

[研究助成 (C)]

人間からの補助情報を利用した学習なしセマンティック セグメンテーション手法の開発

Development of a learning-free semantic segmentation method using auxiliary information from humans

2237002



研究代表者

株式会社センスタイムジャパン リサーチャー 加藤 聡 太
(助成金受領時：名城大学 大学院理工学研究科 博士課程)

[研究の目的]

画像の各ピクセルにクラスラベルを割り当てるセマンティックセグメンテーション（以下セグメンテーション）は、医療分野や細胞生物学分野の画像解析において重要な技術である。従来のセグメンテーションは手動で行われていたため、主観的な結果になりがちであったが、近年では深層学習を用いることにより客観的な結果を自動的に得ることが可能となり、これまでに多くの手法が提案されている。

しかし、新たな画像データに適用するための深層学習モデルを作成する場合、再度モデルの学習を行う必要があり、大量の入力画像と教師画像が必要となる。特に教師画像の作成は人間の専門家の知見が必要であり、多くの時間とコストが発生する。そこで近年、より少ない学習データで効率的な学習を行う Few-shot セグメンテーション、学習を行わず推論のみを行う Zero-shot セグメンテーション、少ない教師画像で学習を行う半教師ありセグメンテーションが提案されている。また近年、Prompt learning という考え方が自然言語処理分野において広まっている。これは大規模な事前学習済みモデルを利用した推論手法であり、いくつかのパターンをモデルに与えることにより、画像認識分野の Few-shot, Zero-shot 学習において従来よりも高い精度を達成することが可能であることが報告されている。

しかし、犬や猫などの一般的な画像データで学習された事前学習済みモデルを異分野の画像データに適用する場合、データのドメインが異なるため、うまく適応できない可能性が考えられる。また Prompt learning を始めとする、近年の few-shot, zero-shot learning 法は、画像情報とテキスト情報の両方を用いることが多いが、異分野データの場合、専門的な用語を扱うため、一般的な言語モデルでは対処しきれない可能性も考えられる。異分野データを用いて Prompt learning を行うためには、よりその分野に特化した学習済みモデルの使用が必要であると考えられる。

そこで本研究では、あらかじめ別の細胞画像データセットを用いた事前学習を行い、学習済みモデルを新たに作成する。さらにこの学習済みモデルを未知の細胞画像データセットに適用することにより、図1に示すような One-shot セグメンテーション法を提案する。また、細胞

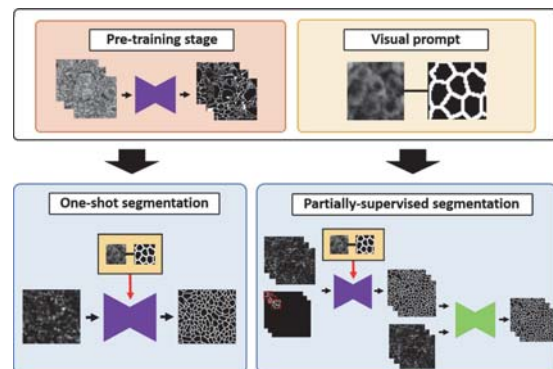


図1 提案手法の概要

画像は、同じ構造が画像内に広がるフラクタルな形状を持つ場合が多い。これを利用し、画像内の一部のアノテーションのみで学習可能な新たな半教師ありセグメンテーション手法を提案する。

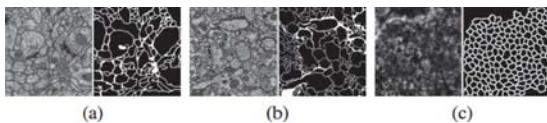
評価実験では、形状の異なる3種類の細胞画像データセットを用いた検証実験の結果、事前学習モデルの有効性と、One-shot セグメンテーションにおける提案手法の有効性を確認した。また提案した半教師あり学習により作成されたモデルを用いて疑似ラベルを作成し、それを用いて学習させた検証の結果、オリジナルデータセットを全て用いた時との精度の差は、全てのデータセットで約5%以内であり、一部のアノテーション付与で十分な教師ラベルを作成できることを確認した。

