

時系列ラベルを用いた弱教師あり学習による 筋組織再生過程の定量化手法の提案

梶村 渉¹ 瀬尾 茂人^{1,a)} 深田 宗一郎² 松田 秀雄¹

概要：筋組織細胞は損傷から再生までの過程で様々な様相を呈する。この筋再生過程の解析は、老化や筋疾患による筋機能研究などへの貢献が期待される。筋組織の再生度合いを評価する方法としては組織切片を撮影した画像から外観検査を行うことが一般的であるが、客観性や定量性の担保といった側面から、深層学習などを用いた画像処理による解析の要求が高まっている。一方で、筋組織の再生についてのアノテーションには専門知識が必要であり、またそのコストも大きい。十分な量と質の学習データを確保することは困難である。そこで本研究では、細胞画像撮影時点で取得可能な時系列ラベルを用いた弱教師あり学習によって、筋組織再生過程の定量化を行う方法を提案する。

キーワード：筋組織画像、深層学習、畳み込みニューラルネットワーク、弱教師あり学習

1. はじめに

近年のバイオイメージング技術 [1] の急速な発展に伴い、多種多様な生体の画像データを入手することが可能になった。この生体画像データは生体やその内部組織を顕微鏡によって選択的に撮影した情報を含んでおり、データ解析を施すことで生命現象に関する諸問題の解明に役立てることが出来ると期待されている。生体画像解析の対象は様々であるが、本研究では筋組織画像に着目する。

筋組織は生体内で張力を発生させるために分化した組織であり、常態では有糸分裂を行わない安定した細胞の集合である。ところが筋組織細胞に損傷が起こると活発的な再生活動を行う細胞へと変化する。このように一般的な細胞とは異なる構造と機能を有しており、その変成や再生過程の解析を行うことで、老化や筋疾患による筋機能低下のメカニズムやその治療法の開発などの研究に役立てることが可能である。筋組織再生過程の研究には古くから遺伝子発現量解析による手法が用いられてきたが、実験に必要なコストの面から画像処理技術を用いた解析が注目されている。細胞画像解析には畳み込みニューラルネットワークを使用した深層学習（ディープラーニング）による手法で高

精度な解析が行えることが知られており、筋組織細胞画像に対しても同様であると考えられるが、筋組織再生過程の定量化においては未だに適切な手法が確立しているとはいえない。

畳み込みニューラルネットワークを用いて画像解析を行う場合、学習データとして大量の画像を用意する必要がある。学習データには筋組織細胞画像と、その再生段階に応じた領域ごとの色の塗分け（ラベル付け）を行った画像の2枚をセットで与える。このラベル付けは専門家が一つ一つ手作業で行うことしか出来ない。特に、筋再生過程の程度の判定は筋繊維の細胞とその周辺領域の特徴から総体的に判断するものであり、ピクセル単位での正確な判定は非常に難しく、畳み込みニューラルネットワークの学習に十分な量の学習データセットを集めることは現実的ではない。そこで本研究では、再生過程における時間経過といった画像に付与された情報を弱教師ラベルとして学習に用い、また物体検出と領域分割のネットワークを組み合わせることによって、筋組織再生過程の定量化を行う手法を提案する。

2. 先行研究

2.1 細胞画像処理と畳み込みニューラルネットワーク

2.1.1 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク（Convolutional Neural Network, 以下 CNN）とは、人間の視覚野における信号処理のネットワークを参考に開発された順伝播型ネットワークの一種である。主に画像認識の研究においてその成果を

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University

² 大阪大学大学院薬学研究科
Graduate School of Pharmaceutical Science, Osaka University

a) senoo@ist.osaka-u.ac.jp

発揮しており、画像から最適な特徴量を抽出することを得意としている。

CNN は畳み込み層とプーリング層と全結合層の三種類で構成される。畳み込み層は特定の形状に反応するようフィルタを用いて画像の特徴量を抽出することを目的としており、プーリング層では入力画像サイズを縮小させ計算量を減らすと共に認識対象の位置変化の柔軟性を向上させることを目的としている。

