

ボクセルデータに基づきワークの表面形状を識別するための 3D CNN の設計と評価

廣野 聖和† 永田 寅臣† 金沖 巧望† 芝 優一郎† 大塚 章正† 渡辺 桂吾‡

† 山口東京理科大学 ‡ 岡山大学

E-mail: nagata@rs.socu.ac.jp

1 緒言

筆者らはこれまでに3次元畳み込みニューラルネットワーク 3D CNN を応用した金属の火花試験の自動評価システムの開発を行い、3D CNN の有用性や解決すべき課題について報告してきた [1, 2]。本研究では、輪郭形状や色が同一でありながら表面形状や表面粗さが異なるようなワークを高精度に分類できるシステムについて検討している。これまでも生産ラインを流れる工業製品の写真をもとに画像処理や CNN の技術を応用して自動的に欠陥検出を行おうとするシステムが多く提案されている [3, 4, 5]。しかしながら、ワークの輪郭形状が同じ場合、カメラで撮影された画像情報のみで CNN モデルを構築するだけでは、ワークの表面形状の違いを識別することは困難である。また、高性能な欠陥検出機能を持つ CNN モデルの学習には膨大な学習データが必要であるが、2D カメラを用いてワークの形状情報を RGB 画像にて取得する際、訓練時とテスト時で照明条件が変化しないように入念に環境を構築していく必要がある。長期的にデータを蓄積していく場合、照明環境が一定となるよう、その都度確認と調整を行うことは非常に煩わしい作業となる。一方、3D カメラは赤外線を照射しワークとの距離情報を取得できるため、照明環境の変化に影響されずに訓練用やテスト用のデータを蓄積できる点で有効である。

本研究では、3D カメラで観測したワークの点群データ (PCD) から得られる深度情報をもとにボクセルデータを生成し、そのようなワークの表面形状をもとに分類を可能とするシステムの検討を行ったので報告する。

2 VoxNet

本研究では、3D カメラから取得した3次元点群データを扱うため、3次元データに対応したネットワークを構築する必要がある。そこで Maturana らの提案する3次元データの学習において有効な VoxNet というモデルを採用した [6]。図1に VoxNet の概要図を示す。このモデルでは、PCD をボクセル化し、3D CNN の学習データとして使用する。ボクセルデータは、各ボクセル

内に3次元データ(点)が存在する場合は1を、存在しない場合は0の情報を持つため、0と1のバイナリコードのみで3次元空間内のオブジェクトの情報が保持され、対象とするワークの形状や構造をより正確に捉えることができる。このため、VoxNet は通常の3D CNN に比べて計算コストとメモリコストが低いという特徴を持つ。3D CNN の構造を持つ VoxNet の学習では、ボクセル化された形状データを入力とし、3次元方向にボクセル単位の畳み込みとプーリングを行う。

本研究では、基本的な評価のために図2のように8カテゴリの3Dモデルを3次元CAD「Creo」でそれぞれ10種類ずつ計80のモデルを設計し、3Dプリンタで造形した。すなわち、造形した実モデルの数も8カテゴリ×10種類=80モデルとなる。次章では、これら80の実モデルの点群データ(PCD)を計測し、点群データが構成する3次元空間内を $32 \times 32 \times 32$ の解像度でボクセル化し、そのボクセルデータを用いてVoxNetの学習と汎化性能の評価を行う。

3 訓練用ボクセルデータの生成

3.1 ワークモデルの作成と3Dカメラによる点群データ(PCD)の取得

2Dカメラでは、モデルを真上から撮影すると、照明などの撮影条件によっては緩やかな傾斜や小さな起伏の特徴などを画像の中に再現性よく安定的に収めることは困難である。このため、3Dカメラで撮影した点群データ(PCD)を用いることとした。図3には本研究で分類対象とする同一色を持つ8カテゴリの3次元モデルの例を示す。それぞれの輪郭形状は同じ矩形であるがモデル上面には異なる表面形状を持っている。なお、構築する3D CNNの汎化性を高めるために、1つのカテゴリに対してその特徴を持つ異なる実モデルを10個ずつ作成した。

