

把持位置学習における中間特徴量を用いた物体形状分類手法の検討

横田泰斗†

鈴木 彼方†

金澤裕治†

竹林知善†

† 富士通研究所

1 はじめに

製造業の分野では、生産ライン自動化のために製品や部品の把持(ピッキングタスク)を行うロボットが広く普及している。ロボットの動作や姿勢制御の設定にはティーチングと呼ばれる専門技術が求められる作業が必要であり、動作や把持対象の物体形状を変更するたびに大きなコストがかかる。近年、深層学習(Deep Learning)をピッキングタスクに適用してティーチング作業を削減する研究が進んでおり[1][2]、特に、深層強化学習[3]や、試行結果から学習データを逐次的に追加する能動学習[4]が注目されている。

一般的にピッキングタスクでは、物体の写った画像に対して把持位置情報を正解ラベルとして付与した訓練データを学習させることで把持位置検出を実現する。しかしながら、実際の工場や倉庫でのピッキングタスクでは複数の形状の物体が混在しているケースがあり、形状に応じて分類することが求められる。そのためには分類用のラベル付けを別途行う必要があり、訓練データの作成に非常に大きな工数がかかる。ラベルを付けずに複数の形状を分類する手法[5]も考案されているが、学習に膨大な時間を要するため実用的ではない。

本稿では上記問題に対し、物体の把持位置検出を学習させた畳み込みニューラルネットワーク(Convolutional Neural Networks, 以下CNN)の中間特徴量を可視化し、中間特徴量を用いて物体形状を分類する手法を検討する。

2 物体形状分類手法

2.1 把持位置検出

本研究では、把持位置検出に Single Shot Multibox Detector (SSD)[6]を用いる。SSDは物体検出に用いられる深層学習モデルの一種で、デフォルトで用意される大量のアスペクト比の異なる矩形領域から正解の矩形領域を出力する方式のため、把持位置 (x, y, w, θ) を矩形表現に変換した上で学習を行う。訓練データの物体形状によっては1つの物体に対して正解の把持位置を複数設定する。また、通常SSDの出力に加えて角度 θ を予測するように学習モデルを拡張しておく。これらは正解との二乗誤差により最適化される。ロボット実機での把持の際は、正解として出力された中で最も信頼度が高い矩形を把持位置に変換して把持動作を試行する。

Object Classification using Hidden Layer Outputs in Grasp Point Detection
Yasuto Yokota, Kanata Suzuki, Yuuji Kanazawa, Tomoyoshi Takebayashi
(Fujitsu Laboratories LTD.)

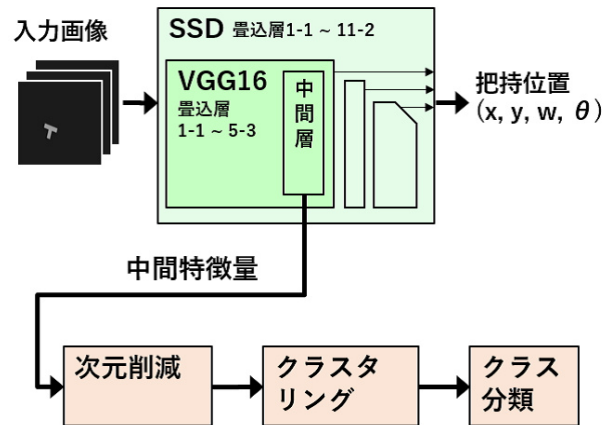


図1: 提案手法の概要

2.2 中間特徴量による分類

提案手法(図1)では、一定数の把持位置検出の試行に対してSSDの中間特徴量について次元削減を行った後、複数の集合にクラスタリングする。各集合のクラスは、その集合内の要素数が最多となる物体形状のクラスとする。これによって、以降の把持位置検出時には中間特徴量の次元削減結果が属する集合のクラスをもとに物体形状を分類することができる。上記を実現するために、把持位置検出の際にSSD各層の中間特徴量を保存しておき、任意の層の中間特徴量に対して次元削減を行うことができる仕組みとする。

なお、クラスタリングの入力には、物体形状の違いにより特徴が表れやすい層、及び次元の特徴量を用いる。これは把持位置検出対象の物体位置に依存する特徴量の影響を軽減し、クラスタリングの際に物体形状に応じた集合に分割させるためである。

3 中間特徴量の可視化実験

3.1 実験設定

上記手法に適した中間特徴量の条件を確認するため、一定回数試行した際の中間特徴量について主成分分析(PCA)による次元削減を行い、主成分可視化とクラスタリングを実施した。

事前作業として、入力画像中の物体の把持位置を予測する学習モデルを作成した。入力画像は、円柱、星形、L字、T字の4種類の形状の物体のいずれか1つがランダムな位置に配置されている画像とする。ここでは、シミュレータ上で入力画像と把持位置の正解ラベルを自動生成して訓練データを800セット用意し、100

