

研究終了報告書

「深層学習の関連タスク学習能力を活用したバイオ画像認識手法の開発」

研究期間：2021年10月～2024年3月

研究者：西村 和也

1. 研究のねらい

深層学習に基づいた画像認識手法の台頭により、一細胞単位の画像認識が実現されつつある。しかし、深層学習モデルを学習するために十分な学習データが必要不可欠であり、学習データの作成コストが高いことが問題となっている。学習データの作成コストを削減するため、細胞画像の特性を用いて簡易に取得可能なメタデータに着目した。例えば、細胞画像を撮影する際に、撮影時刻の情報や実験に使用した細胞種情報は記録されている。これらの簡易取得可能なメタデータは画像認識の直接的な正解ラベルとして使うことはできないが、目標の画像認識タスクの手がかりとなる情報を有していると考えられる。本研究は、このメタデータと人手のアノテーションを組み合わせ、深層学習モデルを学習する方法を検討し、学習データの作成コストが低い画像認識手法の開発を目指す。

2. 研究成果

(1) 概要

本研究期間においては主に3つの試みを行った。まず、2種類の簡易取得可能なメタデータを用いた画像認識に着手した。メタデータは画像認識の直接的な正解を示さないため、単一データでは画像認識に必要な情報が不足する。そこで、複数のメタデータを組み合わせ、学習する手法を開発し、効果的に2種類のラベルを組み合わせることに成功した。次に、人手のアノテーションとメタデータの組み合わせる方法を検討した。メタデータを使う前提で、人手のアノテーション方法を設計することで、最小限のアノテーションコストで高精度な画像認識が可能となることを実証した。最後の試みは応用分野への展開である。データに含まれる簡易に取得可能なメタデータは細胞画像以外のドメインでも存在する。そこで、自身のメタデータを用いた学習技術を医療データおよび天体画像の認識への拡張を試みた。

(2) 詳細

研究テーマA「2種類の簡易ラベルを用いた画像認識」

本テーマでは2種類の簡易取得可能なラベルを用いた細胞形状認識に着手した(図1)。位相差顕微鏡を用いた生細胞の観察では、2種類のラベルは、画像認識のための追加のアノテーションコストなしで取得可能である。第一のラベルは細胞種のラベルであり、細胞を撮影する際に使用した細胞種は記録されているので簡易に取得可能である(図1(a))。第二のラベルは、細胞位置ラベルであり、核染色した細胞を蛍光像と位相差像で撮影することで自動取得可能である(図1(b))。

2種類のラベルからそれぞれ、前景領域(細胞領域)および細胞位置を認識することが可能であるが、細胞形状に必要な細胞間の境界は取得できない。細胞形状認識には個々の細胞形状を認識することが必要である。そこで、図1(c)に示すように単一細胞を探し、その細胞領

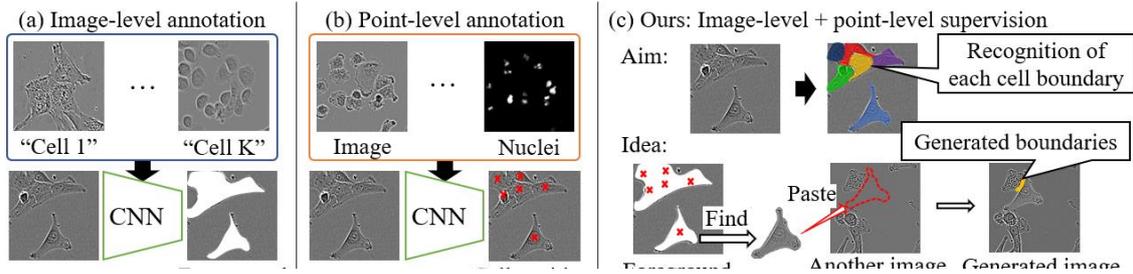


図 1 (a) 細胞種ラベル、(b) 細胞位置ラベル、(c) 提案手法の概要

域を張り付けるという手法を開発し細胞境界のラベルを擬似的に作成することに成功した。作成したラベルを用いて細胞境界認識モデルを学習することにより細胞形状認識を行った。

実験では細胞画像の公開データセット(LIVECell dataset)を用いて、提案手法と従来手法の比較実験を実施した。その結果、教師ありの手法には劣るが従来の簡易取得可能なラベルを用いた手法と比較し高精度に細胞形状を認識可能であることを確認した。