本実験では、図4のようにモデルの真上に3Dカメラを設置し、8カテゴリそれぞれ10個のモデルについて点群データを1ファイル(.pcd)ずつ測定した。3DカメラにはIntel RealSense D435を使用した。なお、モデルが置かれている床の情報を除去するためにZ方向

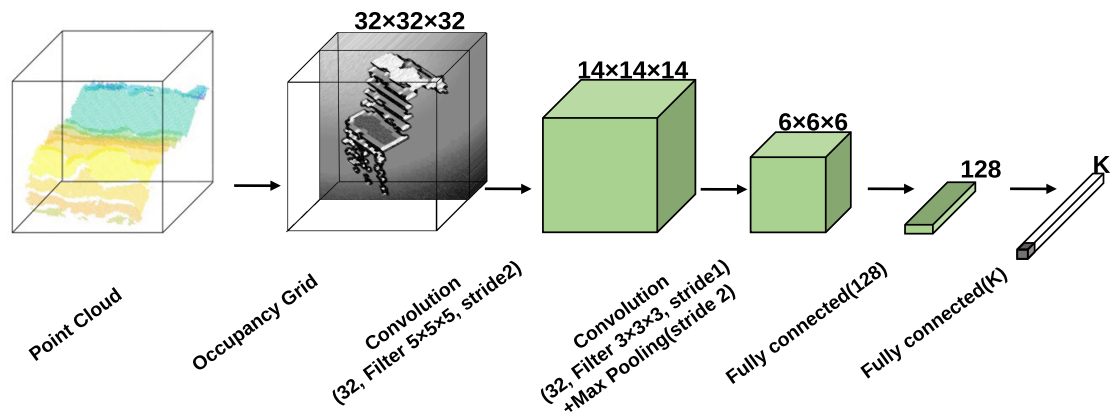


図 1 Overview of VoxNet architecture [6].

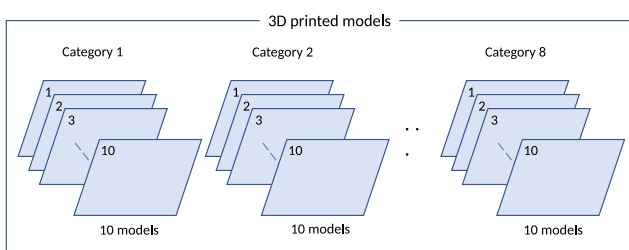


図 2 8 categories \times 10 models = 80 models are 3D-printed.

のスキャン範囲を調整している．取得した点群データに対して，指定した角度で Z 軸周りに回転させるようなデータ拡張法について検討し，回転角度を 1 度に設定することで 1 つの点群データから 359 の拡張ファイル (.pcd) を生成させ，結果的に， $80 \times 360 = 28,800$ のボクセルデータを準備した．

3.2 ボクセルデータへの変換

RealSense D435 で取得できる深度情報 (PCD ファイル) の解像度は 1280×720 ピクセルであり，それから抽出される点群データには大量の 3 次元情報が含まれているため，そのファイル数が増えると効率的な学習処理が困難となる．そこで，点群データ (PCD) をボクセルデータに変換し，それを 3D CNN の学習データとして用いた．今回は，1 つの PCD ファイルを $32 \times 32 \times 32$ のボクセルデータに変換し，3D CNN の学習のために，得られた全ボクセルデータを訓練用とテスト用のデータに分割した．図 5 にはカテゴリ No. 1 の点群データから変換したボクセルデータの例を示す．