このように CNN は画像から特徴量を抽出する際の最適な構造を有している。一般的に単純なクラス分類タスクに用いられることが多いが、領域分割タスクや物体検出タスクにおいても特徴量抽出器として用いられる場合もある。

2.1.2 物体検出

物体検出とは、ある画像の中で特定の物体のクラスとその位置を予測するタスクのことである。物体検出を行うアルゴリズムとしては Haar Cascade 分類器 [2] やテンプレートマッチングを用いた手法などが存在するが、近年では CNN を用いた手法が注目されている。

CNN を用いた物体検出では、候補領域抽出、クラス分類、ボックス回帰の三段階で構成されることが一般的であり、その代表的な物体検出アルゴリズムとして Fast R-CNN [3] や Faster R-CNN [4] が存在する。

ところが上記のような物体候補領域を抽出した後にそれらを分類問題として物体検出を行うアルゴリズムは、それぞれの処理を直列に行う必要があり、処理速度が遅いという問題がある。その問題を解決すべく画像認識を回帰問題として落とし込み、検出と識別を行うアルゴリズムとして YOLO [5] や SSD [6] が挙げられる。その中でも YOLO は頻繁に改良が続けられており、最新のもので YOLOv5 [7] や YOLOX [8] まで発展している。

2.1.3 領域分割

領域分割（セグメンテーション）とは、画像を物体や背景など任意の領域ごとで分割、抽出するタスクのことである。セグメンテーションにも様々なアルゴリズムが存在するが、生物医療画像分野においては U-Net [9] が注目されている。

U-Net は医療画像セグメンテーションを行うために提案された CNN アーキテクチャであり、2018 Kaggle Data Science Bowl [10] における細胞画像セグメンテーションの精度を競うコンペティションでは U-Net を使用したチームが優勝するなど高い精度を誇る。

U-Net のネットワーク構造を図 1 に示す。U-Net は全結合層を持たず、すべて畳み込み層で構成されており、前半のエンコーダーと呼ばれる畳み込みとプーリングによってダウンサンプリングを行う部分と、後半のデコーダーと呼ばれる畳み込みと逆畳み込みという処理によってアップサンプリングを行う部分に分かれている。逆畳み込みとはその名の通り畳み込みと逆の操作を行う処理で、画像サイズ

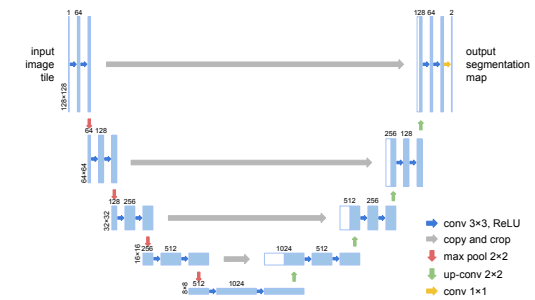


図 1 U-Net のネットワーク構造

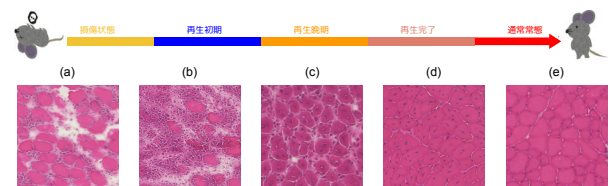


図 2 マウス骨格筋の損傷から再生までの様相

を拡大する効果があるため、畳み込みやプーリングによって縮小された画像をもとのサイズに拡大することが可能である。そして U-Net は一般的な CNN と異なり、エンコーダーの各層で出力される特徴マップとデコーダーの対応する各層の特徴マップに連結するという手法を導入しており、これをスキップ接続と呼ぶ。このスキップ接続が U-Net の最大の特徴であり、特徴が伝播する過程でピクセルのディテールが失われてしまうという一般的な CNN の欠点を補うことが可能である。

