

協働ロボットによる製造自動化技術の開発(第2報)

間瀬 慧* 平野 真**
MASE Satoshi* HIRANO Makoto**

要旨 本研究では少量多品種生産への対応可能なロボットとして、画像認識技術の高度化により自動で部品の位置と姿勢と併せて部品の種類も認識するシステムの開発に取り組む。本報告では今回の把持対象におけるYOLOのバージョンによる認識精度の比較と3Dデータ中にある把持対象を抽出する方法について検討を行った。結果、今回の把持対象においては認識精度、計算速度の面からYOLOv9が適していることと、マスキングを実施した深度画像を点群に変換し外れ点の除去することで3Dデータ中にある把持対象を抽出するが可能であることを確認した。

1 はじめに

近年、製造業における人手不足が深刻化しており、人材確保に何らかの課題のある企業は94.8%となっている¹⁾。特に大企業に比べ中小企業における人材確保の課題は深刻であり国内製造業のうち、人材確保の課題がビジネスにも影響が出ている中小企業は36.2%にも上っている²⁾。これらの課題に対し有効な対応の一つはICTや自動化技術を活用したDX(デジタルトランスフォーメーション)の推進である。

製造業におけるDXを進める代表的な方法の一つが産業用ロボットを活用した製造プロセスの自動化である。産業用ロボットは専用機械に比べ、多様なタスクを行うことが可能となっている。

新しいタイプの産業用ロボットとして協働ロボットに注目が集まっている。従来の産業用ロボットは作業者の安全性を確保するためにロボットを安全柵で囲い、動作中は作業者が柵の中に立ち入ることは出来ない。これに対し、協働ロボットは本体に安全に関するセンサが組み込まれており、適切にリスクアセスメントを行うことで安全柵無しでの設置が可能となっている。安全柵が不要なため設置スペースが従来の産業用ロボットに対して小さく済み、また設置場所の変更が容易であることから複数のロボットの導入が資金的・スペース的に困難な中小製造業での活用が進んでいる。

ロボットを活用した自動化が求められている工程として生産ラインへの部品の供給・搬送がある。一般的に部品はコンテナ内にランダムで配置されている場合が多く事前に動作をプログラムしてロボットに部品を把持させることは難しい。そのため画像認識等を組み合わせて把持対象の位置・姿勢を認識する技術について研究開発が行われている³⁾。

そのような中で、中小企業においては製品の設計変更や複数メーカーの製品など多種多様な部品を扱うことが多い。

そこで本研究では画像認識技術の高度化により自動で機械部品の位置と姿勢と併せて部品の種類・形状も認識するシステムの開発に取り組む。

本報告ではYOLOのバージョンによる認識精度の比較と3Dデータ中にある把持対象の抽出方法について検討を行った結果について報告する。

2 システム構成

ランダムで配置された部品を把持するには部品の3次元での位置と姿勢の情報が必要となる。そこで本研究では図1の3Dカメラ Zivid Two M70(Zivid製)を用いる。その仕様は表1のようになり、色付きの点群データとRGB画像が高精細に出力可能となっている。



図1 導入した3Dカメラ

表1 3Dカメラ仕様

メーカー	Zivid
型式	Zivid Two M70
撮像距離	0.5-1.1m
分解能	0.39mm (撮像距離0.7m)
撮像時間	80ms~1sec
出力データ	3D(XYZ)+Color(RGB)
画像サイズ	1944×1200画素(2.3MP)
インターフェース	10GigE

また今回、深層学習や認識、点群の処理には表2のワークステーションを用いている。

表2 ワークステーション仕様

型式	DAIV FW-X3N40
OS	Windows 11 Pro for Workstations
プロセッサ	Xeon w3-2423 プロセッサ
RAM	64GB
GPU	NVIDIA RTX A4000 16GB