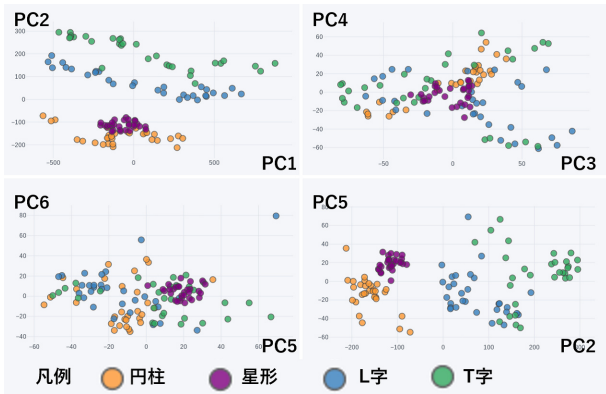


図 2: 実験結果: 5-1 層 主成分

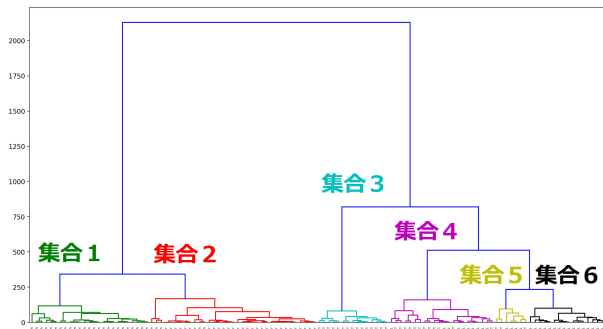


図 3: 実験結果: 5-1 層 クラスタリング

エポック学習させている。SSD への入力は $300 \times 300 \times 3$ [pix] 画像を用い、最適化には Adam を用いた。

実験として、各形状 20 枚ずつ計 80 枚の画像をシミュレータで新規作成し、それらに対して把持位置検出した際の主成分の可視化を行った。可視化対象は SSD の 5-1 層、8-1 層の出力とし、第一主成分 (PC1) から第六主成分 (PC6) までの 6 つの主成分のうち 2 つを選択して散布図を作成した。

また、物体形状に応じて分布している主成分を入力として Ward 法によるクラスタリングを行い、各集合の物体形状のクラスを確認した。なお、集合に分割する際の閾値は要素間距離の最大値の 0.1 倍とした。

3.2 結果と考察

まず、主成分可視化を行った結果を図 2 に示す。5-1 層では、PC2 と PC5 の主成分が物体形状の影響を大きく受けていることが確認できる。PC2 は上から順に T 字、L 字、星形、円柱の主成分が分布している。また、PC5 は主に星形と円柱が中央を境に分離している。一方、8-1 層ではほとんどの主成分が入り混じって分布していた。これは本実験での SSD の最終出力が把持位置であるため、最終層に近い 8-1 層の主成分は物体形状よりも物体位置と把持角度の影響を受けやすいことが要因と考えられる。これらの結果から、5-1 層の PC2、PC5 が提案手法で用いる主成分に適していることが確認できた。

次に、5-1 層の PC2、PC5 の主成分を入力として Ward 法によるクラスタリングを行った結果 (図 3)、6 つの集

合に分割された。各集合と要素に対する物体形状のクラスを表 1 に示す。本実験全体での集合と要素のクラスが合致する確率は 0.85 となった。集合 1, 3, 6 では全ての要素のクラスが合致しており、物体形状ごとに分類できている。また、集合 2 は 35 個中 30 個の要素が星形であり、0.86 の確率で合致している。一方、集合 4, 5 では L 字と T 字がほぼ同じ割合で混在している。T 字の物体形状の一部が L 字形状の相似形であるため、主成分が比較的近い値になることが原因と考えられる。

上記の実験結果から、提案手法での分類に適した中間特徴量の主成分を確認できた。

表 1: 実験結果: 集合と要素のクラス

集合 No.	集合のクラス	要素のクラス				クラス合致率
		円柱	星形	L 字	T 字	
集合 1	円柱	25	0	0	0	1.00
集合 2	星形	5	30	0	0	0.86
集合 3	T 字	0	0	0	15	1.00
集合 4	T 字	0	0	10	12	0.55
集合 5	L 字	0	0	4	3	0.57
集合 6	L 字	0	0	16	0	1.00
実験全体		30	30	30	30	0.85

4 まとめと今後の展望

本研究ではピッキングタスクにおいて、物体把持位置検出を学習させた CNN の中間特徴量を用いて物体形状を分類する手法を検討した。中間特徴量の主成分を可視化することで提案手法に適する主成分を確認し、それらを入力としてクラスタリングすることで物体形状を分類することが可能なことを示した。今後の展望として、提案手法を拡張し、入力画像中に複数物体が存在するケースに適用する。また、分類に適した中間層と主成分を自動で選択するアルゴリズムを設計する。

参考文献

- [1] J.Tobin, et al., "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World," in Int. Conf. on Intelligent Robots and Systems (IROS), 2017.
- [2] K.Bousmalis, et al., "Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping," in Proc. IEEE Int. Conf. Robots and Automation (ICRA), 2018.
- [3] S.Levine, et al., "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," Int. J. Robot. Res., vol.37, no.4-5, pp.421-436, 2018.
- [4] L.Pinto, et al., "Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours," in Proc. IEEE Int. Conf. Robots and Automation (ICRA), 2016.
- [5] E.Jang, et al., "Grasp2Vec: Learning Object Representations from Self-Supervised Grasping," in Conf. on Robot Learning (CoRL), 2018.
- [6] W.Liu, et al., "Ssd: Single shot multibox detector," in European Conf. Computer Vision (ECCV), 2016.