2.2 筋組織と骨格筋再生実験

筋組織は生体内で張力を発生させるために分化した組織である。中でも骨格筋は生体に占める割合が最も大きく、壊死と再生が活発である。骨格筋は通常の細胞とは異なり有糸分裂をしない安定した細胞であるが、一度損傷するなどの刺激を受けると活発な再生活動を行う細胞へと変化する。マウスの骨格筋が損傷してから再生するまでの様相を図 2 に示す。図 2 は CTX (cardiotoxin) と呼ばれるアカドクフキコブラ (Naja Pallida) が持つヘビ毒の一種をマウスの筋組織に注入して数日経過した後撮影した画像である。図 2 左 (a) から 3 日後、5 日後、7 日後、14 日後、筋注前であり、CTX は筋線維のみを限定的に破壊する効力を持つため、筋組織の再生過程を観察する手法として用いられる。

2.3 CNN を用いた筋組織再生過程の定量化の問題点

CNN による手法は大量の学習データを必要とし、筋組織画像とその筋組織に再生段階に応じたラベルを各領域ごとに付与したマスク画像の 2 枚をセットで与える必要がある。このラベル付けは専門家により一枚一枚手作業で行う

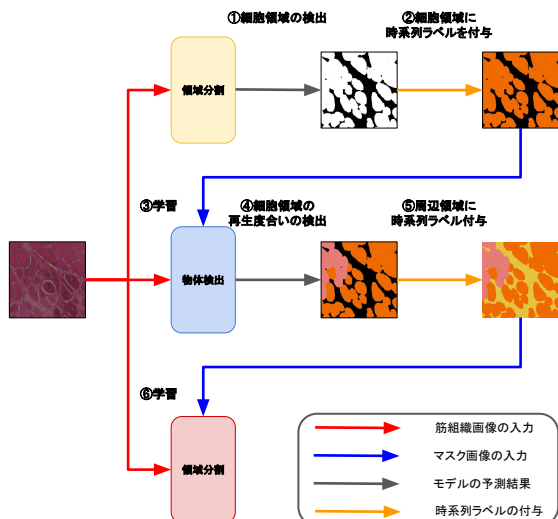


図 3 弱教師ラベルの付与

が、ラベルが付与できる領域は筋組織の再生段階が明確に判断できる部分に限られてしまう。そのため入手できる数少ないマスク画像においても、完全なラベル付けが行われたマスク画像を入手できるケースは稀であり、その多くはラベルが付与されていない領域を含んでいる。加えて手作業の性質上、ある程度大域的な領域の再生段階を判断し、その領域全体に同一のラベルを付与することがほとんどである。そのため、再生段階の異なる領域に対しても同一のラベルが付与されてしまう場合がある。すなわち、CNNで高精度を出すために十分な量のマスク画像を作成することは容易ではない。

また、筋再生過程で筋組織が示す様相は、筋繊維とその周辺領域に大別され、再生段階に関する外観的な特徴もそれぞれで異なっている。筋繊維の再生段階は細胞核の位置など個々の細胞の様相のみで評価することが可能だが、周辺領域は筋繊維の壊死した核や基底膜、浸潤してきた炎症性細胞などの状態を複合して評価する必要がある。このように特徴が異なる領域を単一の量込み込みニューラルネットワークで解析することは適切であるとは言えず、筋組織の細胞領域と周辺領域を個別に解析することが適切であると考えられる。

3. 提案手法

2.3 節の問題を解決するために、本研究では以下の二つアプローチを用いた。

- 弱教師ラベルの付与
- 物体検出と領域分割を組み合わせたセグメンテーション

3.1 弱教師ラベルの付与

筋組織画像のラベル付けが行われた学習データのセット

を集めるのは困難であるが、ラベルが付与されていない筋組織画像のみであれば顕微鏡で撮影するだけでよく、比較的容易に手に入れることが出来る。この教師なし画像を用いて弱教師あり学習を適用する。

3.1.1 弱教師あり学習

弱教師あり学習は学習データを十分な量集めることが困難な場合に有効な手法である。特に教師データが不足している場合に有効であり、弱教師と言われる正確性が乏しいラベルを付与して学習を行うことで実現する。付与される弱教師ラベルはランダムで行われる場合もあるが、一般的にはデータやタスクに関するドメイン知識を用いる場合が多い。

