

# Multi-task DSSD による物体位置と物体把持位置の同時推定

荒木諒介 大西剛史 平川翼 山下隆義 藤吉弘巨 (中部大学)

## 1. はじめに

物流倉庫や生活空間で稼働するロボットには、多品種の物体検出と把持位置検出が求められる。本研究では、深層学習による高速な物体検出モデル [1, 2] をもとに、物体領域のセマンティックセグメンテーションと、吸着ハンド用の把持位置検出を同時に行う手法を提案する。ロボットを用いた把持実験により、提案手法が RGB 画像から出力する把持位置が適切であることを示す。

## 2. 関連研究

Single Shot Multibox Detector (SSD)[1] は、物体候補領域の検出とクラス分類を単一のネットワークで行う。複数解像度の特徴マップを用いて、様々なスケールの物体を検出する。6つの予測レイヤでは、デフォルトボックスと呼ぶ基準矩形に対するオフセットを特徴マップ (Localization maps) として出力し、物体の矩形を回帰する。また、物体のクラス尤度を表す特徴マップ (Confidence maps) を用いて、物体のクラスを識別する。この処理で得られた複数の物体候補領域に対して、Non-maximum suppression (NMS) 処理により重なり合う同じクラスの尤度が低い物体候補領域を除去し、過検出を抑制する。

SSD に Deconvolution 層を追加した Deconvolutional Single Shot Detector (DSSD)[2] は、Deconvolution 処理により特徴マップをアップサンプリングして拡大する。エンコーダに相当する畳み込み処理から得られる特徴マップと1つ前の層から得られる特徴マップを、Deconvolution module で結合する。結合された特徴マップを用いることで、物体検出の高精度化を実現した。

## 3. 提案手法

本研究では、マルチタスク化 (物体検出・セグメンテーション・物体把持位置検出) をした Multi-task DSSD を提案する。本手法のネットワーク構造を図1に示す。DSSD の予測レイヤにセグメンテーション用と把持位置検出用の特徴マップを生成する層を追加した構造である。

Deconvolution module では、図2の左側に示すように、前層の特徴マップと対になるエンコーダの特徴マップを加算して活性化関数 (ReLU) に入力する。こ

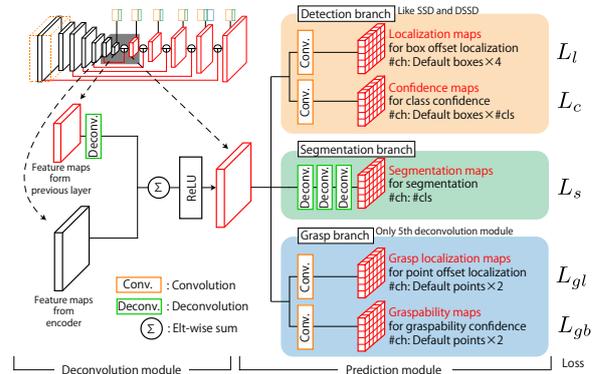


図2 提案手法の Deconv. module と予測レイヤ

のとき、前層の特徴マップを Deconvolution 処理により、エンコーダの特徴マップと同じサイズに拡大する。Deconvolution module の処理後、図2の右側に示すように、予測レイヤの3つの Branch により物体検出とセグメンテーション、把持位置検出を同時に行う。

**Detection branch:** 従来の DSSD と同様に、Localization maps と Confidence maps を出力する。

**Segmentation branch:** Deconvolution を3回行い、入力画像サイズまで拡大した Segmentation maps を出力する。6つの予測レイヤで得た Segmentation maps を結合することで、最終的なセグメンテーション結果を得る。また、各レイヤの Detection branch で得た矩形に NMS を適用し、最も正解に近い矩形に絞り込む。

**Grasp branch:** (5番目の予測レイヤのみ) デフォルトボックスのような基準把持位置を表すデフォルトポイントの回帰を行う Grasp localization maps と、各デフォルトポイントの把持しやすさを表す Graspability maps を出力する。Detection branch と同様に、得られた把持位置候補に NMS を適用し、最も正解に近い把持位置に絞り込む。

## 4. 評価実験

提案手法の有効性を評価するために、Team MC<sup>2</sup> ARC2017 RGB-D Dataset を用いて実験する。データセットは、1,210枚の学習サンプル、200枚の評価サンプルが含まれており、各画像には物体検出とセグメンテーション、把持位置のラベルが付与されている。認

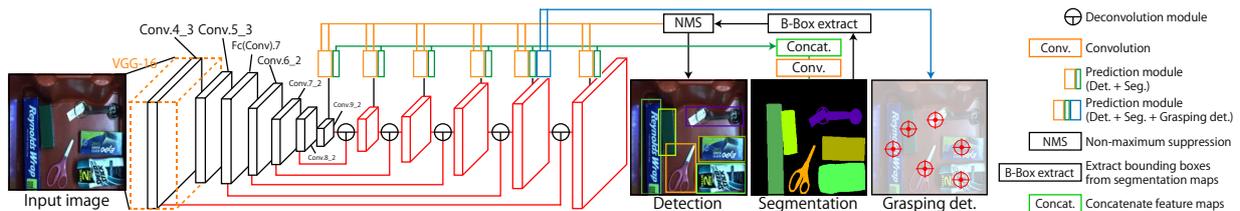


図1 Multi-task DSSD のネットワーク構造

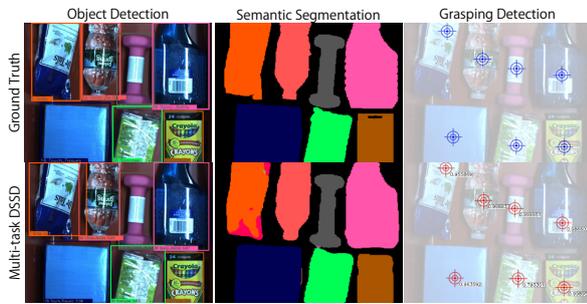


図 3 Multi-task DSSD の検出結果例

表 1 物体検出の結果比較 [%]

