



Title	WSI解析に向けた筋組織の再生過程評価における弱教師あり学習
Author(s)	山岡, 悠
Citation	大阪大学, 2025, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/101769
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

WSI解析に向けた
筋組織の再生過程評価における
弱教師あり学習

提出先 大阪大学大学院情報科学研究科

提出年月 2025年1月

山岡 悠

関連発表論文

1. 学術論文

- 1-1. Yu Yamaoka, Weng Ian Chan, Shigeto Seno, Kanako Iwamori, So-ichiro Fukada, Hideo Matsuda. Quantifying the Recovery Process of Skeletal Muscle on Hematoxylin and Eosin Stained Images via Learning from Label Proportion. *Scientific Reports*, 14, 27044, November 7, 2024 (博士論文第 3 章).

2. 国際会議

- 2-1. Yu Yamaoka, Weng Ian Chan, Shigeto Seno, So-ichiro Fukada, Hideo Matsuda. Learning from Similarity Proportion Loss for Classifying Skeletal Muscle Recovery Stages. *The Medical Image Computing and Computer Assisted Intervention (MICCAI) workshop on Advancing Data Solutions in Medical Imaging(ADSMI)*, October 10, 2024 [oral presentation]. (博士論文第 4 章).

内容梗概

顕微鏡は組織や細胞といった微小世界を拡大して見せることで、肉眼だけでは得られない物体の識別能力を人に与えてくれた。細胞単位の観察から組織異常の有無を評価できるが、顕微鏡レンズを通して見える領域は組織の一部であり、組織全体の状態把握にはすべての視野を隅々まで探索することが望ましい。そこで組織の各領域をデジタル画像化しコンピュータ上で再統合するという、標本全体をデジタル化した全組織スライド画像 (Whole Slide Image; WSI) を用いた解析手段がある。WSI は単位面積あたりのピクセル密度が高く、組織のあらゆる領域を詳細に観察できる反面、ピクセル数の大きい WSI 全域に注意を払うことは大きな労力を伴う。労力を軽減しつつ、高精度な解析のためのユーザ支援を狙い、デジタル画像である点を活かした画像全域にわたる定量的なコンピュータ解析法を構築する。

本研究で扱う WSI は、骨格筋の疾病を克服する医学・薬学研究で用いられている再生過程の筋組織画像である。筋組織の状態は領域ごとに局所性が見られる一方、WSI すべてを観察し評価することは労力を伴うため、コンピュータ支援による手助けが有効である。従来の筋組織 WSI のコンピュータ支援研究では解析対象が限定的であった。特に毒物注入による組織の壊死と再生が始まって数日間は、筋線維が元の形状に戻るまで複雑な分化や成長を経るため、従来の一細胞に関する面積や円形度といった明示的特徴量では再生の段階を分類できなかった。そこで機械学習手法を用いた暗黙的特徴量によって、筋線維が壊れてから元の形状に戻る一連の再生段階の分類を目指す。

機械学習において一般的な教師あり学習には、データ解析の答えとなる教師ラベルが必要となる。専門的な WSI の場合、膨大な教師データを用意することは労力や正確性の観点から困難であるため、画像全体に何らかの事前情報があることを利用した弱教師あり学習は教師作成コストが小さい。そこで筋組織 WSI において、各動物サンプルを安楽死させてから画像化する際に一緒に保存する毒物注入からの「日付」に注目し、日付が浅い個体であれば損傷が激しく、日付が経った個体は組織全体で回復した細胞領域が増えていく、という事前知識を用いる。個々の細胞領域画像の教師ラベル作成をせずとも、筋組織 WSI の概観観察と再生メカニズムの知見から、各日付における再生段階の「クラス比率」を出すことは低コストである。「日付」と「クラス比率」を結びつけた弱教師学習手法を用いて、再生段階の推論モデルを構築した。またモデルを用いた筋組織の再生段階解析ソフトウェアを開発し、ウェブ上に公開することでコミュニティへの貢献を果たした。

筋細胞のクラス分類モデルは画像を低次元に圧縮する特徴量抽出器 (backbone) と特徴量からクラスを推論する分類器 (head) に分かれる。従来のクラス比率学習では backbone に既存の大規模事前学習モデルを用いており、学習によって更新されるのは head に限定されていた。一般画像で事前学習された backbone が筋細胞の特徴抽出に適している保証はなく、実際再生段階の分類クラスを増やすと精度が劣化することが確認された。さらにクラス比率学習では学習時にクラスを名義尺度として扱うため、再生過程における細胞の順序的な類似性を考慮することができない。これら特徴量抽出器の未更新とクラスの名義尺度化の課題に対処するため、類似密度分布による順序尺度類似比率学習 (Ordinal Scale Learning from similarity proportion; OSLSP) を提案し、特徴量空間においてクラス類似度に注意した損失関数の計算を実現した。OSLSP によって特徴量抽出器が筋細胞ドメインに適合したことを確認し、分類精度の向上を達成した。

本研究では慢性的に教師データが不足する WSI 解析のコンピュータ支援において、クラス比率を用いた弱教師あり学習パイプラインを構築し、学習したモデルの妥当性を確認した。また比率学習の課題を解決するための新しい弱教師あり学習である OSLSP を提案し有効性を示した。提案した OSLSP は WSI に留まらず教師のないデータに応用可能であり、機械学習分野の発展に寄与する。

目次

第 1 章 序論	1
1.1 研究背景	1
1.1.1 顕微鏡による微小世界の可視化	1
1.1.2 全組織スライド画像 (Whole Slide Image)	2
1.1.3 古典的画像解析	3
1.1.4 暗黙的特徴量	5
1.1.5 弱教師あり学習	6
1.2 研究目的	9
1.3 本博士論文の構成	9
第 2 章 明示的特徴量による筋組織画像解析	12
2.1 緒言	12
2.2 筋組織画像	14
2.2.1 画像化と解析	14
2.2.2 筋組織の再生過程	16
2.2.3 染色方法の違い	16
2.3 関連研究	18
2.3.1 従来の筋組織画像のコンピュータ支援	18
2.3.2 再生評価における HE 染色の優位性と課題	20
2.4 データセット作成法	22
2.4.1 動物処理	23
2.4.2 検証用データのアノテーション	26
2.5 予備実験	26
2.5.1 学習データの拡張	26
2.5.2 明示的特徴量による分類結果	28
2.6 結言	32
第 3 章 筋組織再生評価に向けた暗黙的特徴量によるクラス比率学習	34
3.1 緒言	34

3.1.1	機械学習手法による推論法	35
3.1.2	日付を弱教師としたクラス比率学習	36
3.2	提案手法	38
3.2.1	提案手法の概観	38
3.2.2	Cellpose のファインチューニング	40
3.2.3	日付ごとのクラス比率計算	42
3.2.4	推論器 (head) のアーキテクチャ設計	43
3.2.5	クラス比率学習	44
3.2.6	リカバリースコアの導入	45
3.3	実験結果と議論	47
3.3.1	実験手法	47
3.3.2	再生過程筋組織のセグメンテーション	48
3.3.3	再生過程の筋細胞クラス分類	52
3.3.4	CTX とグリセロールの回復経過解析	56
3.3.5	免疫染色と手動アノテーションの比較	62
3.4	議論	65
3.5	結言	67
第 4 章	類似比率学習による筋細胞へのドメイン適応	68
4.1	緒言	68
4.2	関連研究	71
4.3	類似比率学習 (Learning from Similarity Proportion) の提案	72
4.3.1	細胞形態の類似度	72
4.3.2	類似比率損失の概要	73
4.3.3	インスタンス類似度による予測密度分布	74
4.3.4	クラス比率の組み合わせによる真密度分布	76
4.4	再生過程の細胞クラス分類実験	78
4.4.1	実験手法	78
4.4.2	クラス分類実験結果	79
4.5	考察	82
4.5.1	検証/学習データのドメインギャップ	82
4.5.2	実用に応じた類似度定義	85

4.6 結言	87
第 5 章 結論	88
付録 A Bag 学習	91
A.1 疑似ラベル法	91
A.2 Learning from Label Proportion: LLP	92
A.3 IIB-MIL	94
付録 B 多数決学習	95
付録 C 多クラス分類の評価指標	98
謝辞	100
参考文献	103

目次

1.1	全組織スライド画像の例	2
1.2	WSI の解析例 [1]	3
1.3	画像の最小単位ピクセルとは [2]	4
1.4	色の平均値による分類例 [2,3]	5
1.5	明示的特徴量が不明な細胞の分類例	5
1.6	機械学習の種類	8
2.1	筋ジストロフィーと健常時の筋組織の比較 [4]	12
2.2	筋組織に含まれる大量の細胞	13
2.3	筋線維の特徴量例	14
2.4	筋組織セグメンテーションの例	15
2.5	平行筋と羽状筋の CSA の違い	15
2.6	CTX 注入による筋組織再生の観察 [5]	16
2.7	筋組織の再生過程	17
2.8	HE 染色とラミニン染色	18
2.9	関連研究の筋組織画像 [6–13]	19
2.10	ラミニン染色用ソフトの HE 染色画像への適用例	20
2.11	明示的特徴量を持つ画像例	22
2.12	HE 染色と免疫染色の画像例	25
2.13	専門家によるクラスアノテーション	26
2.14	画像データ拡張例	27
2.15	各画像の細胞数	28
2.16	WSI に含まれる細胞の面積平均	29
2.17	WSI に含まれる細胞の feret 比 (長径短径比) の平均	30
2.18	WSI に含まれる細胞の輪郭長平均	30
2.19	WSI に含まれる細胞の円形度平均	31
2.20	回復過程の細胞の明示的特徴量によるサポートベクター分類結果	32
2.21	再生過程の細胞特徴	32
3.1	筋組織再生の局所性	34
3.2	日付ごとの細胞状態分布の正規予測	37

3.3	クラス比率学習	37
3.4	提案ソフトウェアの概観	39
3.5	クラス推論のパイプライン	40
3.6	Cellpose の GUI を用いたアノテーション	41
3.7	ラフアノテーション	42
3.8	クラス比率の計算方法	43
3.9	クラス分類器のアーキテクチャ	44
3.10	クラス比率学習	45
3.11	256 ピクセル正方形のセグメンテーション結果例 (文献 [14] の Fig.1(a) 再掲)	48
3.12	検出した細胞数 (文献 [14] の Fig.1(c) 再掲)	49
3.13	WSI に統合後のセグメンテーション結果 (文献 [14] の Fig.1(b) 再掲) . .	49
3.14	セグメンテーション平均 IoU (文献 [14] の Fig.1(d) 再掲)	50
3.15	セグメンテーション F1-スコア (文献 [14] の Fig.1(e) 再掲)	51
3.16	日付ごとの IoU と F1-score	51
3.17	クラス分類結果例 (文献 [14] の Fig.2(d) 再掲)	52
3.18	WSI に統合後のクラス分類結果 (文献 [14] の Fig.2(a) 再掲)	53
3.19	クラス分類の混合行列結果 (文献 [14] の Fig.2(b) 再掲)	54
3.20	クラス分類の recall (文献 [14] の Fig.2(c) 再掲)	54
3.21	クラス分類の precision (文献 [14] の Fig.2(c) 再掲)	55
3.22	クラス分類の F1-スコア (文献 [14] の Fig.2(c) 再掲)	55
3.23	3 分割交差検証 (文献 [14] の Fig.2(e) 再掲)	56
3.24	グリセロールの回復傾向例	57
3.25	CTX とグリセロールの回復傾向の解析 (文献 [14] の Fig.3(a) 改変) . . .	58
3.26	リカバリースコアと細胞領域割合の 2 次元散布図 (文献 [14] の Fig.3(b) 改変)	59
3.27	リカバリースコアと細胞領域割合の経過日数ごとの値 (文献 [14] の Fig.3(c) 改変)	61
3.28	CTX 注入 3 日後の筋組織連続切片画像群 (文献 [14] の Fig.4.(a) 改変) . .	62
3.29	CTX 注入 5 日後の筋組織連続切片画像群 (文献 [14] の Fig.4(b) 改変) .	63
3.30	CTX 注入 7 日後の筋組織連続切片画像群 (文献 [14] の Fig.4(c) 改変) . .	64
3.31	凍結処理に失敗した画像の解析例	66
4.1	複数インスタンス学習 (Multiple instance learning: MIL) の例 [15] . . .	68
4.2	クラス分類のパイプライン	69