研究テーマ B 「簡易ラベルとアノテーションを組み合わせた画像認識」

本テーマでは、人手のアノテーションと簡易に取得可能なメタデータを用いた細胞分裂検出に着手した。メタデータを使うことを前提として、人手のアノテーションを設計することにより、従来よりも少ないアノテーションコストで高精度な画像認識が実現できることを実証した。

具体的には、partial annotation と画像の撮影タイミングを用いて細胞分裂を行った。細胞分裂は時系列のイベントであるため、複数の画像の差分を踏まえたアノテーションが必要である。そのため、注意深く確認を行わなければ、多くの見逃しが含まれやすい。そこで、網羅的な supervised annotation の代わりに partial annotation の使用を考えた。Partial annotation は図 2 に示すように、時系列中のいくつかの分裂イベントにのみアノテーションを付与されており、残りの分裂にはアノテーションを付与しないという方法である。アノテーターはいくつかの分裂イベントに見逃しを気にせずアノテーションを行えば良いため、アノテーションコストは劇的に削減することが可能である。

一方、supervised annotation と違い、全ての分裂にラベルが付与されていないので、分裂しないというラベルがないという問題が生じる。Supervised annotation ではアノテーションが付与されていない領域を細胞分裂が含まれない領域として扱うことができたが、partial annotation ではアノテーションが付与されていない領域は unlabeled data として扱う必要がある。また、そもそもアノテーションの数が少ないため教師ありの条件に比べて与えられている情報が少ない。この二つの問題を解決するため partial annotation から supervised データセットを作成する

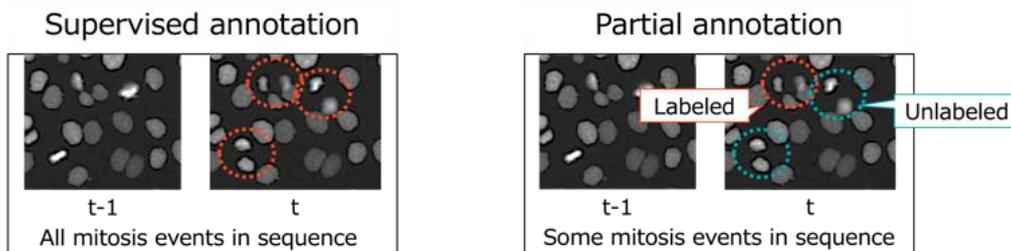


図 2 Supervised annotation と partial annotation

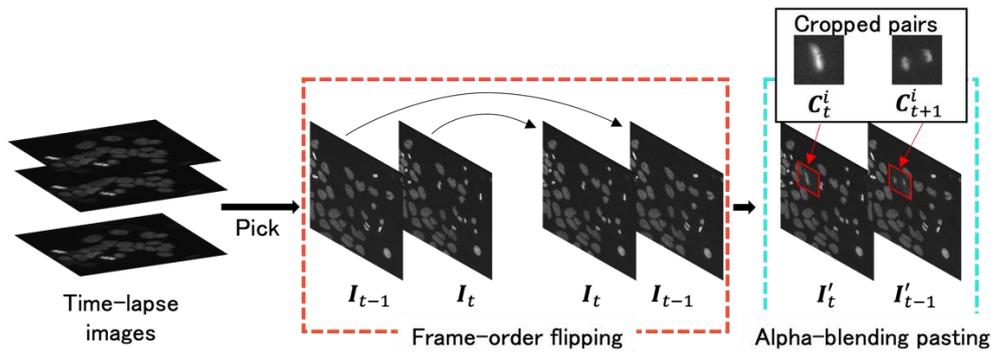


図 3 Dataset generation

手法を提案した(図3)。まず、撮影時に記録されている撮影タイミングを用いて frame-order flipping という撮影時間の順序を反転するという手法を提案した。この手法により unlabeled だった領域は細胞分裂が含まれない領域に変換することが可能である。次に、alpha-blending pasting という自然に分裂ラベルを付与する手法を提案し、データセットの作成に成功した。

実験では、4種類のデータセットを用いて提案手法の有効性の検証を行い、従来手法を上回る性能を確認した。さらに supervised annotation の条件でも提案手法を適用し、supervised annotation の際のデータ拡張としても提案手法が有効であることを確認した。