本実験では，2 種類の分割方法で学習用及びテスト用のデータセットを用意し，それぞれデータセット A (1

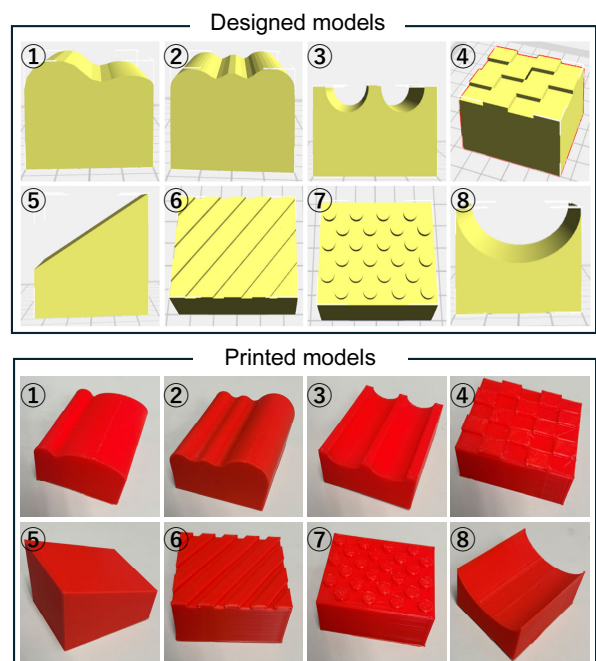


図 3 Examples of 8 categories of 3D models for classification, i.e., upper figures show examples of 3D CAD models, lower figures show their actual 3D-printed models.

セット)，データセット B (10 セット) とした．データセット A では，1 つのモデルから生成した 360 個のボクセルデータのうち 180 個を訓練用データ，残りをテスト用データに分割し，この方法を 80 個すべてのモデルのボクセルデータに対して適用した．一方，データセット B では，図 1 のように，1 カテゴリにつき 9 モデル分のボクセルデータ 3240 個から成る訓練用データセットと，1 モデル分のボクセルデータ 360 個から成るテスト用データセットをそれぞれ作成した．どの

表 1 How to divide all the voxel data into training and testing for dataset B , in this table, all the data of Model 1 are for test and the remained all the data of Model 2 to 10 are for training.

						Test	Training
Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8
Model 1	Model 1	Model 1	Model 1	Model 1	Model 1	Model 1	Model 1
Model 2	Model 2	Model 2	Model 2	Model 2	Model 2	Model 2	Model 2
Model 3	Model 3	Model 3	Model 3	Model 3	Model 3	Model 3	Model 3
Model 4	Model 4	Model 4	Model 4	Model 4	Model 4	Model 4	Model 4
Model 5	Model 5	Model 5	Model 5	Model 5	Model 5	Model 5	Model 5
Model 6	Model 6	Model 6	Model 6	Model 6	Model 6	Model 6	Model 6
Model 7	Model 7	Model 7	Model 7	Model 7	Model 7	Model 7	Model 7
Model 8	Model 8	Model 8	Model 8	Model 8	Model 8	Model 8	Model 8
Model 9	Model 9	Model 9	Model 9	Model 9	Model 9	Model 9	Model 9
Model 10	Model 10	Model 10	Model 10	Model 10	Model 10	Model 10	Model 10

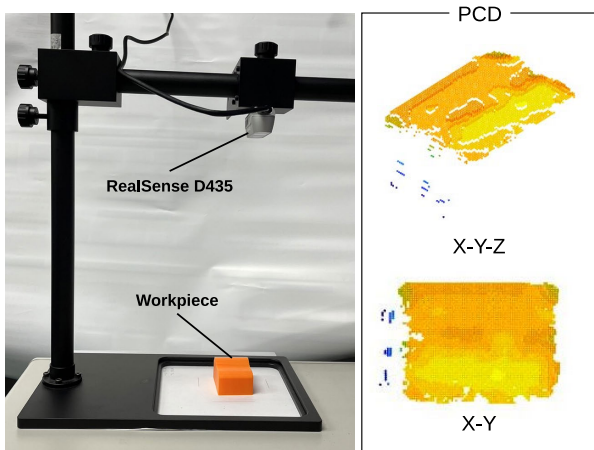


図 4 Measurement scene of a 3D-printed model to obtain a PCD file.