4.3	クラスの類似性	70
4.4	検証データと学習データのドメインギャップ	70
4.5	線維の時間経過による形態変化と類似度	72
4.6	順序尺度類似比率学習 (OSLSP) の概要	73
4.7	インスタンス類似度の確率密度	75
4.8	ガウス展開法と torch.hist の比較	76
4.9	バッグのクラス比率から導出された類似度確率密度	77
4.10	専門家による 5 クラス分類のアノテーション	78
4.11	クラス分類による WSI への色塗り結果	81
4.12	抽出器別の特徴量空間	82
4.13	Day0 と他の日付のインスタンス群コサイン類似度分布	84
4.14	Day3,5,7 と他の日付のインスタンス群コサイン類似度分布	85
4.15	クラス分類器の LLP における test,train 損失	86
1.1	教師ありと弱教師 (クラス比率)	91
1.2	クラス比率学習	92
1.3	分布差を計算する手法の収束性比較	93
2.1	多数決アノテーション	95
2.2	多数決アノテーション結果	96
2.3	各日付の推論クラス比率	96
2.4	確信度スコアの確率密度分布	97

表目次

2.1	染色法の比較	17
2.2	筋組織画像コンピュータ解析の関連研究	21
2.3	本論文で使用した筋組織画像一覧	23
3.1	3 分割検証による分類の汎化性能評価	56
4.1	WSI の複数インスタンス分類の手段一覧.	72
4.2	異なる特徴量抽出器を用いたときのクラス分類器 (head) の推論結果 . . .	80

第 1 章 序論

1.1 研究背景

1.1.1 顕微鏡による微小世界の可視化

顕微鏡とは肉眼で観察できない微小領域を拡大する視覚的な道具である。顕微鏡の発明者は諸説あるものの、16 世紀末から微小な構造の観察が始められており、ロバート・フック (Robert Hooke, 1635 年-1703 年) が残した著 *Micrographia* [16] にはコルク断面の細胞壁、ノミの拡大図といった顕微鏡画像の記録が残っている。

微小領域の可視化は小さな物体の形状把握を助け、肉眼所見だけでは得られない空間識別能力を人に付与してくれた。例えば医学において、病原菌を発見し形状からその種類を特定することができれば、狭域な抗菌薬が投与できる [17]。あるいは工業において、人が持ち歩けるようなコンピュータ^{*1} を作成するための微細な回路設計を可能とした。そして科学研究の場において、組織の構造や微生物の活動とそのメカニズムの理解に役立てられている。

こうした小さな世界の可視化はミリメートル (mm)、マイクロメートル (μm)、ナノメートル (nm)、オングストローム (\AA) のスケールに分けられる。mm スケールでは組織の概観把握、 μm スケールだと組織に含まれる細胞を観察することが可能で、波長 400-700nm の可視光による光学顕微鏡で見ることができる。顕微鏡の拡大スケールは波長による物理的制約があり、より小さな nm スケールの分子やウイルスを観察するには電子 (波長 0.01nm 前後) を用いた電子顕微鏡や原子間の引力を利用した原子間力顕微鏡 (Atomic Force Microscope) の出番となる。mm- μm スケールの観察は細胞の増殖や分裂、移動といった変化を捉え、細胞が集まった組織の機能理解を助ける。nm 以下のスケールは細胞の膜電位、代謝活動といった分子レベルの動きを捉えるために用いる。すなわち拡大スケールの選択は観察目標に依存しており、分子レベルの反応に比べて巨視的な細胞動態の観察は、ある組織が異常なくその役割を果たしているか評価する場合に用いる。ただし組織に含まれるうちのたった一つの細胞をつぶさに観察しても、組織の機能を説明することにはならない。組織の評価には各細胞領域の状態を総合して判断することが求められる。

^{*1} 初期コンピュータの一つである ENIAC は、第二次世界大戦頃に米陸軍の砲弾弾道計算の為に開発され、その大きさは部屋一つ分ほどあった。

1.1.2 全組織スライド画像 (Whole Slide Image)

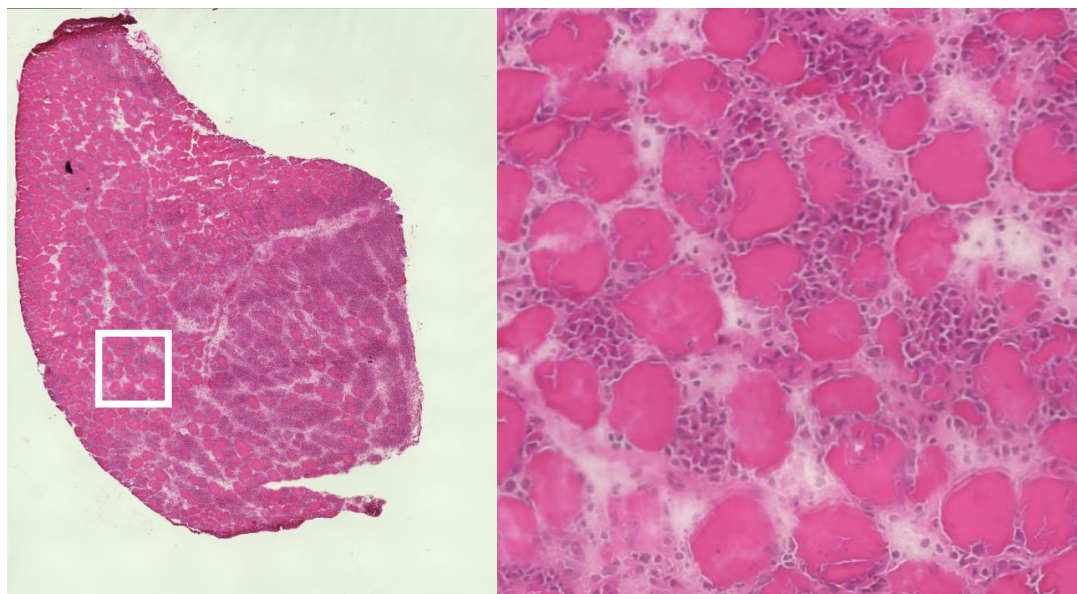


図 1.1 全組織スライド画像の例

組織全体の把握には、標本が載ったスライドガラスを動かしながら高倍率レンズによって各細胞領域をくまなく観察する、という手法が考えられる。このとき低倍率レンズで組織を概観観察して注目すべき領域を限定してから拡大して観察することで負担を軽減する。レンズを通して網膜に届く光は連続なアナログ情報であるから^{*2}、レンズの倍率を変えることで可視化したい領域スケールを変更でき、拡大縮小しても一定の解像度^{*3}を担保する。一方デジタル画像はイメージセンサのピクセルサイズと数に依存した離散情報であり一度保存された情報量が増えることはないが^{*4}、コンピュータ上に無尽蔵に情報を記録でき、かつ他者との情報共有を簡便にする。これらを踏まえると、それ以上拡大しない細胞単位を撮像してデジタル画像化しコンピュータ上で再統合すれば、デジタル画像と顕微鏡を用いた目視、双方の短所を補い合った画像を作ることができる。このようなスライドガラス上にある標本全体をデジタル化したものは全組織スライド画像 (Whole Slide Image; WSI) と呼ばれており、図 1.1 [14] のように組織全体の広範な領域を高いピクセル密度で画像化したものである。例えば人組織切片を染色した WSI は一般に病理画像と

^{*2} 厳密には光はプランク定数を基準にした離散的な性質を持つが、顕微鏡レンズの倍率変更の分解能に比べれば十分連続と近似できる。

^{*3} この解像度は人間の視覚に依存した値となる。

^{*4} 画像超解像技術 (Super-resolution) といった画像の解像度を高めて出力する研究分野もある。

呼ばれ、疾病の進行状況の把握や原因の解明に用いられる。動物実験による薬学研究では安楽死させたマウスの組織を切り取り染色を施して画像化し、組織内の筋細胞のふるまいを観察する。特定の細胞膜や細胞のみに着目する場合は免疫染色と呼ばれる特定の抗体に対する細胞の免疫反応を利用し、特異的に染色を施すことで図 1.1 [1] の WSI のように観察対象を限定する。

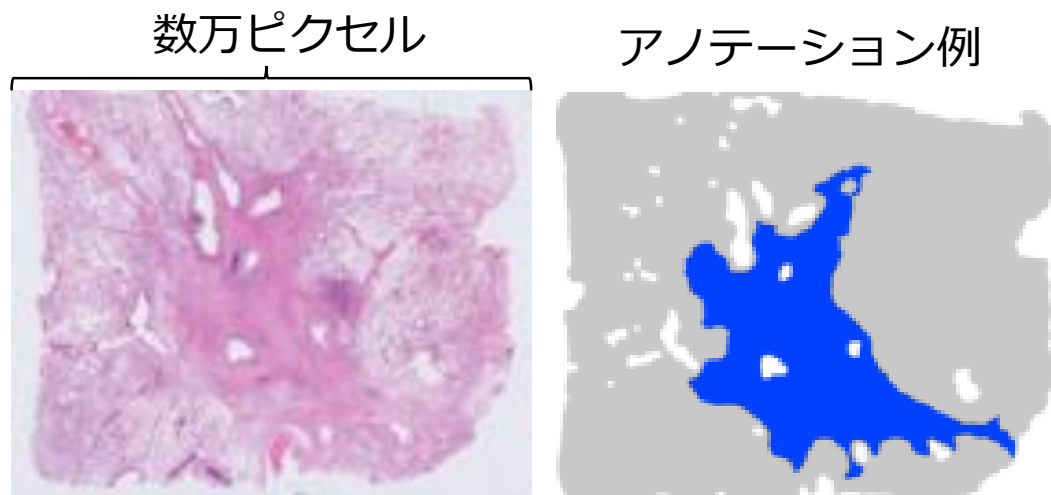


図 1.2 WSI の解析例 [1]

WSI は単位面積あたりのピクセル密度が高く、細部まで詳細に表現できている反面、画像中すべての領域が観察目標に関連するとは限らない。例えば図 1.2 のように縦横のピクセルサイズが数万ピクセルの WSI から、癌腫瘍領域をくまなく探索することには時間を要する。すなわち観察者は疾病への糸口を探るために WSI 全領域に注意を払わなくてはならず大きな労力を伴い、簡便な情報処理であっても繰り返しによるヒューマンエラーの危険がある [1, 17]。そこでデジタル画像化である点を活かし、コンピュータを用いた人間にはない情報記憶能力や画像全領域にわたる定量的な解析能力によって人の観察を支援することで、労力軽減及びより高精度な解析を実現する。

1.1.3 古典的画像解析

図 1.3(りんごは illustAC フリーイラスト参照 [2]) に示すように、コンピュータ上で画像を扱う際の最小単位であるピクセル (pixel) は、カラーであれば Red, Green, Blue の RGB3 チャンネル、白黒であれば 1 チャンネルの値で構成されている。画像の縦幅を $H[\text{pixel}]$ 、横幅を $W[\text{pixel}]$ とすればカラー画像は $H * W * 3$ 次元の情報体としてコンピュータ上で扱うことができる。

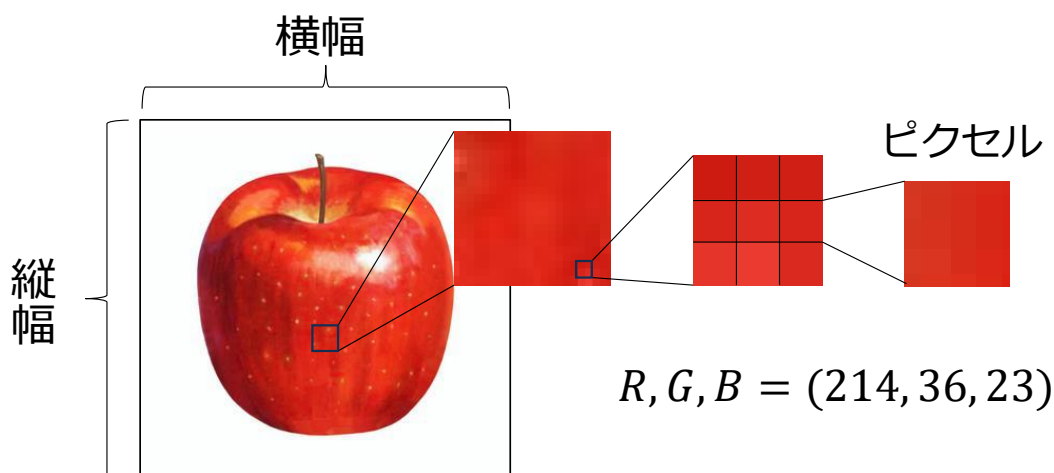


図 1.3 画像の最小単位ピクセルとは [2]

例えばカラー画像から特定の物体が写されているかどうかの 2 値画像分類を行うとする．コンピュータ上では $H * W * 3$ 次元のデータがあり，各次元に閾値を設けて多数決分類してみようとしても，各次元が特定物体の分別にどれだけ相関のある値かどうかは判らない．よって物体を認識するために $H * W * 3$ 次元のデータから特徴量と呼ばれる代表値を計算し，特定物体との関連づく値を探すことを図る．古典的な特徴量の一例は平均である．画像の縦幅を H ，横幅を W ， (x, y) の位置にあるチャンネル c のピクセルの値 $I_{x,y,c}$ とすれば，全ピクセル群の RGB それぞれの平均値 \bar{I}_c を出すことができる．

$$\bar{I}_c = \frac{1}{H \times W} \sum_{y=1}^H \sum_{x=1}^W I_{x,y,c} \quad (1.1)$$