次に本研究に用いる協働ロボットUR5e(Universal Robots製)の外観を図2に示す。またUR5eの主な仕様を表3に示す。

*機械システム係 **電子システム係

UR5eは6自由度のロボットアームとなり可搬重量は5kgで産業用ロボットの安全規格ISO 13849-1に準拠しており協働作業が可能である。



図2 協働ロボット (UR5e)

表3 協働ロボット仕様

型式	UR5e
メーカー	Universal Robots
可搬重量	5kg
リーチ	850mm
自由度	6軸 (ジョイント)
位置繰返し精度	±0.03mm (荷重負荷状態で測定、ISO9283に準拠)
標準TCP速度	1メートル/秒
電源	100-240VAC, 47-440Hz

3 画像認識の処理の流れ

本研究では機械部品の把持を想定している。そのため対象の3DCADデータを有しているものとする。3DCADデータがある場合、3Dカメラ等で取得した3Dデータの中から3DCADデータと一致する部分を見つけることで把持対象の位置と姿勢を推定することが出来る。この推定には特徴量とRANSACを用いたマッチング⁴⁾やICPアルゴリズム⁵⁾を用いることが出来る。ただし3Dカメラ等で取得した3Dデータ全体から一致する部分を見つけることは探索する範囲が大きく計算量も非常に大きくなる。そこで前処理として深層学習を用いて把持対象の可能性が高い範囲を抽出する。今回の目標である部品の位置と姿勢と併せて部品の種類・形状を認識するために図3に示す流れで処理を行うことにした。

RGB-D画像と点群は共に3Dデータであり相互に変換が可能であるが、それぞれの特徴は異なる。RGB-D画像は通常のRGB画像と深度画像(カメラからの距離を画素値として持っている画像)を組み合わせたもので技術的な蓄積が進んでいる2次元画像向けの深層学習を用いた認識が可能となる。これに対して点群は取得した3次元で取得した座標値の集合となっており、並進や回転、座標変換は容易であり点群マッチングやICPアルゴリズムを用いる環境が整っている。しかし深層学習における学習済みネットワークや大規模データセットが整備されておらず深層学習を用いるには学習コストが非常に大きい。そこで深層学習による認識が必要な把持対象の種類を認識、把持対象の

形状の抽出にはRGB-Dを用いて、位置と姿勢の推定には点群マッチングやICPアルゴリズムを用いる環境が整っている点群を用いることにした。

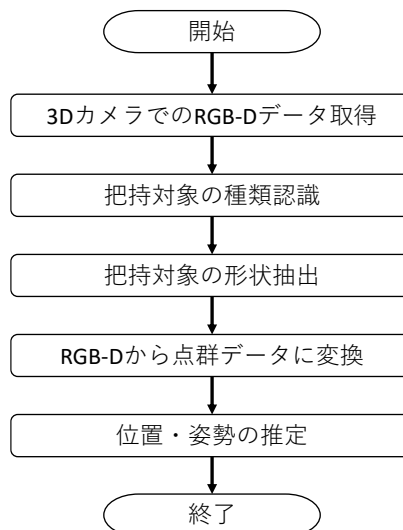


図3 認識フロー

4 YOLOモデルの評価・比較

本研究において把持対象の種類と位置を物体と背景を切り分けて認識するのに深層学習を利用したインスタンスセグメンテーションを用いている⁶⁾。その実装にはUltralyticsが公開しているYOLOを使用している。YOLOは進歩が目覚ましいモデルで日々、新しいバージョンが公開されている。今回はインスタンスセグメンテーションに対応しているYOLOのバージョンのうち、現時点で新しい順に3つのバージョン(YOLOv8、YOLOv9、YOLOv11)での今回の把持対象に対しての認識率を評価、どのバージョンを使用するかを検討した。具体的な方法としては実際に各モデルで深層学習によるインスタンスセグメンテーションのプログラムを作成、独自データセットによる学習と予測を行い、その結果に対して適合率を用いて評価を行った。