[研究の内容, 成果]

1. 提案手法

1.1 事前学習モデルの作成

まず初めに細胞画像のみを用いた学習済みモデルを作成する。本研究で扱うデータセットは図2で示すiRPE, ssTEM, ISBI2012の3種類である。すべてのデータセットは細胞膜とそれ以外を分割する2クラスセグメンテーションであり、推論対象のデータセット以外のデータセット2種類を用いて事前学習モデルを作成し、合計3種類の事前学習済みデータセットを作成する。One-shot セグメンテーションや半教師ありセグメンテーションのネットワークは、対象のデータセットが含まれていない学習済みモデルを使用する。



(a) ISBI2012, (b) ssTEM, (c) iRPE

図2 データセット例

1.2 One-shot セグメンテーション

図3にOne-shot セグメンテーションにおける提案手法の概要を示す。入力画像および入力画像とは別の小領域のサンプル画像をそれぞれセグメンテーションネットワークに入力する。ネットワークの重みは共有されており、出力されたそれぞれの特徴マップを内積し、ソフトマックス関数を用いて正規化されたアテンションマップを作成する。セグメンテーション対象の入力画像の最終出力の特徴 x_i と、サンプル画像の最終出力 x_s を用いて作成されるアテンションマップ β_{ij} は、j番目の領域がi番目の位置にどの程度関与しているかを表している。その後、サンプル画像に対応するone-hotの教師ラベル画像 y_s を用いると、最終出力は $o_j = y_s \beta$ で表すことが出来る。これにより、アテンションマップでは入力画像の特徴とサンプル画像の特徴の類似度を獲得することができ、アテンションマップとサンプル画像の教師ラベル情報を用いて、クラス数分の特徴マップを含む、最終的なセグメンテーション画像を作成することが可能となる。

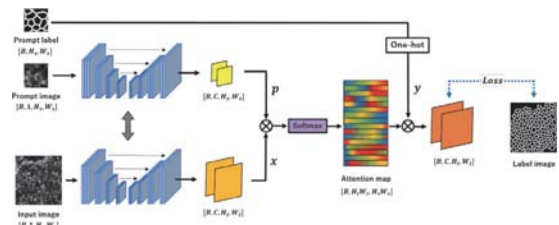


図3 One-shot セグメンテーションの概要

提案手法を用いることで、サンプル画像とそれに対応した教師ラベルの情報を学習と推論両方で活用することができるため、少ない画像データでも精度の高いモデルが作成出来ると考えられる。

1.3 半教師ありセグメンテーション

図4に半教師ありセグメンテーションにおける提案手法の概要を示す。半教師ありセグメンテーションでは、入力画像内で教師ラベルが与えられている箇所が一部であることを仮定し、

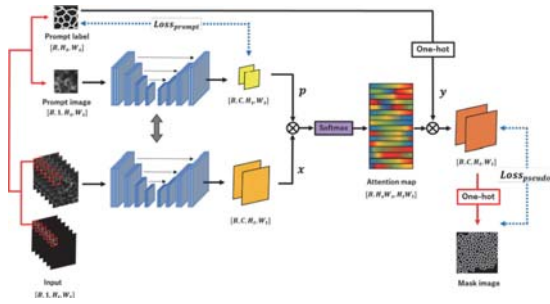


図4 半教師あり学習の概要

該当箇所をクロップした画像をサンプル画像として扱い、3.2節と同じようにセグメンテーションネットワークに入力する。3.2節と同様に、アテンション構造を用いてサンプル画像に対応する教師ラベルを参照したセグメンテーション出力を獲得する。その後 argmax 関数を用いて出力画像を疑似ラベル化し、疑似ラベルを用いて学習する。またサンプル画像の出力に対しては通常の教師画像を用いて学習する。

提案手法を用いることにより、一部しか教師が与えられていない場合でも、アテンション構造により画像全体にサンプル画像の情報を広げることが可能であり、画像全体のセグメンテーションが可能になる。

