

多層に重なる同種物体の Amodal Instance Segmentation の実現

D-12 Amodal Instance Segmentation of the same class objects in multiple layers

広松 太一

Taichi Hiromatsu
芝浦工業大学

Shibaura Institute of Technology

高橋 正信

Masanobu Takahashi
システム理工学部

College of Systems Engineering and Science

1. はじめに

物体のクラス分類と1ピクセル単位の物体領域抽出を行う技術を Instance Segmentation(以下 IS)という。この技術で現在主流となっている手法は 2017 年に発表された Mask RCNN[1](以下 MR)である。しかし、MR には物体の重複や交差が起きた際に物体が正しく IS されないといった欠点がある。これは、物体の前後関係(以下 Occlusion)が考慮できないからである。

多層に重なる多クラスのオブジェクトを前後関係も含めて認識し、隠れた部分も含めて IS する Amodal Instance Segmentation の手法としては、Amodal Instance Segmentation with KINS Dataset[2](以下 AISKD)が提案されている。しかし、この手法の多クラス分類機能は、群衆や畑に植えられている野菜などの単一クラスの画像に対しては不要な機能である。この不要な機能を除き、単一クラスに限定することで、高精度な AIS の実現が期待できる。

2. 目的

本研究の目的は、AISKD で提案されている多層に重なる多クラスのオブジェクトの AIS を実現するフレームワークをもとに、単一クラスに特化して多層のオブジェクトの AIS を実現するフレームワークを実装する。具体的には多層の同種物体の前後関係予測と重複領域の補完機能を実現する。図 1 は単一クラスを車とした場合の理想的な AIS の結果例である。



図1 本手法の理想的な結果例

3. 手法

図 2 に提案手法のフレームワークを示す。これは AISKD をもとにクラス分類に関連する処理を省き簡略化したものであり、Backbone, RoI Align, Multi-Task Branch, Multi-Level-Coding(以下 MLC), Multi-Head の 5 ステップで構成されている。

Backbone は元画像の特徴を畳み込みニューラルネットで抽出する処理である。RoI Align は Backbone で取得した元画像の特徴から関心領域(以下 RoI)を切り出し、RoI の特徴を抽出する処理である。Multi-Task Branch は RoI Align で取得した RoI の特徴を入力とし、並行して Box, OC, Mask Branch に処理を分岐する。順に、物体の矩形領域、物体の前後関係、1ピクセル単位の物体領域の特徴を抽出する処理である。MLC は Box, OC, Mask で個別に学習される特徴を組み合わせ、総合的に学習し易く

する工程である。Multi-Task Branch で取得した Box, OC の特徴と Mask の特徴を連結させる。Box (物体の矩形領域), OC(前後関係分類), Mask(1ピクセル単位の物体領域)のすべての特徴を連結させた特徴を出力する。Multi-Head は Box, OC の特徴からそれぞれ物体の矩形領域、物体の前後関係の結果を出力し、Box, OC, Mask の特徴から1ピクセル単位の物体領域を出力する。この出力結果と正解データとの誤差を逆伝播させ、本フレームワーク全体で学習を行う。

AISKD のソースコードは公開されていないため、MR のソースコードをベースに不足する機能を実装することで全体の処理を実現した。

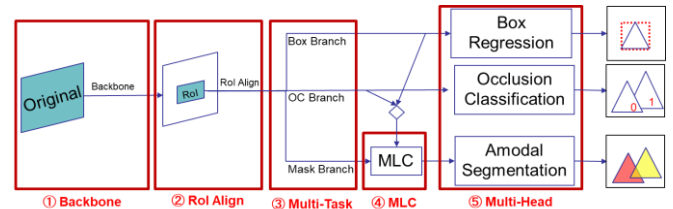


図2 提案手法のフレームワーク

4. 実験

多層の単一クラスの単純な図形(円)を用いた実験を行った。色と座標と半径(20~127)をランダムに設定した円3個を3層に重なるように配置した画像を使用した。また、重複の割合は0.3未満のものとした。学習用画像200,000枚で学習し、評価用画像10,000枚で評価した。

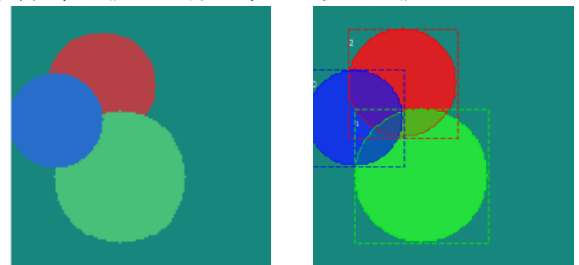


図3 データセットの入力例(左)と正解例(右)

セグメンテーションの精度(Mask の精度)には、正解と出力の物体領域の和と正解の物体領域との一致率(正解と一致した面積 / (正解と出力の面積の和 - 正解と一致した面積))を用いた。実験の結果、Mask の精度として0.86が得られ、重複して隠れた部分を比較的高い精度で補完する機能が実現できた。また、認識された物体の前後関係の順番(OC)の正解率も0.92と高い値が得られた。

今後は複雑な物体や4層以上の重複で実験し汎用性を検証していきたい。また、フレームワークの改善案としては、今回提案したフレームワークを生成器とし、生成したMaskと正解を識別する識別器を実装し、GAN を用いて敵対的に学習させる手法などが考えられる。

[参考文献]

- [1]Kaiming He, et al. : "Mask r-cnn", IEEE-ICCV2017, pp. 2961-2969, 2017.
[2]Lu Qu, et al. : "Amodal Instance Segmentation with KINS Dataset", CVPR2019, pp. 3014-3023, 2019.