研究テーマ C 「応用研究」

本テーマでは、メタデータを用いた研究を天体データ、医療データへの拡張を検討した。

天体画像認識では、ETA detection (地球のトロヤ群の検出)に着手した。この課題は、ノイズの多い画像から小惑星を検出するタスクであり、研究者が着手してきた細胞検出に類似する課題である。研究者は天体画像の専門ではないため、ローレンスリバモア国立研究所へ技術調査に行き、宇宙物理学の研究者と技術交流を実施した。天体画像に人が正解ラベルをするのは難しいため、人が正解ラベルを作成する代わりに合成データを用いて深層学習モデルを学習ことにより、ETA detection を実現できないかを検討した。

天体画像解析では、シミュレーション技術が発展しており、シミュレーションを駆使することで検出の手がかりとなる情報が取得可能である。実際の小惑星の軌道をシミュレーションすることで検出の合成データの作成を行い、深層学習モデルで小惑星検出を行った。実験の結果、シミュレーションデータでは約 95%の検出ができたが、リアルデータでは過検出が生じることが確認された。これは、深層学習がノイズに弱いことに起因しており、シミュレーションデータとリアルデータのドメインギャップによりリアルデータでの性能が低下したと考えられる。

医療データへの拡張では、PDL1 発現の推定を実施した。PDL1 診断は化学療法などに使用される指標であり、臨床現場でも使用されている指標である。病理画像中のがん細胞の中で PDL1 陽性細胞と陰性細胞のうち陽性細胞の比率を推定する課題である。

画像認識で実現するためには、数十万を超える細胞に対して非がん細胞、陽性細胞、陰性細胞にアノテーションを付与し、3class の検出を行い、比率を算出することである。しかし、細胞検出のラベルを付与するのはコストが高く現実的ではない。そこで、臨床現場で付与されている医師が付与する概算の PDL1 expression rate と少量の細胞検出ラベルを用いた、PDL1

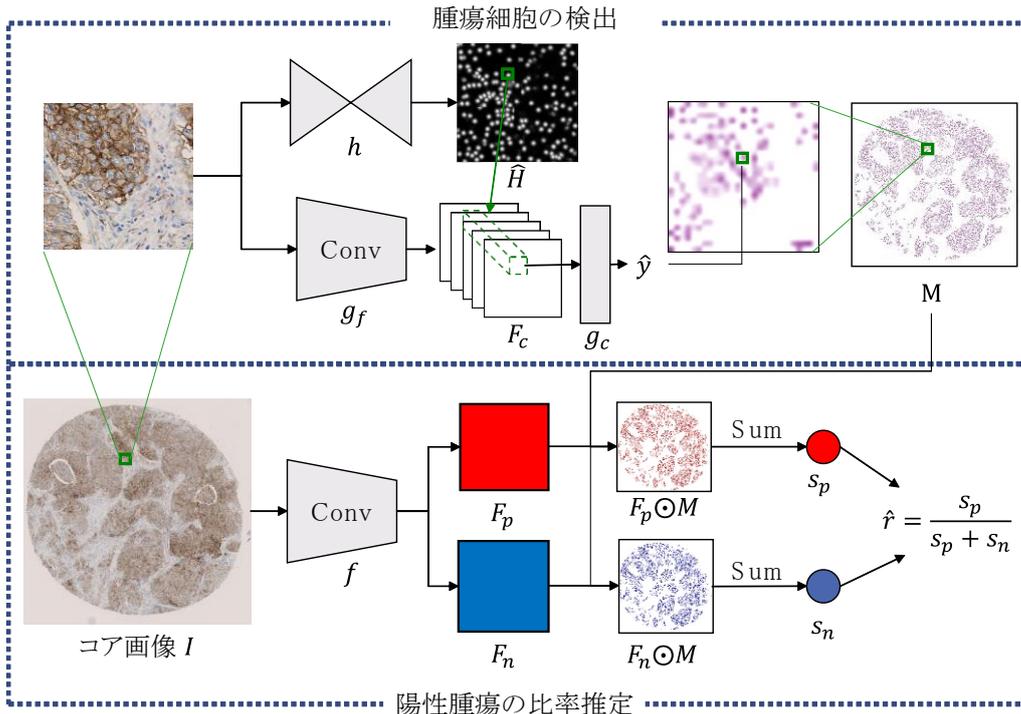


図 4 陽性腫瘍比率推定手法

expression の推定手法を提案する。図 4 に示すように PD-L1 rate が腫瘍細胞中の割合で決定されるため、細胞検出の結果を用いて特徴マップを mask することにより、腫瘍細胞に関する割合の推定を実現する。実験では、解釈性及び性能の面で従来手法より優れていることを確認した。

