

1. はじめに

物体検出は自動運転や産業用ロボットにおいて重要なタスクであり、高精度化が大きな課題である。深層学習による物体検出法として、Deconvolutional Single Shot Detector (DSSD) が提案されている [1]。DSSD による物体検出は中間層の小さい特徴マップより原画像における物体を予測する際に、矩形の大きさや位置に誤差が生じることがある。そこで、本研究では、DSSD を物体検出とセグメンテーションのマルチタスク化し、セグメンテーションの結果を物体検出に反映することで、物体検出の高精度化を図る。

2. Deconvolutional Single Shot Detector

マルチスケールな物体検出法に Single Shot MultiBox Detector (SSD) がある [2]。SSD は VGGNet をベースに後層の複数の層で物体検出を行う。SSD の改良版である DSSD は、ネットワークの後層に Deconvolution 層を追加し、アップサンプリングした高解像度の特徴マップを利用することで、小さなスケールの物体検出が可能となる。

3. 提案手法

提案手法では、物体検出とセグメンテーションを同時に行うために、DSSD の Decoder にセグメンテーション用の Deconvolution 層を追加する。また、Decoder で物体検出とセグメンテーションを行い、Non-Maximum Suppression (NMS) 時にセグメンテーションの結果を物体検出に反映する。提案するネットワークの構造を図 1 に示し、以下に各処理の詳細について述べる。

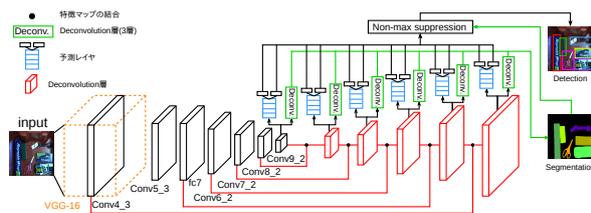
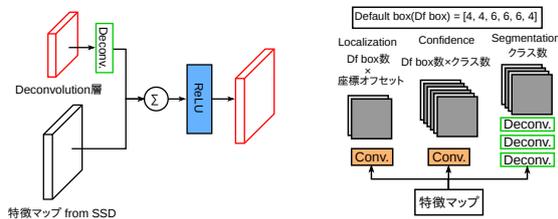


図 1：提案手法のネットワーク構造

3.1. Decoder の改良と各タスクの出力

提案手法では、図 2(a) のように、Encoder の特徴マップと Decoder の特徴マップを要素ごとに総和して結合する。これにより、Decoder の高解像度な特徴マップからの物体検出及びセグメンテーションが可能となる。物体検出とセグメンテーションの出力方法を図 2(b) に示す。結合した特徴マップに対して畳み込みと Deconvolution を行い、物体の位置を予測する Localization map と、物体のカテゴリを予測する Confidence map、セグメンテーションをする Segmentation map の 3 つを出力する。



(a) 特徴マップの結合 (b) 各タスクの出力

図 2：提案手法の Decoder と予測レイヤ

3.2. セグメンテーション結果を反映した NMS

物体検出時に NMS を用いる際、セグメンテーション結果を物体検出に反映することで、正確な矩形の検出が期待できる。セグメンテーション結果より物体の外接矩形 S を求め、物体検出により予測した矩形 D との重なり率 (IoU) を求める。そして、物体のクラス確率 p_c と合わせ正規化するため平均を求める。式 (1) にクラス確率を求める式を示す。

$$score = \left(p_c + \left(\frac{area(D \cap S)}{area(D \cup S)} \right) \right) / 2 \quad (1)$$

ここで、 $area(\cdot)$ は面積を求める関数である。提案手法では上記の $score$ を NMS に用いる。

4. 評価実験

提案手法による物体検出とセグメンテーションの効果を示すために、評価実験を行う。

4.1. 評価方法

評価には 40 種類のアイテムを含む ARC2017 RGB-D Dataset を使用する。学習には、1,210 枚の画像を Data Augmentation した 36,300 枚を用いる。評価には 200 枚を用いる。比較する手法は、物体検出は SSD と DSSD、セグメンテーションは SegNet を用いる。また、提案手法 1 はセグメンテーションを反映しない NMS、提案手法 2 は反映した NMS を用いたネットワークを示す。

4.2. 実験結果

物体検出及びセグメンテーションの比較結果を表 1 に示す。ここで、識別率は検出した矩形のクラス正解率、重なり率は教師と予測した矩形の重なり割合、GA は Global Accuracy, CA は Class Accuracy を示している。提案手法 2 は、DSSD と比べて識別率が約 3% 向上した。提案手法 1 と提案手法 2 を比較すると、重なり率が約 2.7% 向上しており、セグメンテーション結果を NMS に反映した効果を確認した。また、SegNet と比較して、提案手法はマルチタスク化の効果により、Mean IoU を含む全てのセグメンテーション精度を向上させることができた。

表 1：精度比較 [%]

手法	検出			セグメンテーション		
	識別率	未検出率	重なり率	GA	CA	Mean IoU
SSD	86.45	29.86	81.28	-	-	-
DSSD	88.43	30.19	81.37	-	-	-
SegNet	-	-	-	78.19	72.16	53.97
提案手法 1	90.47	29.97	80.95	91.80	87.31	78.30
提案手法 2	91.50	29.68	83.68	91.80	87.31	78.30

物体検出とセグメンテーション例を図 3 に示す。提案手法 2 は、従来の DSSD と比較すると、見切れたペットボトルに対しても検出できている。また、提案手法 1 よりも矩形のずれが補正されていることがわかる。一方セグメンテーションにおいて、提案手法は SegNet と比較して、ノイズが減少していることがわかる。



(a) DSSD (b) 提案手法 1 (c) 提案手法 2



(d) SegNet (e) 提案手法

図 3：物体検出とセグメンテーションの例

5. おわりに

本研究では、物体検出とセグメンテーションを同時に行う DSSD を提案し、この有効性を確認した。今後は、ネットワーク構造の見直しと高速化を目指す。

参考文献

- [1] C. Fu, *et al.*, “DSSD: Deconvolutional Single Shot Detector”, arXiv, no.1701.06659, 2017.
- [2] W. Liu, *et al.*, “SSD: Single Shot MultiBox Detector”, ECCV, pp. 21-37, 2016.