仮に特定物体の二値分類が「赤リンゴかそれ以外か (図 1.4左)」という課題であれば，Red の平均値を見ることで判定可能であろう (いちご絵は StockNova フリーイラスト参照 [3])．このように数式あるいはアルゴリズムによって導出方法が定義されている特徴量を明示的特徴量と呼ぶ (explicit feature)．一般に値の意味を人が解釈しやすく，既に特徴量が明確な物体では，分類目標に合わせて手動で特徴量選択する．このとき分類課題が複雑になると，特徴量の数を増やさなくてはならない．先の「赤リンゴかそれ以外か」という二値分類課題が「赤リンゴか，いちごか，それ以外か (図 1.4右)」だったとする．すると Red の平均値だけでは「赤リンゴ」と「いちご」を区別することができないため，例えばいちごの含まれた緑の葉に注目し，Green の平均値と閾値を導入することで分類を試みる．これは分類する数が増えれば増えるほど，用いる特徴量が増えるということの意味する．

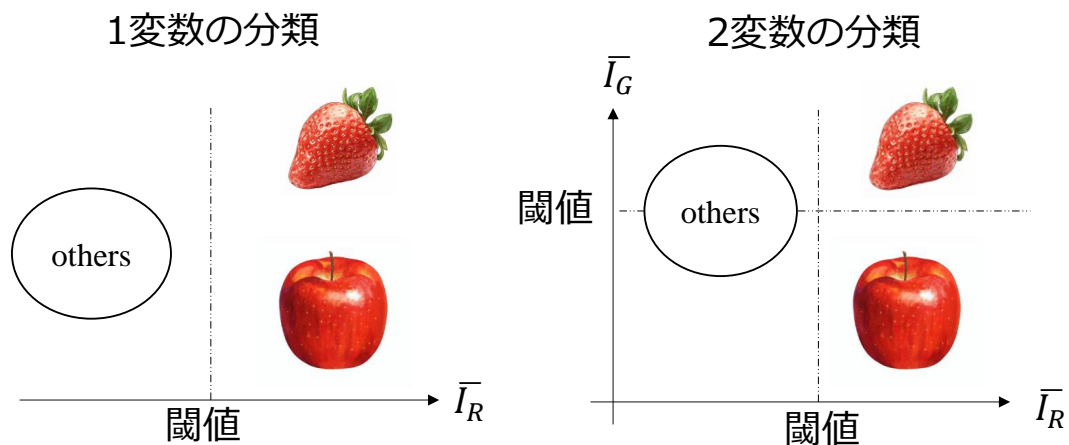


図 1.4 色の平均値による分類例 [2, 3]

1.1.4 暗黙的特徴量

より複雑な分類をするためには、より多くの特徴量を用いなくてはならない。選択された特徴量は分類すべき物体の特徴と符号するものでなければならず、図 1.5にある 5つの異なる細胞の分類といった人が定式化できない特徴を持った物体の場合、明示的特徴量だけでは分類不可を意味する。そこで特徴量と画像的意味の繋がりが明示的ではないものの、大量の特徴量を用いて分類する手段することで、分類不可の壁を打破する。例えば Red の平均値、Green の平均値という風に特徴量を増やしてみても、図 1.3のように、一見赤に見えるピクセルであっても、青や緑の成分が含まれていることから、これらの特徴量は分類にあたって独立な関係ではない。大量の特徴量の関係を相補的に記述する術が求められる。

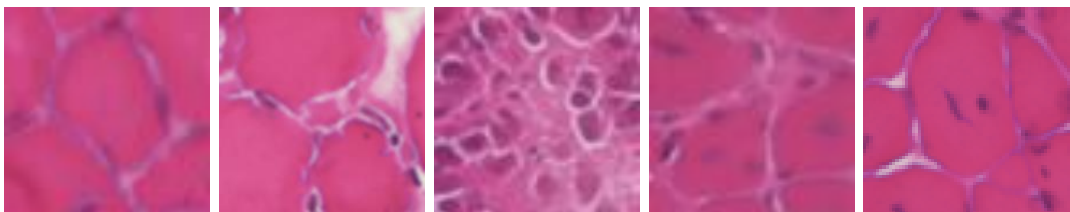


図 1.5 明示的特徴量が不明な細胞の分類例

台頭してきたのは人神経システムを模したニューラルネットワーク [18] を用いた機械学習である。各特徴量は神経細胞に模したパーセプトロン [18] と呼ばれる線形分類アルゴリズムにそれぞれ入力され、パーセプトロンが網目状に接続されることで、特徴量同士

の関係を表現する。機械学習の興隆を支えているのは、入力データと答えとなる教師データを用意するだけで予測モデルを学習する end-to-end と呼ばれる手法であろう。人の手が入るのはハイパーパラメータの調整やデータを準備する作業であり、データ量が豊富かつ高性能計算機を利用できる場合に有効である。データが不足する場合や end-to-end では狙った精度が出ない場合、埋め込み (embedding) と呼ばれる中間特徴量を利用する手法がある。embedding とは画像であれば $H * W * 3$ の高次元データにおいてモデルが扱いやすく (例えば分類であれば境界面を計算しやすく) するために、データを低次元の特徴量ベクトルに圧縮表現することである。end-to-end でも利用される畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) [19], Vision Transformer (ViT) [20] や Mamba [21] といった様々なネットワーク構造 (アーキテクチャ) の中間層においても、入力データは一定の次元数に圧縮されている。この圧縮された次元数の値を、古典的な明示的特徴量に対して、暗黙的特徴量 (implicit feature) と呼ぶ。特徴抽出後は目的の用途に応じて、例えばクラス分類であるならば、ランダムフォレスト [22], k-means 法 [23], あるいは再度ニューラルネットワークに通すといった手法で解析を達成する。上記のようにデータから特徴抽出を行う層を backbone、課題に応じて異なる出力層を head と呼ぶ。

1.1.5 弱教師あり学習

機械学習は予測誤差と呼ばれる推測と観測の差分からネットワークモデルを学習する *5。あるデータから何かの答えを学習するシステムにおいて、推測とは入力データに対するモデルの出力であり、観測はデータの答えに相当する。このデータの答えを教師 (supervisor) と呼び、機械学習では教師のありなしで学習が分かれる。教師がある場合はモデルを更新するための損失関数は推論と教師から計算される一方、教師なしの場合はデータの相関を利用した手法に限定される。教師にはモデルが推論して欲しい結果そのものである完全な教師以外にも、データの一部にのみ完全な教師を持つ半教師、正解であるか不明である不完全な教師である弱教師、データ自身から教師を作る自己教師といった種に分かれる。ただし head 部の学習には教師がなくとも backbone の学習には何らかの教師が用いられている場合もあることに注意する。一般に教師がないという場合、head 部の教師がないことを指している。この教師の種類による学習方法を以下 5 つに分けて詳しく紹介する。

*5 人も遠心性コピーに代表される通り、運動感覚の予測誤差による学習システムが備わっているとされる。

1. 教師あり学習

教師ラベルと呼ばれる目的の解析対象の答えが、画像と一緒にあるデータセットで学習する方法。データの作成の多くの場合、アノテーションと呼ばれる人の手作業によって教師ラベルを付与する。図 1.6では○がデータ点に、○に塗られた色がクラスに相当する。教師あり学習では各データにクラスの色が割り振られており、データを区別するような線を引くことがモデル学習に相当する。

2. 半教師あり学習

データセットの中にアノテーションのあるデータが一部含まれており、残りはアノテーションのない画像で構成されたデータセットで学習する方法。弱教師あり学習とは異なりアノテーションが施されたデータについては教師ラベルが正確であるという保証の下で学習する。例えば図 1.6では一部のデータに赤や青の教師があり、残りは教師がない(黒)。一般にすべての教師データがある場合に比べて推論精度は落ちる。

3. 弱教師あり学習

データセット一部のアノテーションが不正確である、多分類課題において一部のクラスにしかアノテーションが付与されていない、あるいはデータごとにはアノテーションが付与されておらずデータの集まりに対して何らかの情報が付与されている、といった不完全な教師情報に対して実行する学習法である。弱教師となる情報は画像の種類によって分かれており、場合によっては独自の学習パイプラインを作成する必要がある。例えば図 1.6ではあるデータのグループ内に含まれるクラスの比率が与えられている。

4. 自己教師あり学習

教師をデータ自身から生成する手法。主に backbone の学習に用いられており、大量のラベルなしデータから特徴表現を学習する。データに対して何らかの変換を行い、変換時のパラメータを予測するといった課題設定によって学習を進める。

5. 教師なし学習

解析対象の答えと関連する情報が一切ない画像のみのデータセットで学習する方法。例えば図 1.6ではすべてのデータ点に教師がなく、各データ点の近さを用いたクラスタリングが当たる。

ニューラルネットワークの学習はデータ数が多いほど推論精度が上がるため、実用に足るモデルを学習するためにはそれに応じたデータ数が必要となる。ここで教師あり学習の

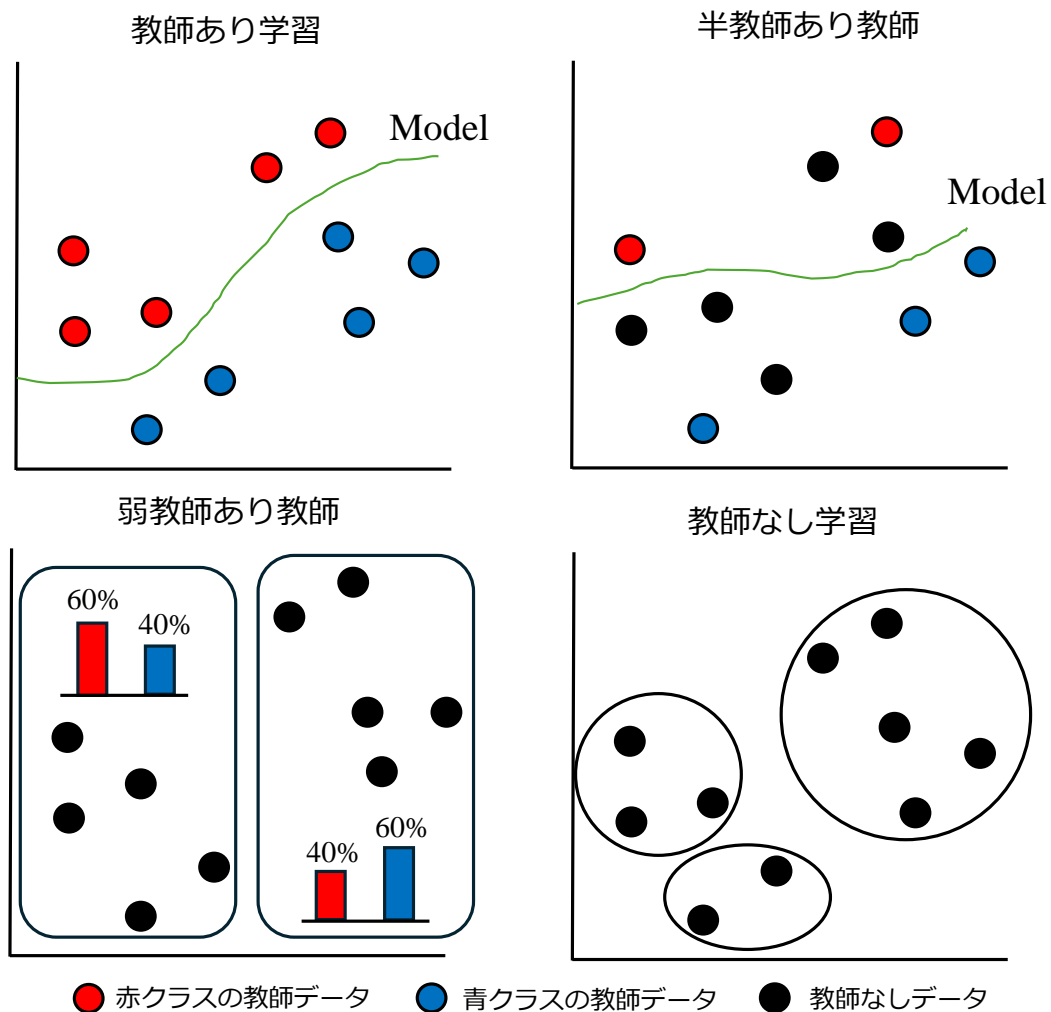


図 1.6 機械学習の種類

手法を用いる場合、WSI のすべての領域に教師を与えなくてはならず、膨大なピクセル群から観察対象の領域をピクセル単位で切り取り教師を付与する作業負荷は計り知れない。また専門性の高い画像である場合、そのアノテーションの正確性の担保についても不明瞭である。そこで WSI によく見られる「画像全体には何かしらの情報がある」ことを利用した半あるいは弱教師あり学習が候補となる。例えば病理画像であれば「組織全体のうちどこかが腫瘍である」あるいは「組織のうち数割が腫瘍である」といった医学的な事前知識に基づいた診断所見が WSI と共に得られることは多い [24]。あるいは科学実験で得られた組織画像であれば、実験条件が弱教師として用いる事前情報となり得る。このように WSI では正確な答えが大量に用意されていることを前提としない教師あり学習以外の手法を用いることで、教師データを作るコストを回避する研究が進められている [25–27]。

