

シードを用いた対象指定による線状物体の抽出 Extraction of Thin Objects by Specifying Targets Using Seeds

菅家 遼平[†]
Ryohei Kanke

高橋 正信[†]
Masanobu Takahashi

1. はじめに

近年の深層学習の発展により、コンピュータビジョンの分野では CNN を用いた手法がさまざまなタスクに応用されている。その中でもインスタンスセグメンテーション (IS : Instance Segmentation) は同種物体も区別してピクセル単位で抽出を行う、基本的なタスクである。IS が可視領域を分割するのに対し、アモダルインスタンスセグメンテーション (AIS : Amodal Instance Segmentation) は可視領域と隠蔽領域の両方を抽出することを目的とするため IS よりも難易度が高い。Mask R-CNN[1]は IS タスクにおいて優れた結果を示したが、隠蔽領域も含めた正解を与えて学習することで AIS でも利用できる[2]。

Mask R-CNN を含む CNN の問題点として、実効的な受容野が限られるため、長距離の関係性をモデル化しにくい点が挙げられる。このため、線状の物体のような空間的に遠い領域が同じオブジェクトに属するような物体の抽出には向いていないと考える。また、Mask R-CNN は、最初に物体候補を含む注目領域を矩形で抽出し、次に矩形内の物体領域をピクセル単位でセグメントするという 2 つのステップからなる。このため、注目領域の抽出精度がセグメンテーション結果に大きな影響を与える問題もある。

そこで、抽出対象の一部を指定し、対象を逐次的に抽出する AIS 手法を提案する。本手法では、注目領域の抽出が不要になるとともに、抽出対象をたどるように抽出することで空間的に離れていても同領域と判断できる利点がある。人工的な線状物体に本手法を適用した結果、Mask R-CNN よりも高精度な結果が得られた。

2. 手法

提案する AIS 手法 (シード法) は抽出対象をシード (種領域) で指定する手法である。物体のクラス数に制限はないが、今回は 1 クラスの線状物体が多数存在する画像に適用した。手法は以下の 4 ステップからなる。

1. 全領域の抽出

原画像からインスタンスを区別せず、全ての線状物体の領域を CNN₁ で抽出する。これをシード候補領域とする。

2. シードで指定した対象の抽出

シード候補領域に点状の小領域を設定しシードとする。そのシードを白、その他を黒とするシード画像を作成する。シード画像を原画像とチャンネル方向に結合して CNN₂ に入力し、シードを含む線状物体領域を抽出する。CNN₂ の出力を 2 値化、膨張、シードを含まない領域を削除、収縮し、仮抽出画像を得る。さらに仮抽出画像を処理して線状のシード画像を作成し、再度抽出を行い抽出画像を得る。

3. シード候補領域から抽出された対象領域を削除

抽出された対象領域をシード候補領域から削除する。

4. シード候補がなくなるまで 2, 3 を反復

シード候補領域がなくなるまで 2, 3 を繰り返し、全ての物体を抽出する。

このように抽出対象をシードで指定して抽出する。さらに、抽出結果から新たなシードを作成して抽出を繰り返すことで抽出領域をたどるように拡張し、受容野を無理に大きくすることなく線状物体を抽出する。

3. 実験

3.1 データ

実験に用いるデータは線状物体を楕円の一部分を用いて模擬的に表現して作成した。描画する弧の指定の方法を図 1 に、作成した画像の例を図 2 に示す。512×512[画素]の画像に、ランダムな長短径 (30~512[画素])、回転角度 (-90~90[度])、開始角度 (0~90[度])、弧の角度 (45~180[度]) で指定される楕円の弧をランダムな位置に描画する。これを 4~16[個]描画し、中央の 256×256[画素]を切り取ったものを線状物体画像とした。

点状のシード画像は物体領域からランダムな画素を選択して作成した。線状のシード画像は物体領域のマスク画像を細線化したものを短くして作成した。

実験に用いた画像は 12,500 枚 (学習用 8,000 枚, 検証用 2,000 枚, 評価用 2,500 枚) であり、1 画像あたりの平均物体数は 10.0[個]、分散は 4.0[個]である。