3.1.2 時系列ラベルから弱教師ラベルの取得

2.2 節で述べたように筋組織画像は筋組織に CTX を中筋後、数日から数週間が経過した状態を撮影することが可能である。骨格筋は筋全体で一様に再生されるわけではないが、この経過日数を時系列ラベルとして使用することで筋組織全体の再生段階をおおよそ推測することが可能である。この時に時系列ラベルとして推測された筋組織の再生段階を弱教師ラベルとして教師なし画像に付与することで、弱教師あり学習を実現する。

3.1.3 弱教師ラベルからマスク画像を作成

セグメンテーションタスクに適用するためには弱教師ラベルを筋組織画像のピクセル一つ一つに対応するように付与する必要がある。図3にその流れを示す。まず筋組織画像から細胞のみを検出する領域分割の事前学習済みモデルを使用して事前に細胞領域の特定を行う(図3①)。その後特定した細胞領域に弱教師ラベルを付与し(図3②)、この弱教師ラベルを使用して細胞領域の物体検出を学習させる(図3③)。この時物体検出で予測されたクラスを新たな細胞領域のクラスとし(図3④)、細胞として検出されなかった周辺領域にも弱教師ラベルを付与する付与する(図3⑤)。このようにして細胞領域と周辺領域にそれぞれ異なる弱教師ラベルを付与することで、マスク画像を作成する。

3.2 物体検出と領域分割を組み合わせたセグメンテーション

領域分割は画像中の全ての画素に対してクラスラベルを予測できる一方で、任意の物体に対して異なるクラスラベルを予測することはできない。つまり周辺領域のような一つの物体として捉えることが出来ない領域には有効だが、細胞のように個々の物体領域として捉え、検出することには適していない。一方で物体検出は物体ごとに異なるクラスラベルを予測できるが、画像中の背景を含め全ての画像にクラスラベルを予測することはできない。そこで細胞領域を物体検出を用いて、周辺領域を領域分割を用いて、二つの CNN を使用し、最終的に出力結果を重ね合わせることで、高精度なセグメンテーションを実現する。この手法

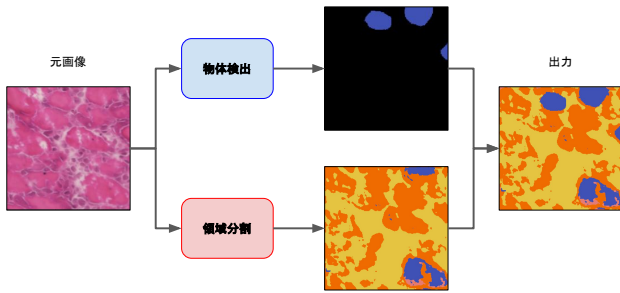


図 4 物体検出と領域分割を組み合わせたセグメンテーション

の学習例を図 4 に示す。

4. 実験

従来手法である通常の教師あり学習の U-Net と、提案手法の精度比較を行う。

4.1 実験環境

4.1.1 実装

物体認識、領域分割ネットワークには 2.1.2 項で述べた YOLOv5 と 2.1.3 項の U-Net をそれぞれ使用した。実装には、YOLOv5 は公式の github から取得したものをそのまま流用し、U-Net にはディープラーニング用フレームワーク Keras[11] を用いた。また弱教師ラベル付与のための細胞領域の検出に用いた領域分割は、事前学習済みモデルの Cellpose[12] を使用した。

4.1.2 実験データ

実験として大阪大学薬学研究科より提供されたマウスの下腿筋組織画像 35 枚を使用し、そのうち 6 枚はテストデータとして評価に用いた。テストデータとして使用した筋組織画像を図 5 に示す。それぞれ CTX を筋注後に (a)3 日経過後、(b)5 日経過後、(c)7 日経過後、(d)11 日経過後、(e)14 日経過後、(f) 筋注前に撮影されたものである。これらのデータを筋組織の再生度に応じて段階に分類し、損傷直後から再生完了にかけて青色、黄色、橙色、桃色、赤色の 5 色に色分けを行った。細胞領域は青色、橙色、桃色、赤色の 4 段階、周辺領域は青色、黄色、桃色、赤色の 4 段階の色分けとなっている。ただし比較のため、物体検出では細胞領域を青色、橙色、桃色、赤色の 4 段階で学習したが、領域分割では既存手法と提案手法ともに青色、黄色、橙色、桃色、赤色の 5 段階で学習した。