手法	認識率	未検出率	平均 IoU
SSD	87.94	29.65	81.05
Single-task	89.81	30.33	81.11
Multi-task	94.00	33.24	83.54

表 2 セグメンテーションの結果比較 [%]

手法	GlobalAcc.	ClassAcc.	平均 IoU
SegNet	78.19	72.16	53.97
U-Net	79.36	70.32	54.88
PSPNet	85.33	80.09	68.34
Multi-task	90.53	85.32	76.27

表 3 把持位置検出の結果比較

手法	正解率 [%]	未検出率 [%]	平均距離 [pixel]
重心	51.86	34.57	38.56
Multi-task	76.32	33.08	28.45

識対象アイテムは 40 種類あり、箱状、ビニール包装、非剛体など形状は様々である。学習サンプルはデータ増強により 30 倍の 36,300 枚に増幅する。

#### 4.1 評価方法

提案手法は 3 つのタスクを同時に行う。各タスクの評価方法は以下である。

- A. 物体検出 検出された物体のうち正しく認識できた割合を表す認識率、全ての物体に対して検出できなかった割合を表す未検出率、検出された矩形と教師レベルの矩形の重なり率の平均を表す平均 IoU を用いる。
- B. セグメンテーション Global Accuracy, Class Accuracy, Mean IoU を用いて評価する。
- C. 把持位置検出 検出された把持位置のうち、正解把持位置ラベルとの画像 (1280 × 960 [pixel]) 上のユークリッド距離が 80 [pixel] 以下である把持位置の割合を表す正解率、全ての正解把持位置に対して把持位置が検出できなかった割合を表す未検出率、全ての正解した把持位置のユークリッド距離の平均を表す平均距離を用いる。

#### 4.2 実験結果

物体検出の評価結果を表 1 に、検出結果例を図 3 左側に示す。提案手法は SSD と比べて識別率が 6.06 ポイント向上した。また、提案手法は Single-task DSSD と比べて重なり率が 2.17 ポイント向上した。次に、セグメンテーションの評価結果を表 2 に、セグメンテーション例を図 3 中央に示す。提案手法は従来手法に比

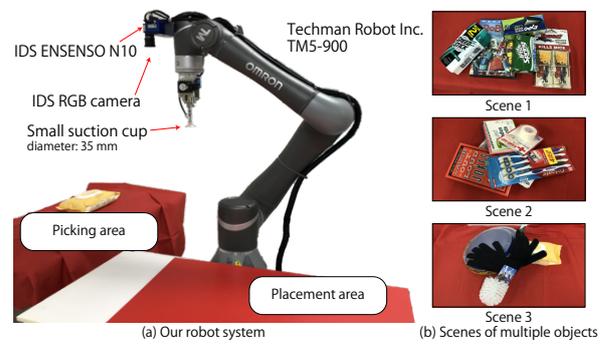


図 4 ロボットシステム

表 4 ロボット実機での把持実験結果

		Scene1	Scene2	Scene3
ピッキング	重心	7	9	19
試行回数	提案手法	4	5	4
把持	重心	57	44	11
成功率 [%]	提案手法	100	100	75

べ、各評価指標において約 5.2 から 22.3 ポイント向上した。把持位置検出の評価結果を表 3 に、把持位置検出例を図 3 右側に示す。青色の印が正解把持位置、赤色の印が検出結果である。セグメンテーション結果の重心を用いた場合と比較すると、検出率は 24.46 ポイント向上し、平均距離は 10.11pixel 向上し、最適な把持位置を検出できた。

#### 4.3 ロボット実機での把持実験

ロボット実機での把持実験により、提案手法が推定する把持位置を評価する。本実験に用いるロボットシステムを図 4(a) に示す。ロボットは、picking area にある物体の把持位置を検出し、得られた把持位置の法線方向にアプローチする。物体を把持し、placement area に移動できたとき成功と判定する。提案手法は、セグメンテーション結果の重心を把持位置として用いる場合と比較する。把持シーンは図 4(b) に示すように、3 つのシーンで実験を行った。

各シーンのピッキング試行回数と把持成功率を表 4 に示す。重心を用いる場合、いくつかのアイテムはうまく把持できない。提案手法は、吸着の難しい領域を避けて把持できることを確認した。

#### 5. おわりに

本研究では、Multi-task DSSD を用いた物体領域と物体把持位置の同時推定を行った。物体や把持位置を高速に検出することで、リアルタイム動作計画への活用が期待できる。

#### 謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものである。

#### 参考文献

- [1] W. Liu *et al.*, “SSD: Single Shot Multibox Detector”, ECCV, 2016.
- [2] C. Fu *et al.*, “DSSD: Deconvolutional Single Shot Detector”, arXiv preprint arXiv:1701.06659, 2017.