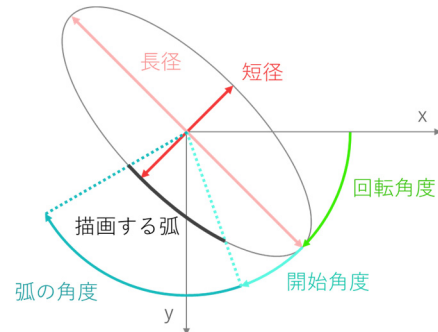


図 1 描画する楕円弧の指定方法



図 2 人工の線状物体画像の例
(左から 4, 8, 12, 16 個)

3.2 実装

実験用データは人工的に作成したものであるため、手順 1 の全領域の抽出は不要であり、原画像を白黒反転したものをシード候補領域として用いた。線状物体の抽出には Attention U-Net[3]を用いた。Attention U-Net はセグメンテーションで優れた結果を示した U-Net に対し、Channel Attention, Spatial Attention を取り入れることで精度向上さ

[†] 芝浦工業大学 Shibaura Institute of Technology

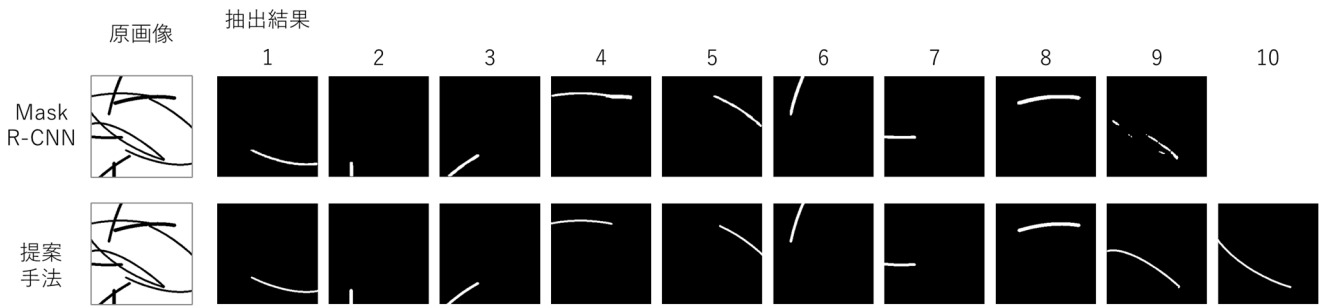


図 3 抽出結果の例

せたモデルである．入力は白黒の原画像とシード画像の 2 チャンネルの 256×256 [画素] の画像．出力は同サイズの抽出された線状物体領域の 1 チャンネルの画像である．その他のパラメータは原論文[3]のものをそのまま用いた．学習は Optimizer に Adam, 学習率は初期 10^{-4} で, 50 [epoch] ごとに半分にして 400 [epoch] 学習した．なお, シード画像は点状のものと同様のものがあるため, CNN₂ として点状, 線状それぞれのシード画像を用いて抽出する 2 種類を使い分けられることも考えられるが, 今回の実験では両方のシード画像を 50% ずつ利用して学習した CNN₂ を用いた．

比較対象には Mask R-CNN を用いた．入力は 256×256 [画素] の白黒の原画像で, 出力は各矩形についてセグメンテーションを行った結果である $256 \times 256 \times$ オブジェクト数 [画素] となるように学習した．学習には Optimizer に Adam を用い, 学習率 10^{-5} で 500 [epoch] 学習した．