1モデルをテストに用いるかにより10通り(10セット)の訓練用データセットとテスト用データセットの組み合わせを用いて学習後の3D CNNの分類性能を評価した。なお、PCDのデータ拡張にはMATLAB上で開発した図6のようなアプリケーションを使用した。

4 ボクセルデータを用いた3D CNNの設計・学習・評価

分類用の3D CNNには2つの3次元畳み込み層と2つの全結合層を含む全12層で構成された表2のような構造を持つ3D CNNを設計した。3D CNNは3次元のフィルタによって時間軸に沿った情報や空間的な情報に含まれる特徴を抽出することができる。学習条件として、ミニバッチサイズ32、学習率0.0001、エポック数60に設定し、データセットAとデータセットBを用いて既にMATLAB上で開発している3D CNNのた

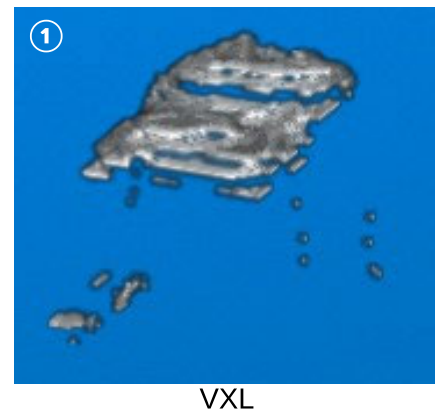


図 5 Example of voxel data transformed from a PCD file.

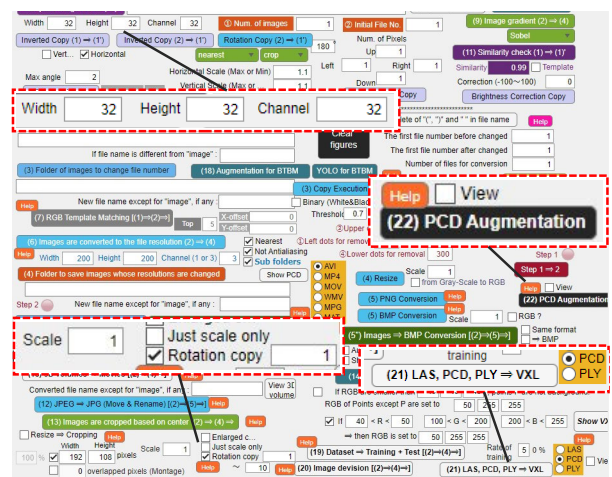


図 6 Data augmentation application developed on MATLAB.

めの設計アプリケーション [7, 8] 上で学習を行った。なお、今回はCPUにIntel(R) Core(TM) i7-12700K 3.60 GHzが、GPUにはNVIDIA GeForce RTX 3080 Tiが

表 2 Structural parameters of the constructed 3D CNN model designed for voxel data classification problem.

Type	Filter size	stride	Output size
Input	32×32×32	-	32×32×32×1
Convolution	5×5×5	2	14×14×14×32
Leaky ReLU	-	-	14×14×14×32
Convolution	3×3×3	1	12×12×12×32
Leaky ReLU	-	-	12×12×12×32
Max Pooling	-	2	6×6×6×32
Fully Connected	-	-	128
ReLU	-	-	128
Dropout	0.5	-	128
Fully Connected	-	-	8
Softmax	-	-	8
Output	-	-	8

Category 1	360							
Category 2		360						
Category 3			360					
Category 4				360				
Category 5					360			
Category 6						360		
Category 7							360	
Category 8	11							349
	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8

図 7 Example of Classification result given by confusion matrix, in which voxel data of Model 4 is used for testing.

搭載されたメインメモリ 32 GB の PC を用いた。

学習後、各データセットのテスト用データを用いて分類実験を行い、3D CNN における 3 次元モデルの分類における汎化性能を評価した。データセット A の訓練データを用いて訓練した 3D CNN は、データセット A の 8 カテゴリーのテストデータについて精度 100% で分類することができた。一方、データセット B の訓練データを用いて訓練した 10 器の 3D CNN モデルの中には図 7 の混同行列に示すように 99.6% の分類精度を示すものもあったが、10 器の分類結果の概要は図 8 のとおりであり、その分類精度の平均値は 75.2% 程度であった。図 8 より、どのモデルをテスト用データに用いるかによって分類結果にばらつきが生じている。この結