プログラムの開発環境は表2のワークステーション上に表4に示す環境に構築した。

表4 開発環境

言語	Python3.11.9
並列計算	CUDA12.1
実行	Jupyter Notebook

学習については表5に示すUltralyticsが公開している表5に示すインスタンスセグメンテーションに対応した学習済みモデルyolo8x-seg、yolov9e-seg、yolo11x-segに対して、同一の独自データセットによる学習を行った。

独自データとしてトレイ上に複数の歯車を配置したものを10枚用意し、labelmeでアノテーションを行ったものに対して、roboflowを用いて訓練用、検証用、テスト用で8:1:1

に分割、訓練用データを表5の条件で3倍にデータ拡張を行ったものを用いた。

表5 データ拡張条件

Auto-Orient	○
Resize	640×640
Flip	Horizontal, Vertical
Crop	0%~20%
Rotation	-15° ~15°
Shear	±10° Horizontal, ±10° Vertical
Saturation	-25% ~25%
Brightness	-15% ~15%
Noise	Up to 0.1% of pixels

このデータを用いて学習済みモデル(yolov8x-seg)に対して転移学習を100エポック行った。

学習中に最も精度が良かったときのパラメータを用いてテストデータでの予測を行った。テストデータとしては今回、トレイ上に歯車の数を1個、5個、8個、9個と変えながらランダムに配置した状態で3Dカメラにより撮影したRGB画像を3枚用意。それぞれに対して認識を行った。その結果を機械学習モデルにおける予測の正確さの評価に用いられる適合率により評価を行った。

その結果をまとめたものを表6に示す。これらの結果はそれぞれの条件で3回認識させたときの適合率の平均値となる。適合率は0から1の値となり、1に近づくほどその機械学習モデルの予測が正確であることを示す。今回のテストではyolov8x-segとyolov9e-segは認識対象の見落としや見間違いなどの誤認識は無く、非常に高い性能を示した。これに対してyolo11x-segは図6のような認識対象とは別の部分を検出するような誤認識が常に1~2個発生していたためやや性能が落ちる結果となった。

表6 形状抽出結果

歯車の数	yolov8x-seg	yolov9e-seg	yolo11x-seg
1個	1	1	0.3
5個	1	1	0.8
8個	1	1	0.8
9個	1	1	0.9

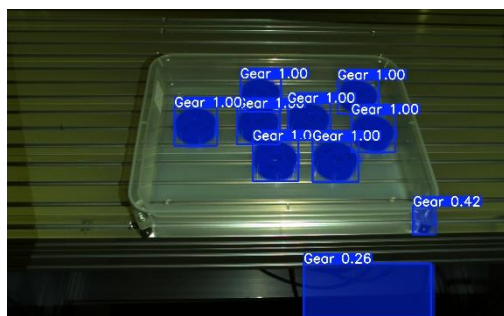


図6 yolo11x-seg誤認識

本結果より結果より適合率に関してはYOLOv8またはYOLOv9は適合率としては同等となっている。これに対し

て表7にUltralyticsで公開されている各モデルで用いられているパラメータの数、計算速度を示す。これによりYOLOv9の方がパラメータの数が小さく、計算速度が速いため今回はYOLOv9を用いることとする。

5 把持対象の3D形状抽出

前報でインスタンスセグメンテーションにより認識対象のマスク画像を生成することが可能であり、これを用いてRGB画像上での形状の抽出が可能であることを確認した(6)。これに対して、マスク画像を用いて把持対象の3D形状の抽出を行う方法について検討した。今回の検討に使用するRGB画像と深度画像を図7、図8に示す。