4.1.3 比較手法

弱教師あり学習と通常の教師あり学習の比較と、物体検出と領域分割の組み合わせによる学習の有用性をそれぞれ確認するために提案手法を 3 パターンで行い、比較手法を含めた 4 つの手法で実験を行った。

比較手法 通常の教師あり学習による U-Net で 5 段階の再生度で学習する。

提案手法 1 弱教師あり学習による U-Net で 5 段階の再生

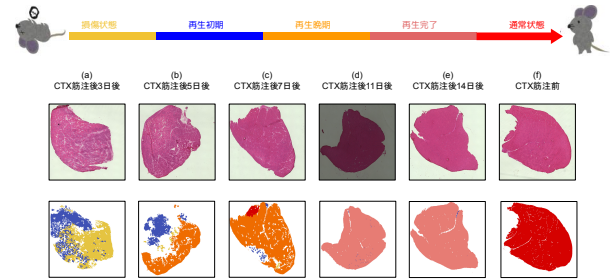


図 5 テストに使用したマウスの下腿筋組織画像

度で学習する、

提案手法 2 通常の教師あり学習による U-Net で周辺領域を 5 段階の再生度で学習し、YOLO で細胞領域を 4 段階の再生度で学習した結果を重ね合わせる。

提案手法 3 弱教師あり学習による U-Net で周辺領域を 5 段階の再生度で学習し、YOLO で細胞領域を 4 段階の再生度で学習した結果を重ね合わせる。

4.1.4 評価指標

実験精度の評価のために 2 種類の評価指標 Pixel Error (PE) と mean Intersection over Union (mIoU) を用いる。PE は予測した結果と正解とのピクセル単位で不一致率を評価した値である。予測結果が正解と一致したピクセル数を T 、一致しなかったピクセル数を F とした時、PE は式 1 で表される。

$$PE = \frac{T}{T + F} \quad (1)$$

mIoU は正解クラスが i であるピクセルをクラス j と予測した数を N_{ij} 、クラス数を k とした時、式 2 で表される。

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{N_{ii}}{\sum_{j=1}^k N_{ij} + \sum_{j=1}^k N_{ji} - N_{ii}} \quad (2)$$

またより詳細な評価を行うために、この PE と mIoU をそれぞれ 2 種類の手法によって評価を行う。

評価 1 全ピクセルに対して評価指標を計算する。

評価 2 正解ラベルのうちラベルが付与されていない（白色）のピクセルを除いて評価指標を計算する。

4.2 実験結果

比較手法と 3 種類の提案手法に対して、それぞれの評価を行った結果を表 1、表 2、表 3、表 4 に示し、そのグラフを図 6、図 7、図 8、図 9 に示す。また実際に予測された出力結果を図 10 に示す。

4.3 考察

評価 1,2 において、テスト画像 (d) と (e) 以外の画像に対するほぼすべての画像で PE と mIoU が比較手法よりも提

	従来手法	提案手法 1	提案手法 2	提案手法 3
(a)	0.577	0.677	0.605	0.695
(b)	0.578	0.599	0.623	0.609
(c)	0.588	0.627	0.546	0.553
(d)	0.941	0.796	0.966	0.943
(e)	0.962	0.801	0.968	0.913
(f)	0.446	0.799	0.613	0.797

表 1 評価 1 による PE の結果

	従来手法	提案手法 1	提案手法 2	提案手法 3
(a)	0.577	0.677	0.605	0.695
(b)	0.578	0.599	0.623	0.609
(c)	0.588	0.627	0.546	0.553
(d)	0.941	0.796	0.966	0.943
(e)	0.962	0.801	0.968	0.913
(f)	0.446	0.799	0.613	0.797

表 2 評価 2 による PE の結果

	従来手法	提案手法 1	提案手法 2	提案手法 3
(a)	0.226	0.269	0.267	0.296
(b)	0.200	0.213	0.212	0.234
(c)	0.182	0.242	0.226	0.242
(d)	0.307	0.250	0.322	0.315
(e)	0.384	0.269	0.324	0.307
(f)	0.151	0.273	0.208	0.273