また人間情報の観点 [28] から、知覚システムは教師なしで学習する系が備わっているとされており [29,30]、人システムと符号する点でも弱教師・教師なしの手法は興味深いと同時に可能性を秘めている。

1.2 研究目的

本研究では WSI の一つである筋組織画像において、教師ラベルのない条件下で画像化までの実験情報を弱教師としたアノテーションコストの小さい学習法に取り組む。また用いた弱教師あり学習の手法の問題点について言及し、新たな損失関数を提案し手法の改善をすることで、他 WSI への転用時の精度向上に寄与する。筋組織の観察は人における病気や加齢による筋力低下のメカニズムを解明する生物学的研究において、実験効果の大きさを判定する上で欠くことのない解析手段の一つである。生物学的研究における現在最も一般的な画像化の手段は安楽死させたマウスの対象領域の筋切片を染色したのちに顕微鏡で撮影する手法である [31–33]。筋組織に含まれる筋線維やそのほかの細胞の形状や数は筋の持つ運動能力と関連していることから解析の対象となっている [12]。しかし例えばマウス太腿の筋組織には数千単位の筋線維が備わっており、組織領域すべてを人の目によって観察することは、大きなコストを要し (1)、多量の線維を一つ一つ観察することはヒューマンエラーの生む危険性が残り (2)、そして線維の形状を定量的に扱えない (3)、といった WSI 全般に通ずる課題が筋組織画像にも当てはまる。また豊富な教師付き学習データを作成することは困難であるため、弱教師あり学習から筋組織画像解析に挑む。本研究の目的は WSI の一つである筋組織画像において、完全な教師データを用意せずにモデル学習を行う弱教師あり学習手法の導入と改善に試み、その精度向上を確認することをもって、不完全データの多い WSI 解析へ貢献することを狙う。薬学・医学研究領域における貢献は、創薬における評価法の武器となる解析ソフトウェアの開発と、低コストな WSI 解析の学習・推論パイプライン構築である。情報工学領域における貢献は、新しい弱教師あり学習を提案した点にある。

1.3 本博士論文の構成

本博士論文は全 5 章で構成される。第 2 章ではこれまでの筋組織画像のコンピュータ画像解析について俯瞰し、その中でも再生過程における筋組織の画像解析の方法論および意義について述べる。特に解析に用いる特徴量が明示的か暗黙的であるかについて焦点を当て議論し、暗黙的すなわち機械学習手法の必要性を説く。第 3 章では機械学習手

法を特徴量抽出器 (backbone) とクラス分類器 (head) に分け、head の学習法について説明する。再生過程の筋組織内にある細胞を含んだ小さい領域の画像分類を行うために、クラス比率を用いた弱教師あり学習を導入する。クラス比率学習 (Learning from Label Proportion; LLP) は、モデル学習時に、個々のデータ点であるインスタンス (instance) を多数集めたバッグ (Bag) と呼ばれるインスタンス群を形成し、バッグに含まれるインスタンスのクラス比率ラベルのみが与えられる、という問題設定を解くための学習手法である。LLP を再生過程の筋組織画像に適用しその推論性能を確認し、開発したソフトウェアの有用性を検証すると共にコードを公開し、筋組織画像解析コミュニティへの貢献を行った。第 4 章では機械学習手法を特徴量抽出器 (backbone) の学習について述べる。第 3 章で用いた LLP が持つ 2 つの課題、(1) 特徴量抽出器の未更新と (2) クラスを名義尺度化に言及する。(1) 特徴量抽出と呼ばれる画像の低次元圧縮段階において、LLP では大規模な一般物体画像データセットを用いた事前学習モデルを用いており、筋細胞に適した特徴抽出がなされている保証はなかった。(2) また、再生過程段階のクラス分類という本来順序性を持つ画像に LLP を適用すると、クラス間の類似情報が失われてしまう。実際 (1,2) の問題で、分類クラスを増やした際に分類結果及び特徴量空間での可視化から 3 章の手法の限界が明らかとなった。よって当該課題を解決するために、インスタンス群とクラス比率ラベルから類似比率損失 (Similarity Proportion Loss) を提案し、クラスの順序性を考慮した比率学習法「順序尺度類似比率学習 (Ordinal Scale Learning from Similarity Proportion; OSLSP)」を提案した。提案手法は (1,2) の問題を従来法に比べて改善されていることが特徴量空間での可視化で確認され、分類精度結果からも提案法の妥当性を得た。第 5 章ではこれまで LLP が WSI 解析に適用された事例を紹介しつつ、新しく提案された OSLSP がこういったデータに応用されるかの展望を述べる。

本論文では数式の表記は以下に従う.

- ベクトルはゴシック体で統一表記する.

例: $\mathbf{x} = [x_1, x_2, \dots, x_n]$

- スカラはイタリック体である.

例: クラス数 $K = 5$

- 集合はダブルストライクキャピタルで統一表記する.

例: 自然数集合 \mathbb{N} , 整数集合 \mathbb{Z} , 実数集合 \mathbb{R} , クラス $\mathbb{C} = \{red, blue\}$

第 2 章 明示的特徴量による筋組織画像解析

2.1 緒言

加齢に伴う筋力の低下であるサルコペニア [34] や骨格筋が壊れやすく再生されにくい筋ジストロフィー [35] といった疾病は人間の運動能力を削ぎ、運動機能の低下は自立した生活を困難にさせ、生命活動の制限を引き起こす。

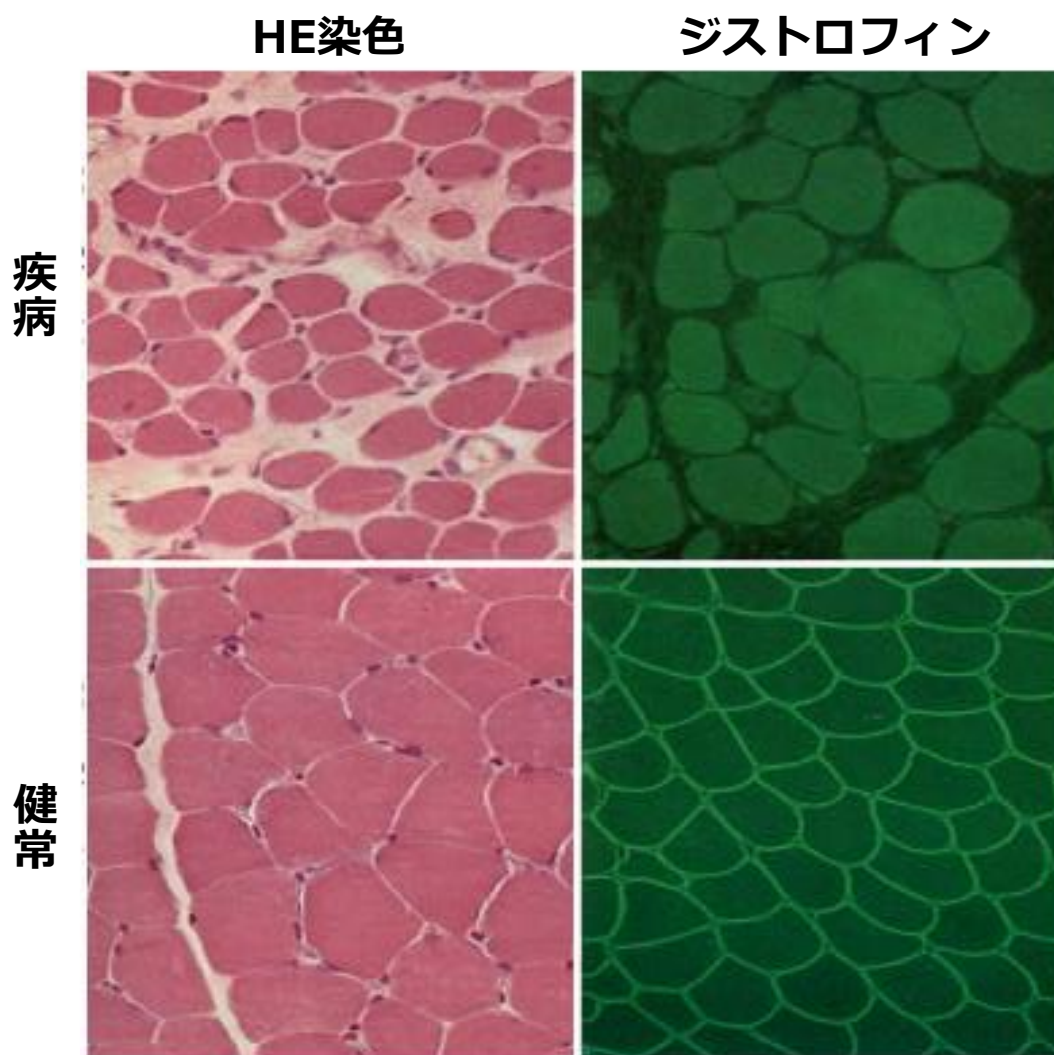


図 2.1 筋ジストロフィーと健常時の筋組織の比較 [4]

骨格筋の疾病を打破する手段の一つとして筋組織の筋肥大や再生機能のメカニズムを解明し [36-38], 適切な薬剤処方によって疾病の進行を妨げるというものがある [39]. これらの薬学研究では筋組織切片を染色したのちに画像化し、組織内の筋線維の形状や大きさが

ら疾病の進行状態を評価する．例えば図 2.1 [4] が示すように筋ジストロフィー (muscular dystrophy) と健常なマウスの筋組織を一般的な組織全体の可視化に用いるヘマエキトシンエオジン (Hematoxylin and Eosin ; HE) 染色とジストロフィン (dystrophin) 染色の 2 つの染色法で見比べる．ジストロフィン染色とはその名の通り筋ジストロフィーの診断や研究で用いられる染色法で，細胞膜に存在するジストロフィンと呼ばれるたんぱく質の可視化に用いる．健常なマウスは筋線維で組織内が満ちているが，筋ジストロフィーのマウスは白い脂肪部が散見されることが HE 染色画像の比較から判る．また線維の大小についても健常なマウスは均一な線維面積である一方，筋ジストロフィーのマウスは線維の面積にばらつきがみられることがジストロフィン染色画像から判る．このように筋組織の染色画像を比較することで特徴を見出し疾病の進行を判断することができる．

またこのような病気状態のマウス (muscular dystrophy X-linked: mdx) の解析 [6, 8, 9, 40] のみならず，健常なマウスを用いた遅筋と速筋の分布解析 [10, 11]，筋肉トレーニングの経過観察 [12]，破壊された筋組織の再生観察 [13] など画像解析の目的は多岐にわたる．本章では WSI を用いた全組織領域の再生段階評価において，画像化の手順，染色法およびコンピュータ解析手法の検討を行う．

