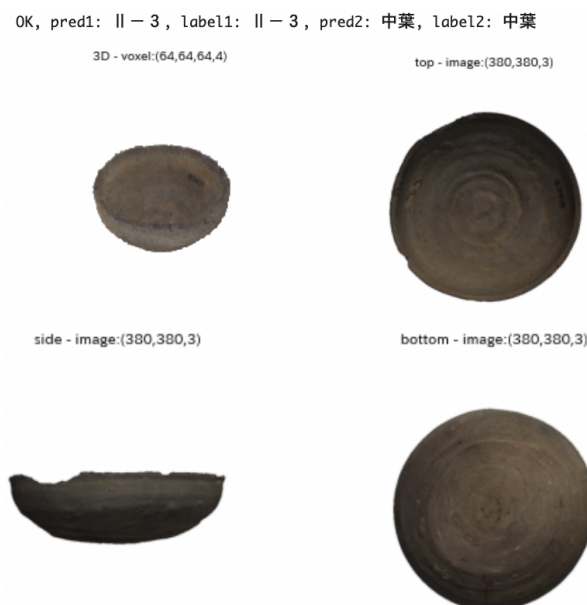


図 2 3D+3x2D 入力データ

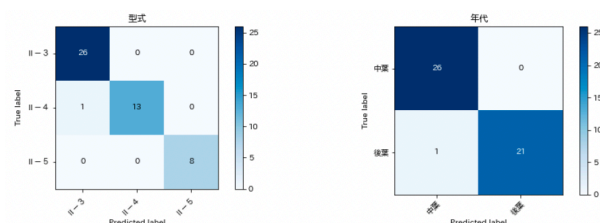


図 3 型式推論の混同行列

6. 考 察

今回はスキャナー計測データの入力データである obj ファイルから Voxel データ作成, Muliti View Base による 3 方向からの透視投影 2 D データの作成, これらの 4 種類のデータを入力して 3D-CNN と 3x2D-CNN の並列構造で, 2 種類の分類出力が正常に機能するかを検証することが主たる目的であったので, スキャンしたサンプル数も 24 と少数であった。しかしながら, データ拡張をとりいれることにより, データ数を訓練用, 検証用, テスト用に確保し, 型式および年代の分類精度も高い結果となったことから, モデルの統合的な性能は評価できると考えられる。今後, 3D-CNN の性能を上げていくための検討が必要である。今後検討をおこなう資料は各地の窯から出土した年代基準となる資料（大阪府陶邑古窯址群、同吹田窯跡群、愛知県猿投古窯跡群等）に東京国立博物館が所蔵する 6 世紀の完形・略完形の須恵器坏 600 点程度である。

◇ 参 考 文 献 ◇

- [Su 15] Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E.: Multi-view Convolutional Neural Networks for 3D Shape Recognition, <https://arxiv.org/pdf/1505.00880.pdf> (2015)
- [市川健太 21] 市川健太 板垣正敏, 山本亮, 藤田晴啓: 須恵器 3D 計測データによる分類解析モデルの検討, 考古文化財ディープラーニング研究会報告書, 新潟国際情報大学機関リポジトリ (2021)
- [中村 81] 中村 浩: 出土須恵器の編年的考察, 柏書房 (1981)



図4 3x2D+3D マルチインプット・マルチタスクモデル (部分)