表 3 評価 1 による mIoU の結果

	従来手法	提案手法 1	提案手法 2	提案手法 3
(a)	0.248	0.313	0.295	0.343
(b)	0.268	0.293	0.284	0.319
(c)	0.196	0.266	0.247	0.270
(d)	0.308	0.251	0.323	0.318
(e)	0.386	0.271	0.326	0.309
(f)	0.152	0.277	0.210	0.276

表 4 評価 2 による mIoU の結果

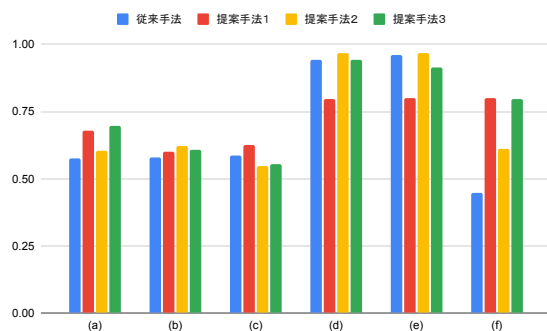


図 6 評価 1 による PE の結果 (グラフ)

案手法の方が高い結果になった。(d) と (e) に関しては僅かに比較手法の方が高い精度を発揮した。ところが必ずしも比較手法の方が精度が高いとは言いがたい。比較手法の通常の教師あり学習で使った学習データには、教師ラベルのうち桃色が占める割合が非常に高く、データの不均一性から (d) と (e) の結果に桃色が出力されたことが考えられ

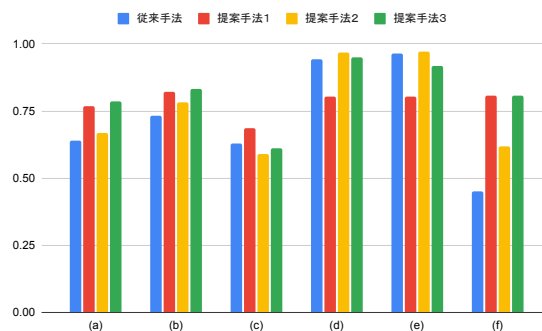


図 7 評価 2 による PE の結果 (グラフ)

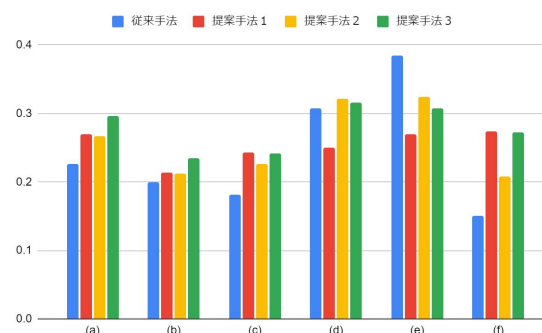


図 8 評価 1 による mIoU の結果 (グラフ)

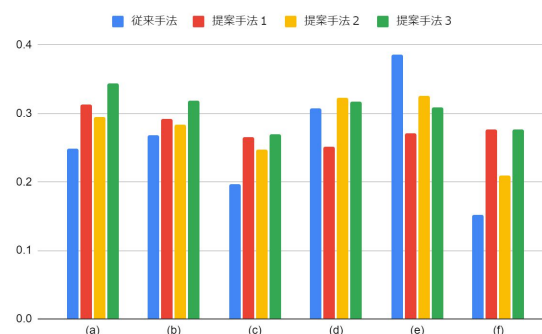


図 9 評価 2 による mIoU の結果 (グラフ)

ラベル	青	黄	橙	桃	赤	白
ピクセル数 (10 ⁵)	80	33	52	313	2	1196

表 5 比較手法で用いた教師ラベルの詳細

る。実際にすべての画像に対して桃色と分類された割合が高く、エラー値であると考えられる。表 5 に比較手法で用いた教師ラベルのそれぞれのピクセル数を示す。それに対して提案手法で時系列ラベルをもとに均一に弱教師ラベルを付与したことにより、ラベルの不均一性が解消された。その根拠として、同じ U-Net を使用した比較手法と提案手法 1 を比較すると、提案手法 1 ではすべての画像において桃色と予測される割合が減っていることが確認できる。