3. 今後の展開

本研究では、データに付随しているメタデータを画像認識に活用することにより、人手の Annotation コストの少ない画像認識の提案を行い、細胞画像に関してはいくつかの成果をあげること成功した。この研究を通して、研究者のメタデータを用いた学習技術は、正解データの作成が困難なデータ解析においてより需要があると考えた。そこで今後は医療データへの展開を考えている。

まず、臨床データでは画像に対して、診断ラベル、撮影機器、患者情報などの多数のメタデータが付与されている。これらのメタデータを活用した画像認識技術は開拓する余地が残っており、研究者の技術が必要になる分野だと考えている。医療分野では信頼性や性能の面から AI の活用が進んでいない分野であるため、医療データにおいて AI 活用を推進できたらと考えている。

さらに、近年空間解像度の高い遺伝子発現データが画像と同時に撮影可能となった。新しく収集されたデータを駆使して新しい発見ができないかを検討予定である。

4. 自己評価

研究目的の達成状況・研究の進め方:

研究目標である、応用分野に存在する簡易に取得可能なメタデータを活用した画像認識に関して、想定通りいくつかの成果を上げることができた。研究の進め方に関しては、分野が目まぐるしいスピードで進展していることもあり、方向性に迷うこともあったが、計画通り応用分野への展開を検討できるフェーズに差し掛かっている。研究費執行状況はコロナの時制もあり、旅費の見積もりに誤差がでたが、概ね順調に執行できた。

研究成果の科学技術及び社会・経済への波及効果:

研究成果に関しては code を公開することにより、分野の進展に寄与することができている。経済への波及効果に関しては研究段階の技術であり、経済活動への直接的な寄与は現状では目処がついていない。

研究者ネットワークの形成に関して:

ACT-X の領域会議を足がかりに、自身の研究を飛躍させるためのネットワーキングができた。遺伝子発現データを扱っている研究者と共同研究のプロジェクトが持ち上がり自身のキャリアパスにも寄与するネットワークを構築できた。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 4件

1. [Kazuya Nishimura](#), Ryoma Bise, “Weakly Supervised Cell-Instance Segmentation with Two Types of Weak Labels by Single Instance Pasting”, In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, 3185–3194

深層学習に基づく手法が高性能な細胞形状認識を達成しているが、撮影環境毎に学習データが必要である。そこで、本研究では簡易に取得可能な二種類のラベル(細胞核の位置と細胞種ラベル)を用いた、細胞形状認識手法を提案する。二種類のラベルは細胞形状認識の手がかりを持つが、細胞間の境界を示す情報が不足する。そこで、細胞間の境界ラベルを self-supervised で作成する single instance pasting を提案し、従来手法より優れた細胞形状認識手法を提案した。

2. [Kazuya Nishimura](#), Ami Katanaya, Shinichiro Chunma, Ryoma Bise, “Mitosis Detection from Partial Annotation by Dataset Generation via Frame-Order Flipping”, In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, 2023

学習データが十分にあれば深層学習により高精度な細胞分裂検出が実現できる。しかし、動画中の全ての分裂した細胞にラベルを付与することは時間がかかり、見逃しも発生する。そこで本研究では、少数の部分的な分裂位置ラベルと撮影時刻の情報を用いた細胞分裂検出手法を提案する。部分的な分裂位置ラベルを用いる場合、分裂しない画素を示すラベルが不足する。そこで、分裂しない画素を示すラベルを self-supervised で作成する frame-order-flipping を提案し、高精度な分裂検出を実現した。

3. [Takumi Okuo](#), [Kazuya Nishimura](#), Hiroaki Ito, Kazuhiro Terada, Akihiko Yoshizawa, Ryoma Bise, “Proportion Estimation by Masked Learning from Label Proportion”, In Proceedings of

the MICCAI workshop on DALI, 2023

PD-L1 rate の推定は、腫瘍細胞中の陽性細胞の割合を推定するタスクである。本研究では、少数の細胞検出ラベルと比率ラベルを用いて、アノテーションコストの低い PD-L1 rate estimation method を提案する。PD-L1 rate が腫瘍細胞中の割合で決定されるため、細胞検出の結果を用いて特徴マップを mask することにより、腫瘍細胞に関する割合の推定を実現する。実験では、解釈性及び性能の面で従来手法より優れていることを確認した。

(2)特許出願

研究期間全出願件数： 0 件(特許公開前のものは件数にのみ含む)

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

学会発表:4件