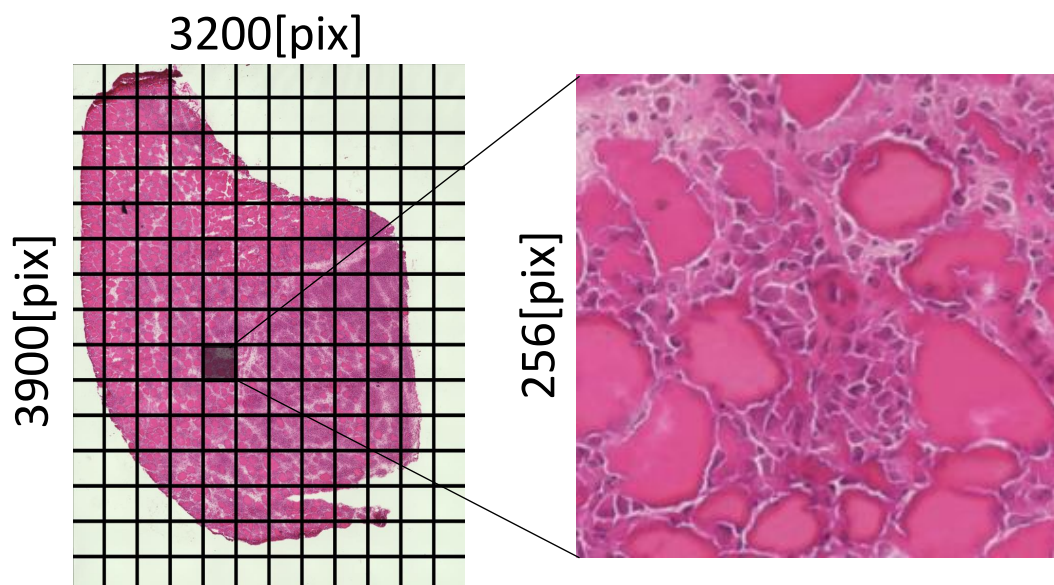


図 2.2 筋組織に含まれる大量の細胞

2.2 筋組織画像

2.2.1 画像化と解析

画像化にあたっては，前脛骨筋 (Tibialis anterior) などを切断しスライス状にした組織切片を染色してから行う．画像化した組織標本例の図 2.2 のように一つの筋組織には数千本の筋線維やその他の細胞で構成されており，これらの細胞の面積，Feret 比や円形度といった特徴量を図 2.3 のように一つ一つ計算することが定量的解析に相当する．同じ組織内であっても各線維の状態にはばらつきがあるため，組織の状態を評価するときは局所的な筋線維や細胞のみに着目するのではなく，組織全体を解析することが望ましい．

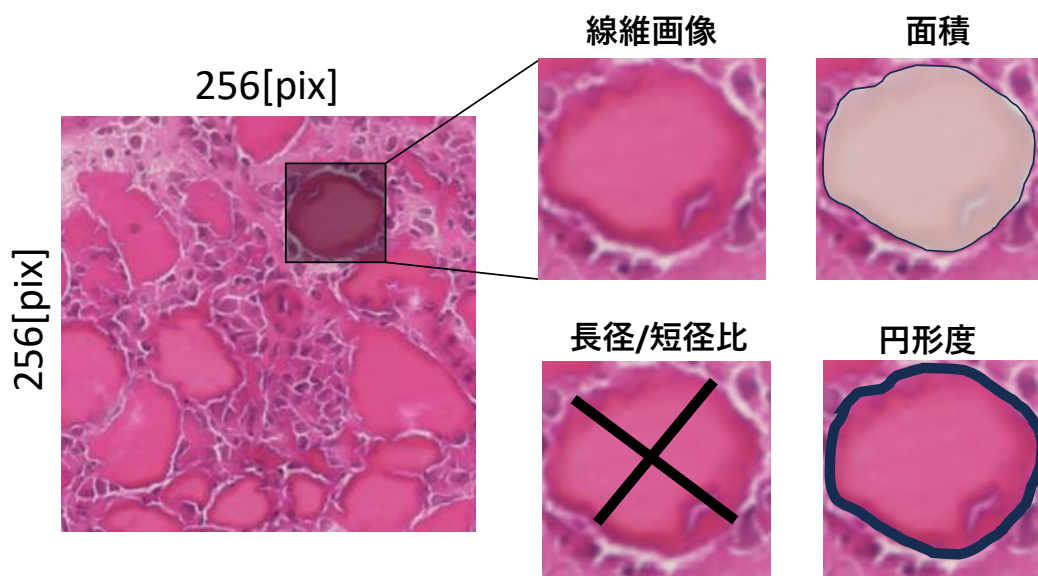


図 2.3 筋線維の特徴量例

図 2.3 の特徴量をコンピュータで自動計算する場合図 2.4 に示すセグメンテーションと呼ばれる細胞領域の検出が必要となる．

組織評価について例えばセグメンテーションをしたのちに筋線維の断面積 (Cross Sectional Area: CSA) を計算するという手法がある．一つの切断面から計算される線維面積は解剖学的断面積 (Anatomical Cross Sectional Area: ACSA) と呼ばれており，図 2.5 が示すように平行筋であれば生理学的断面積 (Physiologic Cross Sectional Area: PCSA) と等しいが，羽状筋である場合は羽状角 θ に応じて下式のように $\cos\theta$ による補正をかけることで計算される [41]．

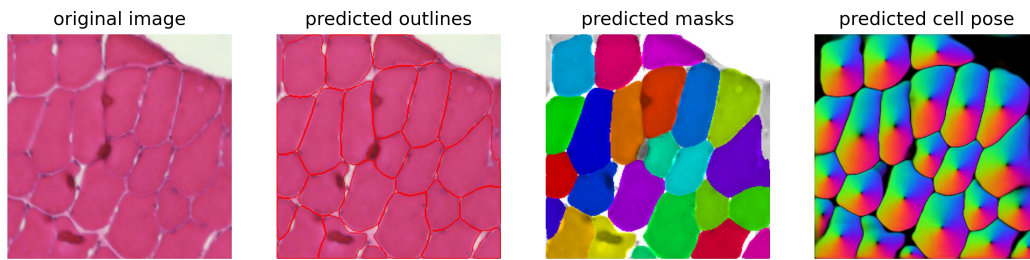


図 2.4 筋組織セグメンテーションの例

$$\text{PCSA}[cm^2] = \frac{\text{筋質量}[g]}{\text{筋密度}[\frac{g}{cm^3}] * \text{筋線維長}[cm]} * \cos\theta \quad (2.1)$$

CSA は筋出力と正の相関があることが知られており，筋組織画像から直接線維面積を計算することで筋出力を近似的に算出することが可能となる．なお筋密度は哺乳類の場合， $1.05g/cm^3$ である [41]．

平行筋

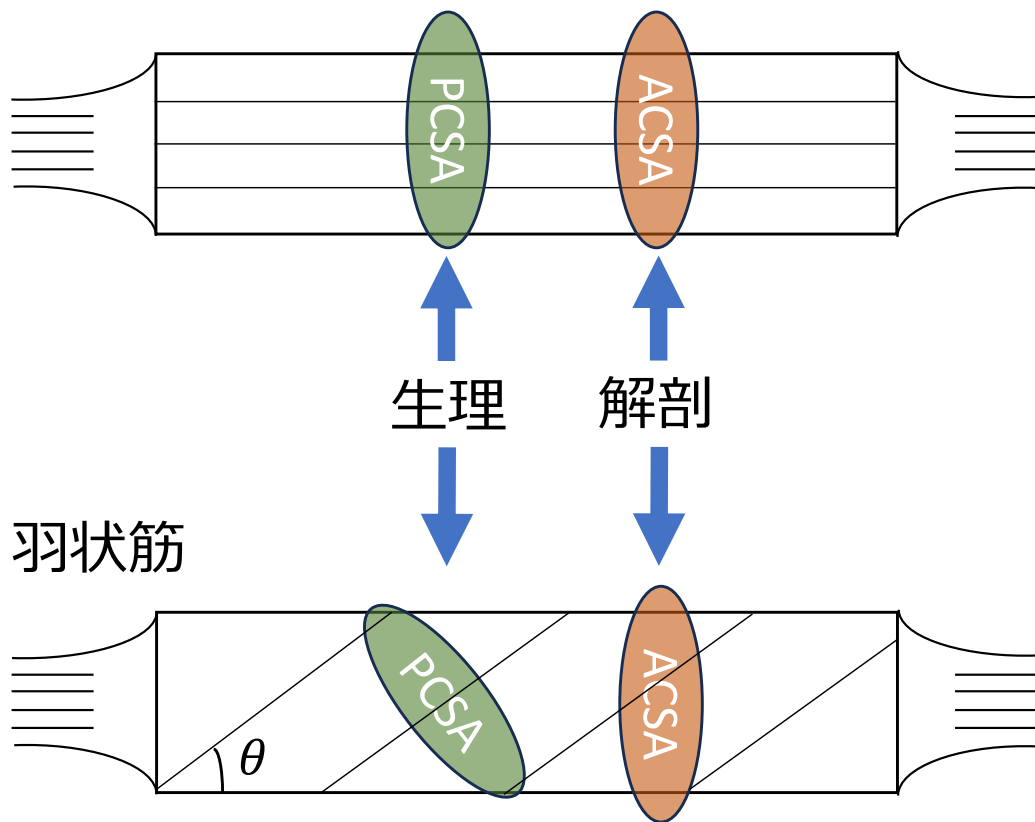


図 2.5 平行筋と羽状筋の CSA の違い

2.2.2 筋組織の再生過程

筋組織は損傷しても適切な条件下であれば再生することが知られており、1961 年の筋サテライト細胞の発見 [42] を皮切りに再生メカニズムの研究が加速した。この筋組織再生の解析は老化や疾患における筋力低下のメカニズムや治療法に関する研究に役立つことが期待される。筋組織の再生を解析する手段としてコブラ (*Naja pallida*, *Naja nigricollis*) の持つ毒 (cardiotoxin; CTX) を注入し選択的に筋組織の壊死と再生を誘発する手法がある [43](図 2.6)。

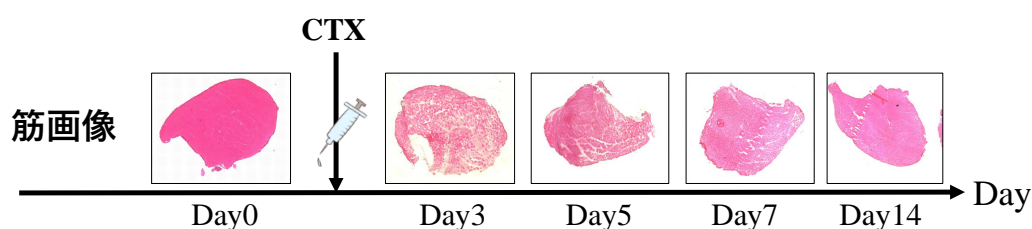


図 2.6 CTX 注入による筋組織再生の観察 [5]

CTX はコブラ (*Naja pallida*, *Naja nigricollis*) が持つヘビ毒の一種であり、毒素によって筋線維のみを選択的に傷害する [43]。再生に必要な筋サテライト細胞や血管、神経、コラーゲンなどの基底膜は保たれることから、すみやかな再生が誘導される。損傷が起こると、まず筋組織を構成する主たる細胞である筋線維が壊死する。その後、浸潤してきた炎症性細胞により壊死した筋線維が除去され、ゴーストファイバーと呼ばれる筋線維を囲っていた基底膜だけが保持された状態となる [44]。壊死した筋線維由来の damaged-myofiber-derived factors (DMDFs) [45] などによって筋サテライト細胞の活性化が誘導され、増殖したサテライト細胞は萌芽細胞 (myoblast) と呼ばれる細胞になる。一連の再生過程によって増殖した筋芽細胞は細胞分裂を停止し、隣接する筋芽細胞同士が融合していくことで多核の筋管細胞 (Myotube) となる。筋管細胞がさらに成熟することで筋線維へ成長し、再生が完了する頃には成熟した筋線維によって隙間が埋め尽くされ、組織内でひしめき合うようになる。こういった一連の流れは図 2.7 のように進行する。

2.2.3 染色方法の違い

本論文では表 2.1 の通り、筋組織の染色法を HE 染色と免疫染色 (immunostaining) に大別する。HE 染色は染色手順が簡便かつ迅速に行うことができ、組織全体を観察に適し

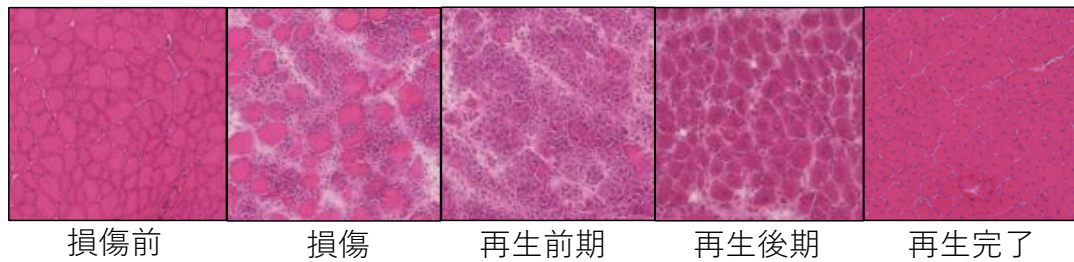


図 2.7 筋組織の再生過程

ている．一方免疫染色は染色目標となる細胞の抗体を用いることで特定の分子または細胞の可視化に優れる．例えばラット抗マウスラミニン α 2 (Enzo, Clone 4H8-2, Cat# ALX-804-190-C100) を用いた免疫染色の一種であるラミニン染色によって筋線維の細胞膜を特異的に染色することを行う．観察する際は物体の反射光をレンズに通して拡大する光学顕微鏡や，励起光と呼ばれる特定の波長を検体に照射し，発生した蛍光を観察する蛍光顕微鏡が挙げられる．ただし同じ組織切片に HE 染色と免疫染色を両方行うことはできないため，仮に同じような切片に異なる染色を施す場合は同じ下腿筋 (tibialis anterior; TA) の連続した切片を用いる．