(a) と (b) の画像に注目すると、比較手法と提案手法 2、提案手法 1 と提案手法 3 で共に後者の方が高い精度を示している。これは領域分割が上手く働き、細かな細胞の再

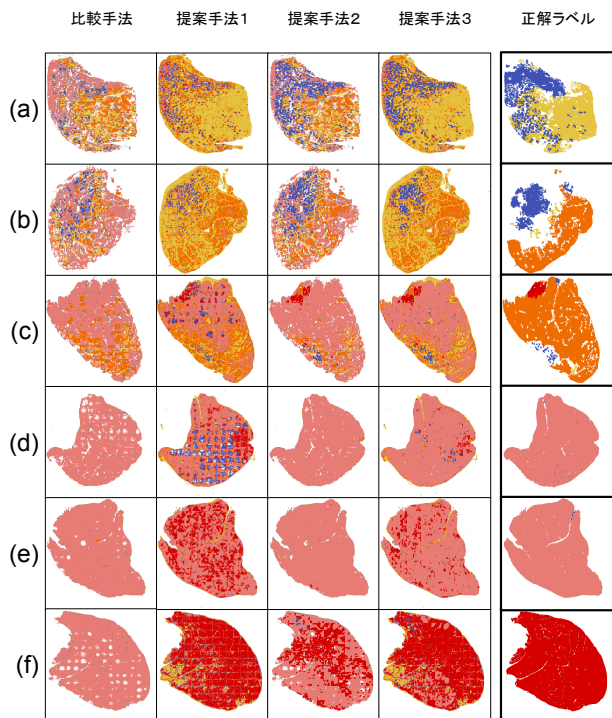


図 10 実験結果画像

生段階の異なりが検出できていることが要因だと考えられる。一方で (c) の画像で提案手法 1 と提案手法 3 を比較すると、物体検出を適用した提案手法 3 の精度が低下していることが確認できる。橙色と桃色の細胞の違いは細胞の密度である。そのため細胞領域のみに注目する物体検出でその違いを区別するのは困難だが、その周辺領域にも注目する領域分割の方が細胞の密度の違いを認識できたと考えられる。

5. おわりに

本研究では、時系列ラベルから取得した弱教師ラベルを弱教師あり学習に適応し、CNN による物体検出と領域分割を組み合わせたセグメンテーションにより、筋組織画像に対する再生度に応じた分割精度を向上させることが出来ると示した。ドメイン知識を利用した弱教師あり学習や複数の CNN を組み合わせる学習手法はその他の生体画像データにも適応でき、本研究はその一例となり、実用化への一助になると考える。

参考文献

- [1] Uchida, S.: Image-informatics for biology and biology for image-informatics, *Journal of the Institute of Electronics, Information and Communication Engineers*, Vol. 98, No. 7, pp. 597–603 (2015).
- [2] Viola, P. and Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features, *IEEE Conf Comput Vis Pattern Recognit*, Vol. 1, pp. I–511 (2001).
- [3] Girshick, R.: Fast r-cnn, *Proceedings of the IEEE inter-*

national conference on computer vision, pp. 1440–1448 (2015).

- [4] Ren, S., He, K., Girshick, R. and Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, Vol. 28, pp. 91–99 (2015).
- [5] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788 (2016).
- [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector, *Lecture Notes in Computer Science*, p. 21–37 (2016).
- [7] Ultralytics: YOLOv5, , available from <https://github.com/ultralytics/yolov5> (accessed 2022-2-1).
- [8] Ge, Z., Liu, S., Wang, F., Li, Z. and Sun, J.: YoloX: Exceeding yolo series in 2021, *arXiv preprint arXiv:2107.08430* (2021).
- [9] Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241 (2015).
- [10] Caicedo, J. C., Goodman, A., Karhohs, K. W., Cimini, B. A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C. et al.: Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl, *Nature methods*, Vol. 16, No. 12, pp. 1247–1253 (2019).
- [11] Gulli, A. and Pal, S.: *Deep learning with Keras*, Packt Publishing Ltd (2017).
- [12] Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation, *Nature Methods*, Vol. 18, No. 1, pp. 100–106 (2021).