3.3 評価

評価には AIS タスクの評価指標に一般的に用いられる Mean Average Precision (mAP) を用いた．正解領域, 抽出領域それぞれを含む最小の矩形を求め, それらの IoU が一定以上ならば抽出成功とする．画像それぞれについて Precision, Recall を求め, 横軸を Recall, 縦軸を Precision とする PR 曲線を引くと, AP は以下の式で定義される．

$$AP = \int_0^1 p(r) dr$$

これを全クラスについて平均したものが mAP だが, 本研究ではクラス数は一つのため単なる AP で評価を行った．IoU 閾値が 0.50, 0.75 の場合と, [0.50, 0.95] の間を 0.05 ずつステップさせて平均をとった場合をそれぞれ AP50, AP75, AP で表す．

4. 結果と考察

抽出結果の例を図 3 に, 線状物体密度を 6.0~14.0 [個/画像] の間で変えた場合の AP を図 4 に, 密度 10.0 [個/画像] の場合の AP, AP50, AP75 を図 5 に示す．

図 4 より, 線状物体密度によらず提案手法により AP は改善されることが分かる．特に密度が高いほど改善幅が大きい．また, 図 5 より, AP50 は Mask R-CNN とほとんど変わらないが, AP75 は改善され, AP では改善幅がさらに大きいことが分かる．これは IoU の大ききなところほど提案手法の精度が高いことを示している．本手法は物体を含む矩形領域を抽出する手法ではないが, 結果的には線状物体を含む矩形領域を Mask R-CNN より正確に推定できることを示している．

図 3 の抽出結果を見ると, Mask R-CNN では 2 つの物体

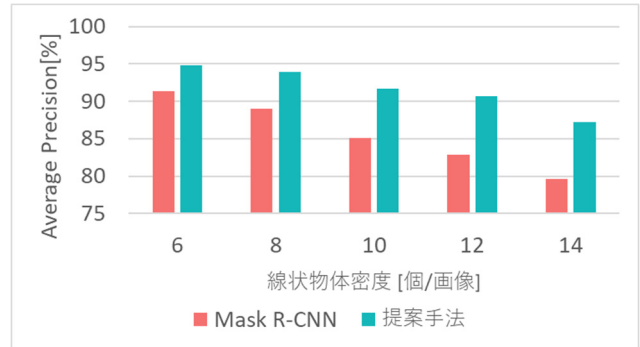


図 4 物体密度ごとの精度 (AP)

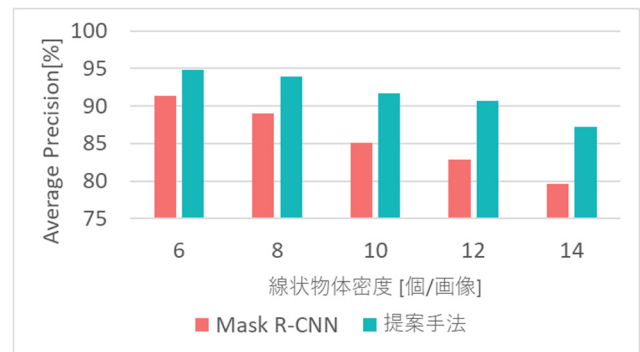


図 5 精度評価結果 (10.0 [個/画像])

を 1 つの物体として誤抽出している (抽出結果 4) . また, 長い物体が 2 本近接する場合も抽出できていないが (抽出結果 9) , 提案手法では何れも正しく抽出できている (抽出結果 9, 10) .

5. おわりに

本研究では AIS タスクにおいて, 抽出対象の一部をシードで指定し, そこから抽出領域をたどるように抽出するシード法を提案した．これにより空間的に離れていても同領域と判断できるようになり, また矩形で注目領域を抽出する処理も不要となった．人工的な線状物体を用いた実験で有効性を確認できた．今後は頭髮など実際の線状物体へ適用する予定である．

参考文献

- [1] Kaiming He, et al., "Mask R-CNN", Proceedings of the IEEE International Conference on Computer Vision, pp. 2961-2969, (2017).
- [2] Lu Qi, et al., "Amodal Instance Segmentation with KINS Dataset", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3014-3023, (2019).
- [3] Ozan Oktay, et al., "Attention U-Net: Learning Where to Look for the Pancreas", arXiv preprint arXiv:1804.03999, (2018).