表 2.1 染色法の比較

	HE 染色	免疫染色
手軽さ	○	×
特定の細胞の可視化	×	○
組織全体の把握	○	×
領域検出の難易度	△	○

また図 2.8の染色画像の違いからも判るように免疫染色の一つであるラミニン染色では細胞膜の輪郭がはっきりしていることに比べ，HE 染色は膜の領域に色のグラデーションがあり境界線が判然としない．これによって画像処理の一つである領域検出 (セグメンテーション) の難易度が上がってしまう．ただし HE 染色では線維自体や周りの細胞についても染色がなされていることから，線維の周辺環境の情報を得られることが判る．特に回復過程では筋線維以外の細胞が回復度合いの判定には必要であり，ラミニン染色から得られる線維膜の情報だけでは図 2.7の分別は困難である．よって本論文では主に HE 染色画像から回復過程段階の推論を試みる．

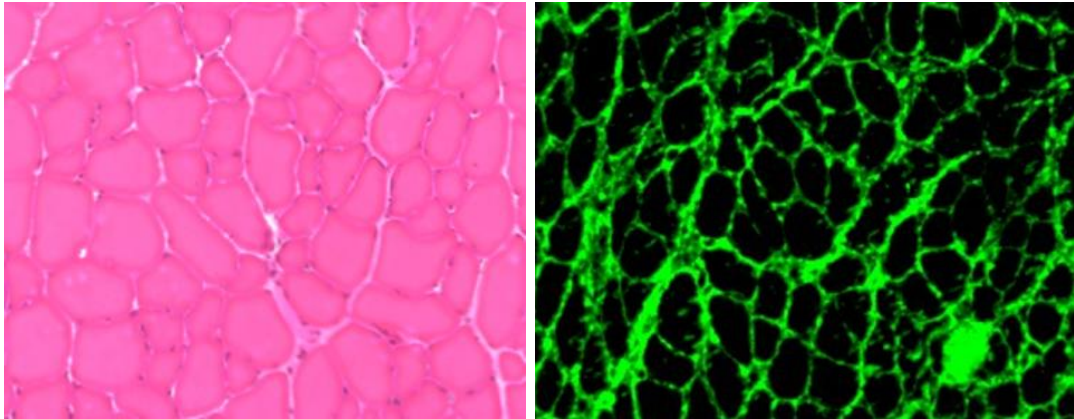


図 2.8 HE 染色とラミニン染色

2.3 関連研究

2.3.1 従来の筋組織画像のコンピュータ支援

本節では染色，解析目標および手法の観点から，コンピュータ支援による筋組織画像解析の関連研究について述べる．表 2.2の通り筋組織画像のコンピュータ解析は主にラミニン染色された画像を中心に研究されてきた [6, 7, 9–13, 46–48]．この理由として，筋組織の解析では主に Cross-sectional area(CSA) を計算することが求められており，健全な筋線維であればラミニン染色を施すことで線維の輪郭がはっきりと判るため，線維面積がコンピュータ解析上求めやすいからである．特に用いられる解析手法は図 2.4のような領域検出 (セグメンテーション) であり，組織内の各細胞の輪郭をなぞるように領域を検出する．セグメンテーションによって図 2.3のような筋線維の明示的な特徴量を計算することが可能になる．HE 染色に比べてラミニン染色画像のようにバイナリ化された画像ではセグメンテーションに必要な情報に限定されているため扱いやすいと言える．またこのようにして得られた特徴量を元に病理マウスの判定 [6–9]，遅筋と速筋の分類 [10]，細胞の分類 [11]，筋肉トレーニングの経過観察 [12]，そして再生過程の観察 [13] などに用いられてきた．

個別の手法ごとに見ていくと，SMASH [46]、MuscleJ [6]，および Muscle2View [11] では，酸化系アデノシン三リン酸 (adenosine tri-phosphate; ATP) 供給によってエネルギーを得る遅筋 (slow-type oxidative fibers) と解糖系 ATP 供給によってエネルギーを得る速筋 (fast-type glycolytic fiber) などに線維種類を分け，セグメンテーションに

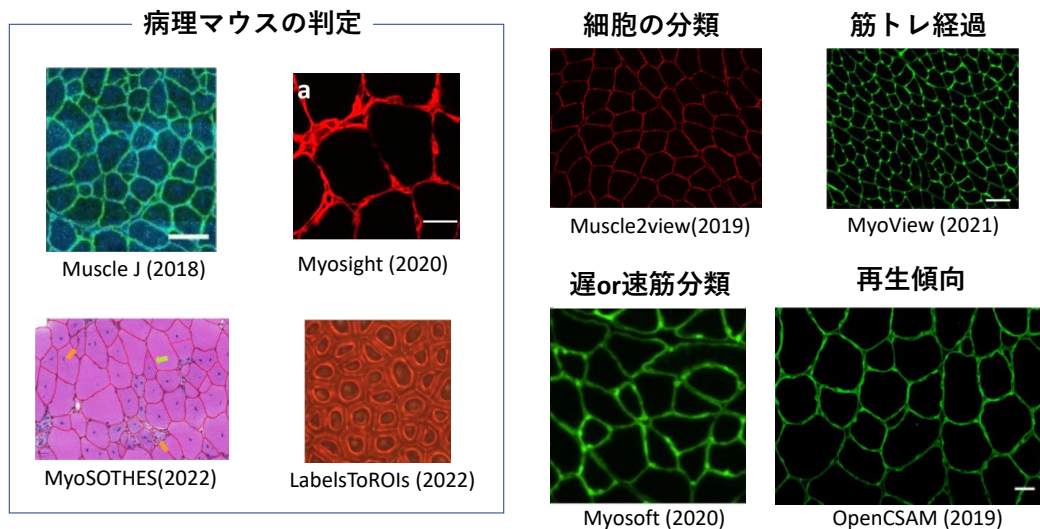


図 2.9 関連研究の筋組織画像 [6–13]

よって得られる明示的な特徴量で分類を行っている．OpenCSAM [13] では CTX により壊死と回復を誘発されたラミニン染色の筋線維セグメンテーションの精度に着目している．OpenCSAM [13] は特にコブラの毒 (CTX) によって引き起こされた壊死および回復している筋組織におけるラミニン染色筋線維のセグメンテーションの精度に焦点を当てている．MyoView [12] では High-intensity interval training と呼ばれる筋肉の強化プロトコルを行ったマウスの組織について観察しており，日数が経過したマウスの筋線維を人のマニュアルと同等の精度でセグメンテーションすることに成功し，筋肉トレーニングしたマウスの平均 CSA が大きくなったことを確認した．また MyoView [12] は，openCSAM [13]，MuscleJ [6]，SMASH [46]，および MyoVision [48] などの他のセグメンテーションツールと比較して，健全な筋線維のセグメンテーションにおいて最も高い精度を達成したことを報告している．Myosoft [10] では 筋肉の代謝や収縮に関する性質 (metabolic and contractile properties) に基づいて線維を種類分けし，面積・円形度 (Circularity)・最小フェレ径 (Min Feret) の特徴量を用いて分類することに成功している．またラミニン染色のために開発されたソフトウェアを他の染色法に適用することは図 2.10の予備実験結果からも判る通り困難であるため，LabelsToRois [9]，Laghi *et al.* [49]、Liu *et al.* [50]、および MyoSOTHES [8] などの代替ソフトウェアが開発されている．

MyoSOTHES [8] は本研究と同様に HE 染色された画像に焦点を当てているが，解析対象は健常な筋線維に限定的であり，回復過程で現れるゴーストファイバー，筋芽細胞，

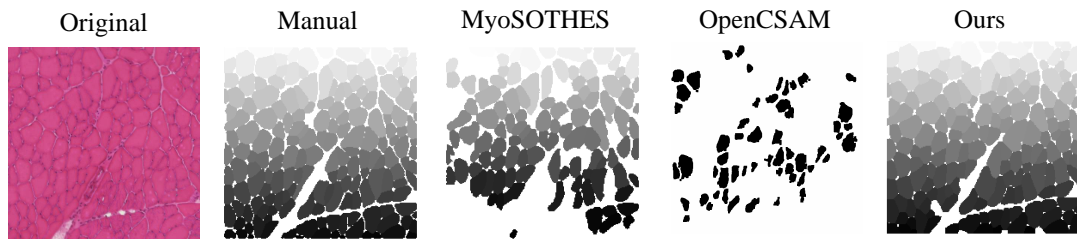


図 2.10 ラミニン染色用ソフトの HE 染色画像への適用例

筋管細胞など他の細胞をセグメンテーションすることはできていない。

セグメンテーションの手段にはおおまかに画像特徴量について閾値や条件式を組み合わせるアルゴリズム方式と既存ソフトウェアである cellprofiler [51] を採用する手段、あるいは機械学習を基盤としたモデル Cellpose [52] を用いるものがある。

2.3.2 再生評価における HE 染色の優位性と課題

表 2.2に示す通り、これまでの研究の多くは無傷な筋組織に免疫染色を施した画像を主な対象としており、回復経過中の HE 染色画像をその解析対象としてこなかった。免疫染色は細胞膜の可視化に適しており定常状態のセグメンテーションがしやすい反面色味情報は失われるため、回復経過中の組織を解析する場合にはゴーストファイバー (ghost fiber)、萌芽細胞 (myoblast) や筋幹細胞 (myotube)、そもそも筋線維でないものを誤って定常状態筋線維 (myofiber) と解析してしまうリスクが伴う。一方最も一般的かつ手軽な HE 染色が施された筋線維は免疫染色にはない線維の周辺・輪郭情報 (特にサテライト細胞の分布やゴーストファイバー (ghost fiber) のもやのような輪郭など) や色味を拾うことが可能であり、各線維の再生状態を含めた評価に適する。ただ画像解析側にとっては色味や周辺情報が増えることや画像上で細胞膜のエッジが際立ちづらい都合から画像処理としての難易度が上がってしまう。これまでの研究では細胞壁が損傷していない安定状態な筋線維のラミニン染色 [9–12] や HE 染色 [8], Picrosirius red staining 画像 [49] のセグメンテーションを扱ってきたが、損傷初期から回復末期の HE 染色された筋線維についてはその困難性から画像解析の対象外であった。当課題の解決には 2 つの段階があり、1 つ目は Myofiber 以外の線維のセグメンテーション、2 つ目は線維の状態推定すなわち分類である。

1 点目のセグメンテーションについて、これまでの研究では HE 染色画像において無傷な線維のみを領域検出することは確かめられていたが [8], 回復期における筋芽細胞やゴーストファイバーなどの他の細胞については対象外であった。また図 2.10が示すように、ラ

表 2.2 筋組織画像コンピュータ解析の関連研究

名前	染色法	目標	解析項目	手段
SMASH [46]	ラミニン	健常, mdx	分類 (fiber types)	アルゴリズム
MuscleJ [6]	ラミニン	健常, mdx	分類 (fiber types)	アルゴリズム
MuscleAnalyzer [47]	ラミニン	健常, mdx	領域検出	CellProfiler [51]
MyoVision [48]	ラミニン	Training, disuse atrophy	領域検出	アルゴリズム
Muscle2View [11]	ラミニン	健常	分類 (fiber types)	CellProfiler [51]
OpenCSAM [13]	ラミニン	回復過程 (days 0,8-24), mdx	領域検出	アルゴリズム
MyoSight [7]	ラミニン	健常, mdx	領域検出	アルゴリズム
Myosoft [10]	ラミニン	健常, mdx	分類 (fiber types)	アルゴリズム
MyoView [12]	ラミニン	Training (days 0,28,56)	領域検出	アルゴリズム
LabelsToRois [9]	ラミニン, etc.	健常, mdx	領域検出	Cellpose [52]
Laghi <i>et al.</i> [49]	Picrosirius red	健常, mdx	領域検出	CellProfiler [51]
Liu <i>et al.</i> [50]	HE	健常, mdx	領域検出	アルゴリズム
MyoSOTHEs [8]	HE	健常, mdx	領域検出	Cellpose [52]
Ours	HE	回復過程 (days 0, 3-14)	分類 (回復段階)	Cellpose [52], DINO [53]

ミニン染色画像によって学習されたセグメンテーションモデル [13] を HE 染色画像に直ちに転用することは難しい。よって再生過程に出現する細胞についてはファインチューニングを行う必要がある。