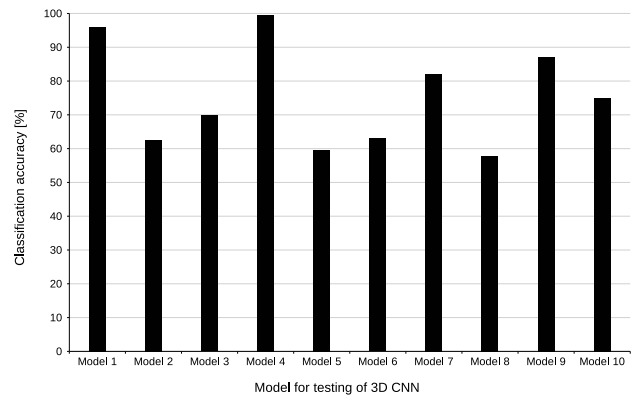


図 8 Classification accuracies predicted by 10 types of 3D CNN models, in which the combination of training and testing voxel data in dataset B is varied.

果から、訓練に用いていない未知のモデルに対する分類性能については十分な汎化性能が得られなかったと考えられる。誤分類の原因としては訓練データ、すなわち、各カテゴリーの類似モデル数が不十分であることにより、各カテゴリーに属するモデルの持つ広範な特徴を十分に捉えることができていない、あるいは各カテゴリー間で共通するような形状的特徴を強く認識してしまっていることなどが考えられる。

5 結言

本研究では、3D カメラで観測した深度情報をもとにボクセルデータを生成させ、ボクセルデータに基づく 3D CNN を設計し、表面形状の異なる 8 カテゴリーのワークの分類性能を評価した。データセット A の場合は、8 カテゴリー × 10 種類 = 80 のモデルから測定したボクセルデータを訓練用とテスト用に分けて使用したため、高い分類精度が得られたと思われる。一方、10 モデルのうち、どのモデルをテスト用に用いるかにより準備した 10 パターンのデータセット B については、汎化性能に大きなばらつきが見受けられた。今後は、識別が困難である形状的特徴を評価しながら対象モデルのカテゴリーとそのカテゴリーに含まれるワークの種類を増やすことで訓練用のデータセットを充実させ、より汎化性の高い表面形状識別のための 3D CNN モデルの構築を試みる。また、点群データまたはボクセルデータに対して形状的な特徴を強調させる拡張処理についても検討していく予定である。

参考文献

- [1] 永田, 廣野, 米原, 加藤, 渡辺, “動画解析と 2D CNN を用いた金属の火花試験の自動評価シス

- テム”，産業応用工学全国大会 2022 講演論文集，
pp 41–42，2022
- [2] 廣野，永田，米原，加藤，大塚，渡辺，“金属の火花試験の自動評価システム— 3D CNN と 2D CNN の分類性能の比較評価—”，第 23 回計測自動制御学会システムインテグレーション部門講演会論文集，pp. 325–328，2022
- [3] 田中，笠原，“画像を用いた自動外観検査技術”，*Ricoh Technical Report*, No. 42, pp. 26–33, 2017
- [4] 鎌田，中塚，山本，“外観検査に適した新たなニューラルネットワークの提案”，*Panasonic Technical Journal*, Vol. 66, No. 2, pp. 31–36, 2020
- [5] 長江，佐々木，吉田，片平，“機械学習によるモータコアの傷検出に関する研究”，*愛知電機技法*，No. 44, pp. 3–14, 2023
- [6] D. Maturana and S. Scherer, “VoxNet : A 3D Convolutional Neural Network for real-time object recognition,” *Procs . of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, pp. 922–928, 2015
- [7] 永田，渡辺，“畳み込みニューラルネットワーク (CNN)・畳み込みオートエンコーダ (CAE)・サポートベクタマシン (SVM) のための設計支援ツールの開発”，*画像ラボ*，Vol. 32, No. 12, pp. 20–26, 2021
- [8] 永田，阿部，有馬，清水，“Grad-CAM による工業材料の欠陥領域の可視化性能を向上させる画像拡張法と CNN による分類時のスコアの影響”，*画像ラボ*，Vol. 34, No. 5, pp. 1–8, 2023