図7 テスト用RGB画像

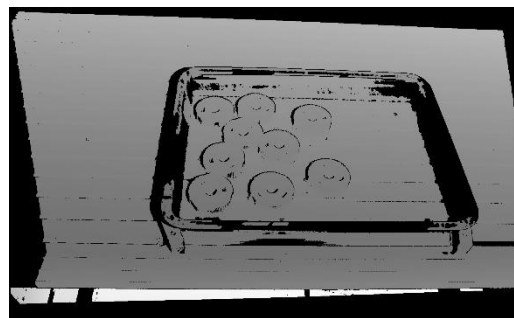


図8 テスト用深度画像

RGB画像と深度画像の両方に対してインスタンスセグメンテーションにより生成した図9のマスク画像を適用。これらを用いてRGB-D画像を生成して、これをピンホールカメラモデルにより点群に変換した。今回、深度画像から点群への座標の変換に必要なスケールやカメラパラメータは既知である。



図9 マスク画像

その結果が図10のようになり視野に沿って一部の点群が視野に沿って配置されている。

謝辞

本研究においてご助言頂いた滋賀県立大学 工学部 宮城茂幸准教授、山野光裕准教授に謝意を申し上げます。

参考文献

- 1) 経済産業省. “ものづくり白書 2019年版 第1部 第2章 第1節 我が国製造業の足元の状況”. 2019. p. 38-39. URL : https://www.meti.go.jp/report/whitepaper/mono/2019/honbun_pdf/pdf/honbun_01_02_01.pdf (参照日 : 2022年3月14日)
- 2) 経済産業省. “ものづくり白書, 2019年版 第1部 第2章 第3節 世界で勝ち切るための戦略-Connected Industriesの実現に向けて-”. 2019. p. 186-187. URL : https://www.meti.go.jp/report/whitepaper/mono/2019/honbun_pdf/pdf/honbun_01_02_03.pdf (参照日 : 2022年3月14日)
- 3) 山崎公俊. “物品掴み上げ作業で求められる視覚機能”. 日本ロボット学会誌. 2020. Vol. 38 No. 6 p.538-541
- 4) 金崎朝子, 秋月秀一, 千葉直也. 詳解 3次元点群処理. 講談社. 2022. p. 91-103.
- 5) 金崎朝子, 秋月秀一, 千葉直也. 詳解 3次元点群処理. 講談社. 2022. p. 64-68.
- 6) 間瀬慧, 平野真. “協働ロボットによる製造自動化技術の開発”. 令和5年度 業務報告書. 滋賀県東北部工業技術センター報告. 2024. p. 76-79.
- 7) 金崎朝子, 秋月秀一, 千葉直也. 詳解 3次元点群処理. 講談社. 2022. p. 37-38.



図10 RGB-D画像から点群への変換結果

この原因はインスタンスセグメンテーションで出力されるマスク画像が実際の画像のサイズより小さく、拡大処理をしたため本来各画素に対して0か1の値のみを持つマスク画像の周縁部に補間処理が入り0~1の値を持つことになったためであった。ただこれらの点はバラバラの位置になっており外れ値除去の処理で十分に取り除くことが可能である。

3Dセンサで取得したデータには精度の出る視野範囲をよりも近い、または遠い点群やノイズや反射などによるエラー一点が含まれ、外れ値を除去する処理は必須となる。今回は図10のデータに対して以下の2つの処理を行い、不要な点を削除した。

1. カメラの撮像範囲外の点を除去
2. 半径外れ値除去⁷⁾により不要な点を除去

半径外れ値除去で設定する半径と点の数については試行錯誤により決定した。その結果が図11のようになり認識対象の3D形状が抽出出来ていることが確認できた。



図11 外れ値除去の結果

6 まとめ

把持対象の位置と姿勢と併せて、種類も認識・把持するシステムの開発の目的として今回の把持対象におけるYOLOのバージョンによる認識精度の比較と3Dデータ中にある把持対象を抽出する方法について検討を行った。結果、今回の把持対象においては認識精度、計算速度の面からYOLOv9が適していることと、マスキングを実施したRGB-D画像を点群に変換し外れ点を除去することで3Dデータ中にある把持対象を抽出することが可能であることを確認した。