2 点目の回復段階の分類について、クラス分類を行うときセグメンテーションから得られる選択的な明示的特徴量データは限定的である。たとえば Myosoft [10], MuscleJ [6], Open-CSAM [13] などがあり、これらは HE 染色ではなくラミニン染色された筋組織を対象としており、セグメンテーションされた筋線維から面積や円形度などの特徴を抽出する。Open-CSAM [13] は、CTX 注射後の 8 日目から 28 日目までのラミニン染色された画像の断面積 (CSA) を計算する。ユーザーは経過日数、マウスの年齢、健康状態に応じて細胞の大きさや円形度の閾値を手動で設定する必要がある、誤分類されたり未検出の筋線維は手動で追加しなければならない。この処理プロセスはユーザーの専門知識のレベルや撮影条件によって影響を受ける。Myosoft [10] も Feret アスペクト比や最小 Feret 距離などの形状に関連する特徴を使用して筋線維のタイプを判定する。こういったアプローチは手動での特徴選択がドメイン知識に依存するためドメインギャップに対しての頑健性が不足していると言える。例えば図 2.11 のように地面からの高さや葉っぱの数のみで成長が定義される画像であれば、成長を分類する際に必要な特徴量は 2 つである。一方回復経過中の筋組織ではこういった明示的な特徴量がない。再生程度の指標となる定量的な数値が定義されていないことが、再生評価の課題である。

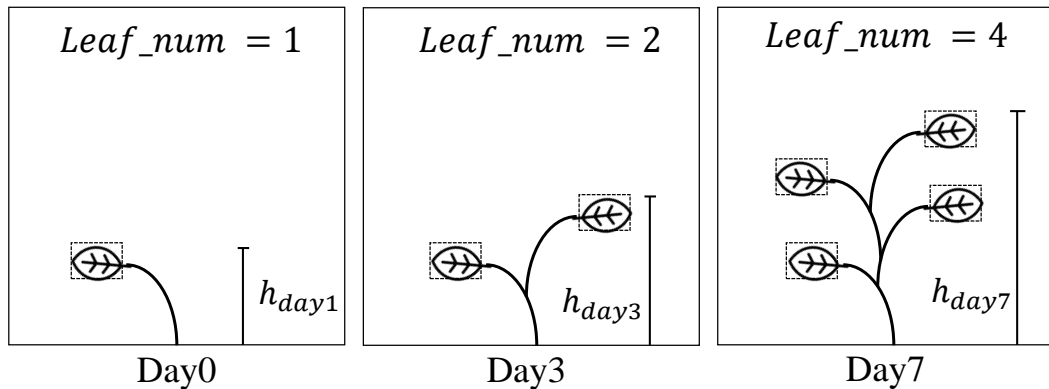


図 2.11 明示的特徴量を持つ画像例

2.4 データセット作成法

再生過程筋組織の細胞及び周辺領域の再生段階を局所的に評価する実験を行うための、解析対象となるマウスの回復過程画像データセットの作成方法について述べる。表 2.3 に

表 2.3 本論文で使した筋組織画像一覧

目的	注入	Day0	Day3	Day5	Day7	Day14
学習用データ	CTX	5	6	6	6	3
アノテーション済み検証用データ	CTX	1	1	1	1	1
比較実験用データ	CTX	3	3	3	3	3
比較実験用データ	Glycerol	0	3	3	3	3
連続切片の HE 染色データ	CTX	0	1	1	1	0
連続切片の免疫染色データ (Laminin)	CTX	0	1	1	1	0
連続切片の免疫染色データ (DAPI)	CTX	0	1	1	1	0
連続切片の免疫染色データ (MyoD)	CTX	0	1	0	0	0
連続切片の免疫染色データ (eMyHC)	CTX	0	0	1	1	0
連続切片の免疫染色データ (Perilipin)	Glycerol	0	0	0	1	1
連続切片の免疫染色データ (Collagen)	Glycerol	0	0	0	1	1

において、学習用データとはモデル学習のために使したデータである。アノテーション済み検証用データとはモデル評価のために専門家によるアノテーションを行った検証用データ (Test data) である。比較実験用データとは、作成したモデルを用いたソフトウェア MyoRegenTrack の性能評価のためのデータである。CTX とグリセロール (Glycerol) の異なる薬剤による回復過程の比較実験を行う。CTX とグリセロール (Glycerol) の回復傾向が異なることが知られているため、本ソフトウェアによる解析を行いその傾向の違いを確認する。上記 4 つのデータ項目については HE 染色画像を用いた。

連続切片のデータとはアノテーションの正確性を担保するために同じ前脛骨筋 (Tibialis Anterior; TA) の連続切片に異なる染色を行ったデータである。取得した画像の詳細は表 2.3 に示す。

2.4.1 動物処理

画像化に用いた C57BL/6J マウスは、神奈川県横浜市にあるチャールズリバー研究所から購入された。マウスは温度 $24 \pm 2^\circ\text{C}$ 、湿度 $50\% \pm 10\%$ に制御された環境下で、12 時間の明暗サイクルで飼育された。飼育中のマウスには滅菌された標準飼料 (DC-8, 日本クレア, 東京, 日本) と水が自由に与えられた。筋組織の再生を誘導するために、 $100 \mu\text{L}$ の CTX (カーディオトキシン; 生理食塩水中 $10 \mu\text{M}$; Latoxan, Valence, France) ま

たは 50% v/v のグリセロールを 29 ゲージの針を使用して前脛骨筋に注射した [31–33]. この際、麻酔にはメデトミジン、ミダゾラム、およびブトルファノールのカクテルを使用した. [31–33].

マウスの安楽死には、熟練した研究者が大阪大学の動物実験委員会の指導に従い、頸椎脱臼法を使用した. その後、前脛骨筋 (Tibialis Anterior; TA) は解剖され、練り合わせたトラガカントゴム (和光純薬工業, 大阪, 日本) を使用してコルクに固定し、液体窒素で冷却したイソペンタン (和光純薬工業) で 1 分間フラッシュ凍結した. イソペンタンを蒸発させるため、ドライアイス上で 1 時間インキュベーションした後、筋肉は密閉容器に入れて -80 °C で保管した. 10 μ m の厚さの横断的な凍結切片を作成し、ヘマトキシリンおよびエオシン溶液で染色した.

免疫染色データについては TA 筋の横断凍結切片 (6 μ m 厚) を MyoD 染色の場合は 4% パラホルムアルデヒド (PFA) で固定し、胚性ミオシン重鎖 (eMyHC) 染色の場合は冷却アセトンで 10 分間固定した. 5% 脱脂乳でブロッキングを行った後、一次抗体を用いて 4° C で一晩染色した. eMyHC 染色では、内在性のマウス IgG 抗体をブロックするために M.O.M. キット (Vector Laboratories, Burlingame, CA, USA) を使用した. 本研究で使用した一次抗体は、ラット抗マウスラミニン α 2 (Enzo, Clone 4H8-2, Cat# ALX-804-190-C100), ウサギ抗マウス MyoD (Abcam, Cat# ab133627), マウス eMyHC (DSHB, Clone F1.652), およびウサギ抗コラーゲンタイプ I (Bio-Rad, #2150-1410) 抗体である. 洗浄後、Alexa Fluor 488, 546, または 647 (Molecular Probes, Eugene, OR, USA) で標識された二次抗体でインキュベーションを行った. 洗浄したサンプルは、DAPI を含む VECTASHIELD 封入剤 (Vector Laboratories, #H-1200) で封入した. なお切片の厚さを正確に保つため、組織切片の厚さを変更する際には、最初の 1 枚または 2 枚の切片を廃棄した.

染色された組織画像は、Plan Apochromat (キーエンス社) および 20x 対物レンズ (ニコン社) を使用して BZ-X Analyzer で撮影した. 撮影された画像の幅は 2877 から 4606, 縦は 2720 から 4355 で、合計 12.5 メガピクセルのサイズである. これらの画像は、CTX 注射後の経過日数に関する情報とともにタグ付き画像ファイル形式 (TIFF) で保存した.

取得した画像の詳細は表 2.3 に示す. 学習用データと Day0 の比較実験用データでは、各マウスの両足を切断し両方の前脛骨筋の撮影を行ったが、それ以外の場合では 1 つの画像が 1 つのマウスサンプルに対応している. 凍結処理や染色に誤りがあった画像はデータセットから除外しており、例として学習用データと比較実験用データが該当する. した

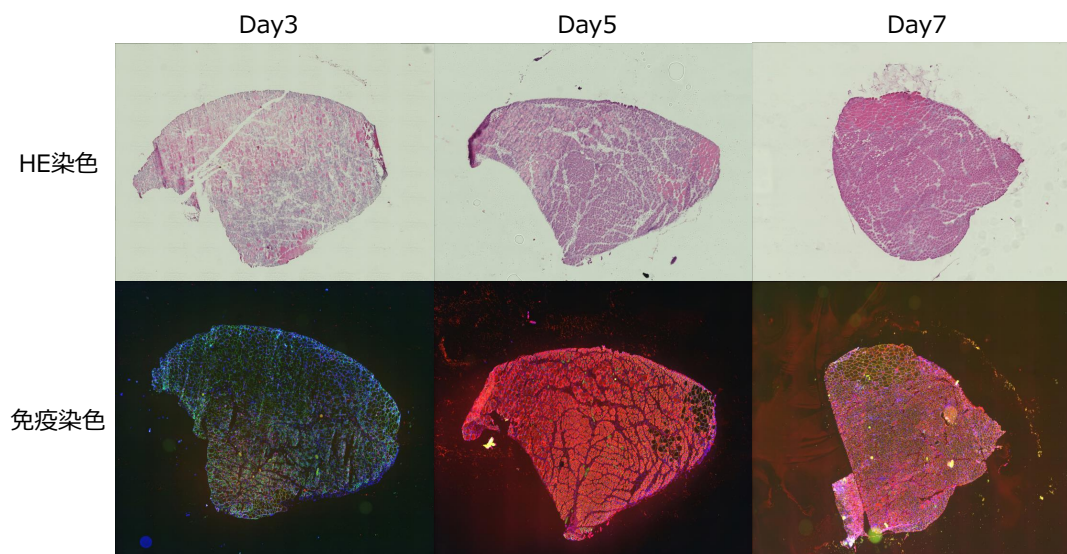


図 2.12 HE 染色と免疫染色の画像例

がって、学習用データは 14 匹のマウス、アノテーション済み検証用データは 5 匹のマウス、CTX またはグリセロールの注射を受けた比較実験用データは 26 匹のマウスで構成されている。連続切片の免疫染色画像については同じマウスの連続切片を使用する関係から同じ注入かつ日付である画像は同じマウスから切り取られた TA の画像である。よって連続切片画像は CTX で日付ごとに 3 匹、Glycerol は 2 匹である。

すべての実験動物に対する手順は、大阪大学の実験動物使用委員会（承認番号：R02-3）によって承認され、すべての方法は関連するガイドラインや ARRIVE ガイドラインに基づいて実施された。

2.4.2 検証用データのアノテーション

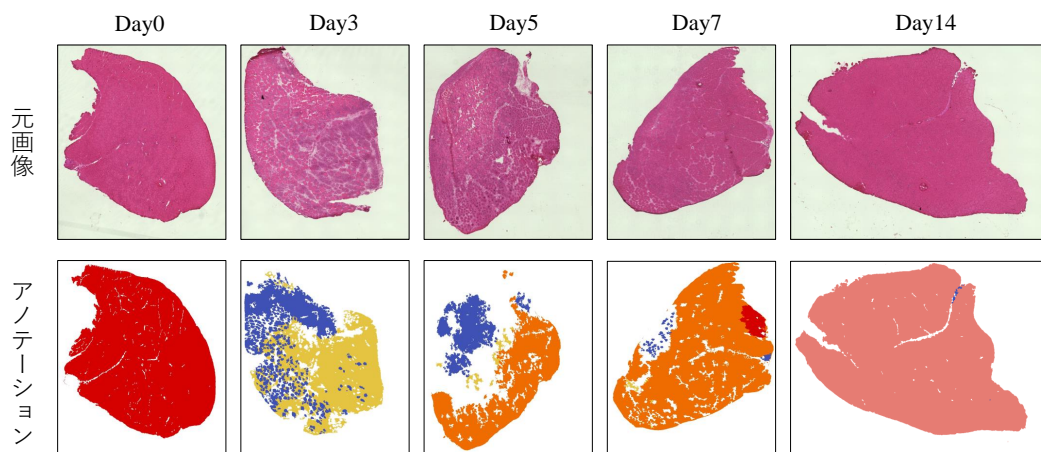


図 2.13 専門家によるクラスアノテーション

図 2.13の通り CTX 注入データのうち Day0,3,5,7,14 の画像を一枚ずつ選択し線維状態を以下の 5 つのフェーズ: 安定 (赤), 損傷 (青), 再生前期 (黄色), 再生後期 (オレンジ), 再生完了 (ピンク) に分け, Labelboxを用いて専門家によるマニュアルでの領域アノテーションを行った. 判断のつかない箇所についてはアノテーションをせず白塗りとした. 筋組織の外側の領域も白塗りでマスクした. 青色クラスは MyoD [54] *⁶ の発現がないまたは低い初期の再生を示し基底膜はあるが核は見られない領域を示す. 黄色クラスは顕著な MyoD 発現が見られる再生中期を表す. オレンジクラスは中央に筋核を持つ小型 (eMyHC-high) および大型の筋管 (eMyHC-low) の両方を含む再生後期を示す [55]. 赤色は損傷する前や回復が完了した段階である.