2. 評価実験

2.1 データセット

実験では Absorbance microscopy images of human iRPE cells (iRPE), Serial sectioning transmission electron microscopy (ssTEM), ISBI2012 を使用した。iRPE の画像枚数は 1,032 枚であり、画素数は 256×256 画素である。教師画像には背景 (background) と細胞膜 (membrane) の 2 種類のクラスが含まれている。ssTEM の画像枚数は 240 枚であり、画素数は 256×256 画素である。教師画像には背景 (background) と細胞膜 (membrane) の 2 種類のクラスが含まれている。ISBI2012 の画像枚数は 1,032 枚であり、画素数は 256×256 画素である。教師画像には背景 (background) と細胞膜 (membrane) の 2 種類のクラスが含

まれている。

2.2 実験結果

図 5 に 3 つのデータセットに対する one-shot セグメンテーションの精度比較を示す。赤字の数値は最も良い精度を示している。提案手法を用いた結果、iRPE で 4.06%, ssTEM で 4.20%, ISBI2012 で 2.41% の平均 DSC の精度向上が確認できた。また学習済みモデルを使用した場合、iRPE で 4.60%, ssTEM で 2.42%, ISBI2012 で 3.47% の精度向上を確認した。また図 7(c), (d) にセグメンテーション画像の結果を示す。提案手法を用いることにより、クラスの境界箇所が鮮明に表れることが分かる。

以上の結果から、提案手法を用いた One-shot セグメンテーションと、事前学習済みモデルの有効性を確認することができた。

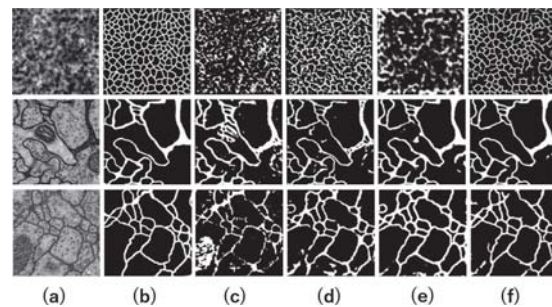
図 6 に 3 つのデータセットに対する半教師ありセグメンテーションの精度比較を示す。提案手法を用いた結果、iRPE で 3.24%, ssTEM で

Dice score	iRPE			ssTEM			ISBI2012		
	Average	Background	Membrane	Average	Background	Membrane	Average	Background	Membrane
Original learning									
Baseline	74.39 ± 0.15	84.60 ± 0.14	64.17 ± 0.07	91.60 ± 0.14	96.54 ± 0.05	86.65 ± 0.15	86.66 ± 0.10	94.04 ± 0.15	79.28 ± 0.08
One-shot learning									
Baseline	59.25 ± 0.08	76.96 ± 0.04	39.53 ± 0.04	78.92 ± 0.14	92.56 ± 0.15	69.28 ± 0.07	77.82 ± 0.18	90.06 ± 0.15	65.57 ± 0.08
Ours	63.31 ± 0.10	78.48 ± 0.06	43.14 ± 0.04	83.12 ± 0.04	92.96 ± 0.04	80.23 ± 0.10	80.99 ± 0.04	93.47 ± 0.05	74.06 ± 0.06
Baseline (pre-trained)	59.25 ± 0.08	76.96 ± 0.04	39.53 ± 0.04	82.42 ± 0.15	92.71 ± 0.07	72.13 ± 0.07	77.54 ± 0.07	90.21 ± 0.07	64.86 ± 0.10
Ours (pre-trained)	63.89 ± 0.08	78.56 ± 0.05	45.15 ± 0.03	84.84 ± 0.15	93.46 ± 0.06	78.23 ± 0.04	81.01 ± 0.05	91.91 ± 0.05	70.91 ± 0.09

図5 One-shot セグメンテーションの精度比較

Dice score	iRPE			ssTEM			ISBI2012		
	Average	Background	Membrane	Average	Background	Membrane	Average	Background	Membrane
Original learning									
Baseline	74.39 ± 0.15	84.60 ± 0.14	64.17 ± 0.07	91.60 ± 0.14	96.54 ± 0.05	86.65 ± 0.15	86.66 ± 0.10	94.04 ± 0.15	79.28 ± 0.08
Semi-supervised learning									
Baseline	53.47 ± 0.06	76.77 ± 0.06	31.18 ± 0.10	72.22 ± 0.10	88.92 ± 0.07	55.52 ± 0.06	76.19 ± 0.06	84.86 ± 0.06	57.52 ± 0.06
Ours	56.71 ± 0.17	77.49 ± 0.06	36.62 ± 0.06	85.71 ± 0.10	94.25 ± 0.06	77.17 ± 0.05	83.47 ± 0.07	93.28 ± 0.07	74.06 ± 0.06
Baseline (pre-trained)	53.47 ± 0.06	76.77 ± 0.06	31.18 ± 0.10	72.22 ± 0.10	88.92 ± 0.07	55.52 ± 0.06	76.19 ± 0.06	84.86 ± 0.06	57.52 ± 0.06
Ours (pre-trained)	67.99 ± 0.10	82.87 ± 0.06	53.40 ± 0.04	86.06 ± 0.06	94.42 ± 0.05	77.71 ± 0.05	82.45 ± 0.05	92.77 ± 0.05	72.13 ± 0.10

図6 半教師ありセグメンテーションの精度比較



すべて事前学習済みモデルを用いた場合である。上から iRPE, ssTEM, ISBI2012。(a) 入力画像、(b) 正解画像、(c) One-shot セグメンテーション (U-Net)、(d) One-shot セグメンテーション (提案手法)、(e) 半教師ありセグメンテーション (U-Net)、(f) 半教師ありセグメンテーション (提案手法)

図7 セグメンテーション画像の比較

Dice score	iRPE cell	ssTEM	ISBI2012
Original learning	74.39 $\pm$ 0.19	91.60 $\pm$ 0.14	86.66 $\pm$ 0.50
Ours	73.30 $\pm$ 0.25	87.45 $\pm$ 0.19	85.19 $\pm$ 0.08

図8 疑似ラベルを用いた実験結果

13.49%, ISBI2012 で 7.48% の平均 DSC の精度向上が確認できた。また学習済みモデルを使用した場合, iRPE で 14.52%, ssTEM で 13.84%, ISBI2012 で 6.26% の精度向上を確認した。図 7(e), (f) にセグメンテーション画像の結果を示す。図 7(e) から, 一部の小領域の教師ラベルのみを用いた学習では画像全体の予測が困難であることが分かるが, 提案手法を用いることにより, (f) のように画像全体の予測が可能である。また図 8 に提案手法により作成した疑似ラベルによる学習結果を示す。オリジナルデータセットを用いた学習と比較した結果, ssTEM で 4.15%, ISBI2012 で 1.47% の精度低下であり, 疑似ラベルを用いた学習でも十分学習可能であることが分かった。以上の結果から, 提案手法を用いた半教師あり学習の有効性を確認することができた。

[まとめと今後の展望]

本研究では, 細胞画像セグメンテーションのための新たな One-shot セグメンテーション手

法と半教師ありセグメンテーション手法を提案した。3 種類の細胞画像データセットを用いた検証実験の結果, 提案手法を用いることにより, 少ない画像枚数や少ないアノテーションでも精度の高いモデルを作成することが可能であり, また別の細胞画像で学習した学習済みモデルは有効であることを確認した。今後はサンプル画像をさらに増やすことでさらなる精度向上を考えている。また多種類の細胞画像データを学習させ, より高精度かつドメインが大きく異なるデータに対しても適用できるような学習済みモデルの作成を考えている。これにより, 学習を一切行わない Zero-shot セグメンテーションに対しても有効性を示すことが出来る可能性がある。

[成果の発表, 論文など]

1. 加藤聡太, 堀田一弘, “事前学習済みモデルを用いた細胞画像に対する One-shot/Semi-supervised Semantic Segmentation”, パターン認識・メディア理解研究会 (PRMU), 2023 年 03 月, 北海道函館市.
2. Sota Kato, Kazuhiro Hotta, “One-shot and Partially-Supervised Cell Image Segmentation Using Small Visual Prompt”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshop (CVMI), pp.4294-4303, June, 2023, Vancouver (Oral).