2.5 予備実験

2.5.1 学習データの拡張

すべての学習用データはコンピュータ上メモリで扱うために 256 ピクセルの正方画像の切り取った. すべての正方画像データは OpenCV4.8.1(Python3.7.13) の rotate 関数を用いて 90, 180, 270 度の回転画像と Flip 関数を用いた鏡像反転で 8 倍にデータ数を増やした. これは細胞は画像上ならば回転させてもその形態特徴は変化しない事前知識を利用している. 補足ではあるが例えば人であれば画像を回転させた場合そ

*⁶ MyoD は再生医療の始まりに発見された転写因子であり筋分化の運命を決定する

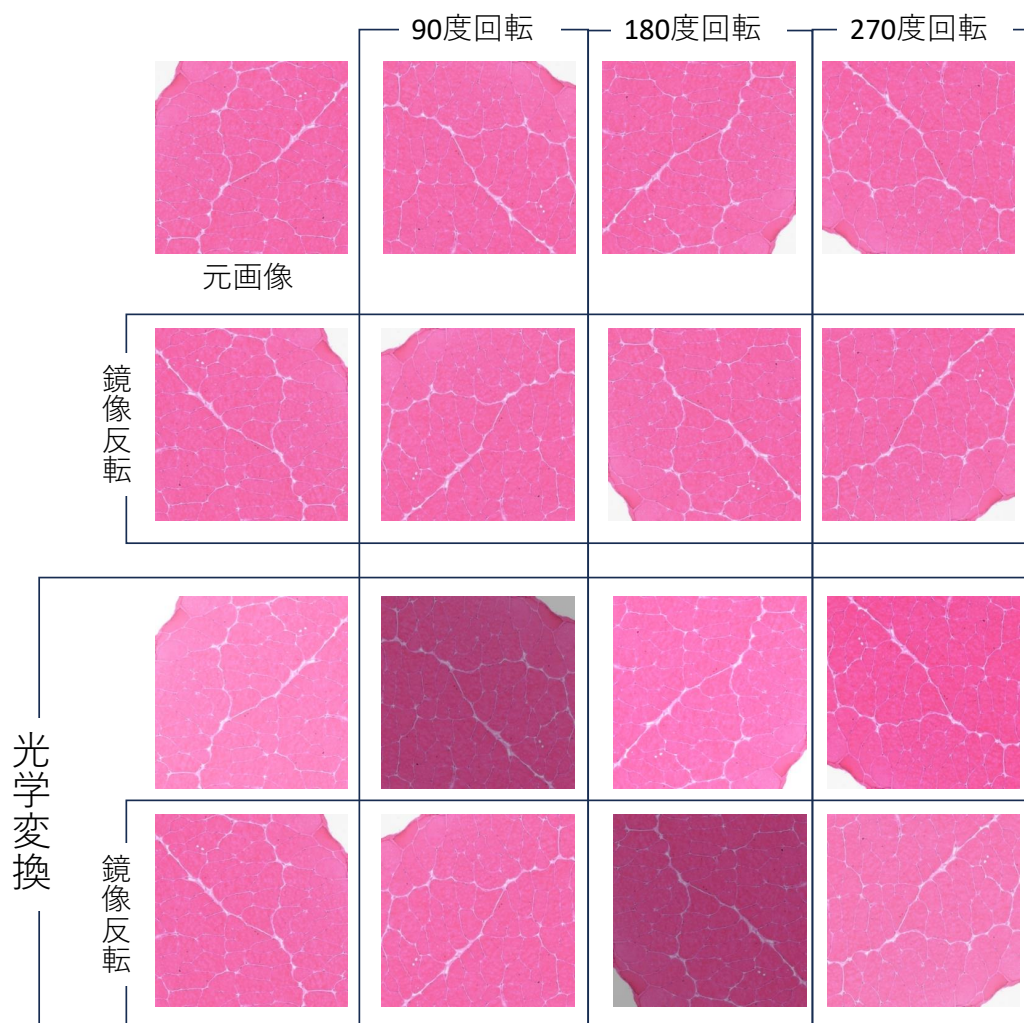


図 2.14 画像データ拡張例

の特徴は失われるが，鏡像反転であれば目や鼻，口の位置関係などに齟齬は生じない為可能であるとされている．撮影画像は撮影場所の照明，Light-Emitting Diode(LED)の光量，カメラの露光時間や色味の設定，顕微鏡や鏡筒など撮影ハードウェアの型番，検体の濃淡などの影響を受ける．こういった環境依存の条件をドメインと言い，画像に現れる差をドメインギャップと言う．一般にこれらの撮影条件の統制を執ることは難しく，画像処理側でドメインギャップの吸収を行う必要がある．そこでランダムな光学条件で撮影されたことを仮定したデータ拡張を albumentations v1.3.1 [56] で行い更に画像数を 2 倍に増やした．使用した Compose は RandomBrightness($p=0.5$), RandomContrast($p=0.5$), RandomContrast($p=0.5$), RandomGamma($p=0.5$) である．従ってデータ拡張は図 2.14 の通り合計で 16 倍となる．

2.5.2 明示的特徴量による分類結果

これまでの筋組織画像のコンピュータ解析研究 [6–13, 46–50] では，細胞領域の検出 (セグメンテーション) をしたのちに細胞の明示的特徴量を計算することで，細胞の種類や状態を分別していた．そこで本節では同様にセグメンテーションで検出した各細胞の明示的特徴量を導出し，日付や WSI ごとに並べることでその傾向を見る．また明示的特徴量を用いた分類を試み，その成果を確認する．

得られた表 2.3 に示す WSI を Cellpose [52] を用いてセグメンテーションをし，画像ごとに細胞数，細胞面積の平均，円形度の平均について先行研究の手法によるパイロットスタディを行った．細胞のセグメンテーションには Cellpose [52] の cyto モデル (Version: 2.0.3, チャンネル=[0,0], ミニサイズ=40, 残りはデフォルト設定) を用いた．図 2.15, 2.16, 2.17, 2.18, 2.19 ではセグメンテーションの結果から OpenCV (バージョン 4.8.1) を用いて各検出オブジェクトの細胞数，細胞面積，feret 比 (長径短径比)，輪郭長，円形度を計算し，WSI の画像名ごとに平均化した．

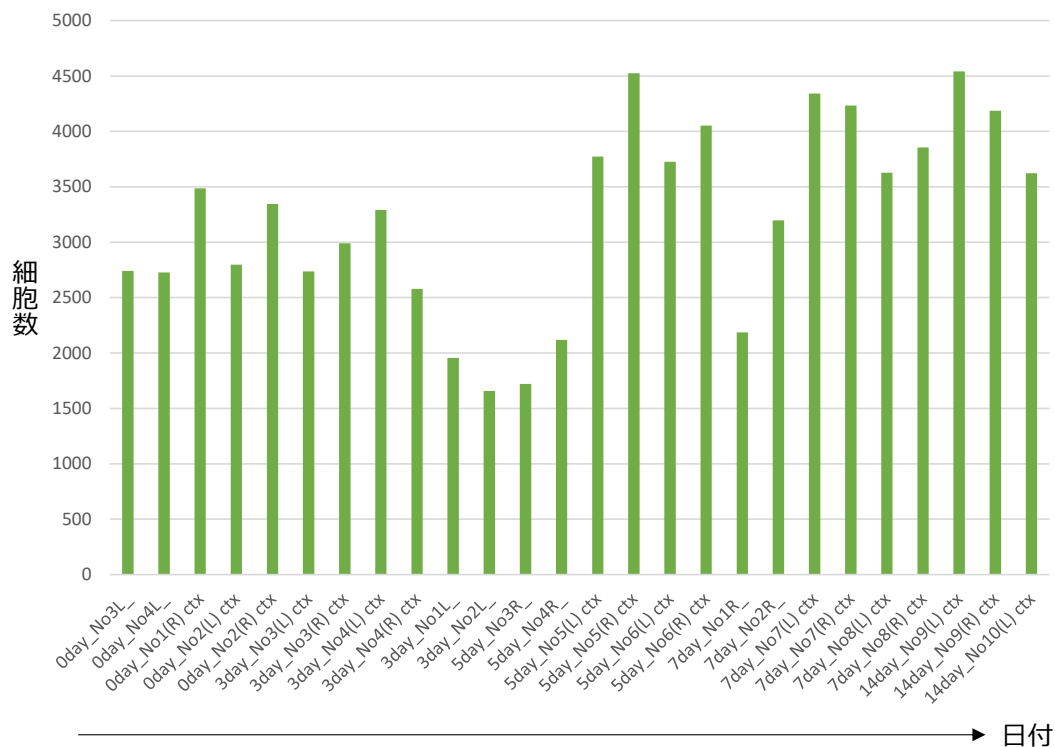


図 2.15 各画像の細胞数

図 2.15 の細胞数はセグメンテーションの検出したオブジェクト数であり，各 WSI の

図 2.16, 2.16, 2.17, 2.18における標本数を示している。細胞数には筋線維だけでなくゴーストファイバー、萌芽細胞、筋幹細胞などが含まれており、一般に組織切片の個体差に依存する数なため、細胞の数と回復の度合いを結び付けられる因果関係はない。筋線維は萌芽細胞や筋幹細胞よりも面積が大きいため [6]、図 2.16から、Day0 の細胞面積が CTX 注入後の Day3,5 の面積に比べて大きい傾向にある。なお筋ジストロフィーの再生障害では筋線維の面積が小ささまざまとなり値の分散が高い傾向になる [35]。なおワイルドタイプ (wildtype:WT) と筋ジストロフィーマウス (mdx) の面積分布を比較すると、mdx マウスの面積分布が WT に比べて正規性が失われ偏った分布となることが報告されている。[10]。サルコペニアの場合は全体的に筋線維の面積が低下する [34]。CTX 注入の場合、上記のような疾病症状がでない代わりに、筋線維以外の細胞が発現されるため、筋線維か否かの判断が再生評価には必要となる。

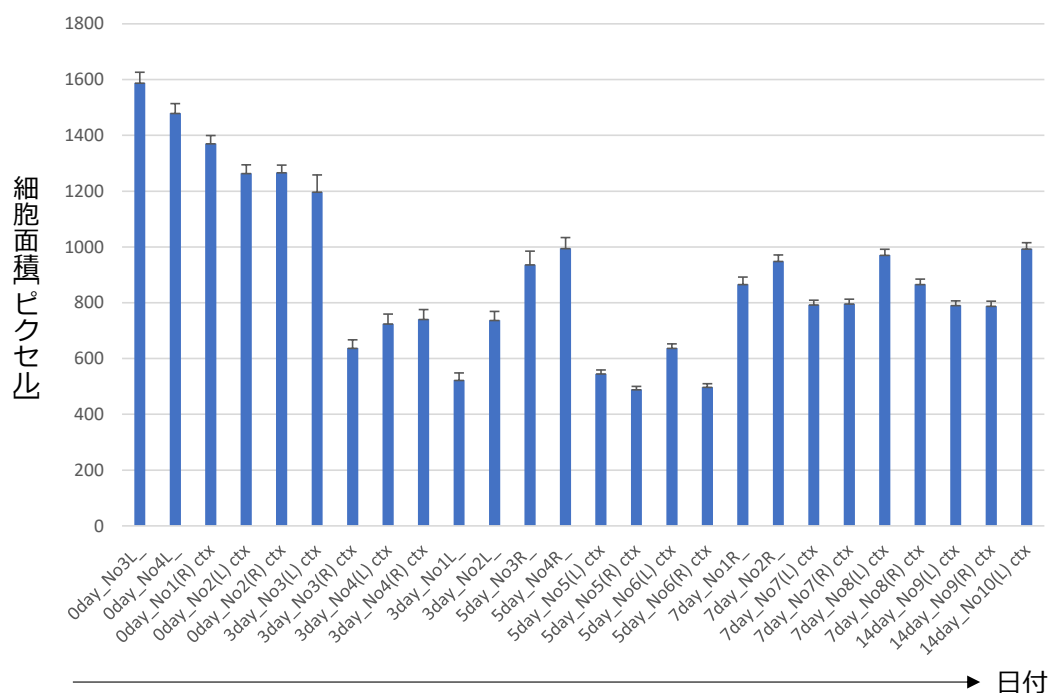


図 2.16 WSI に含まれる細胞の面積平均

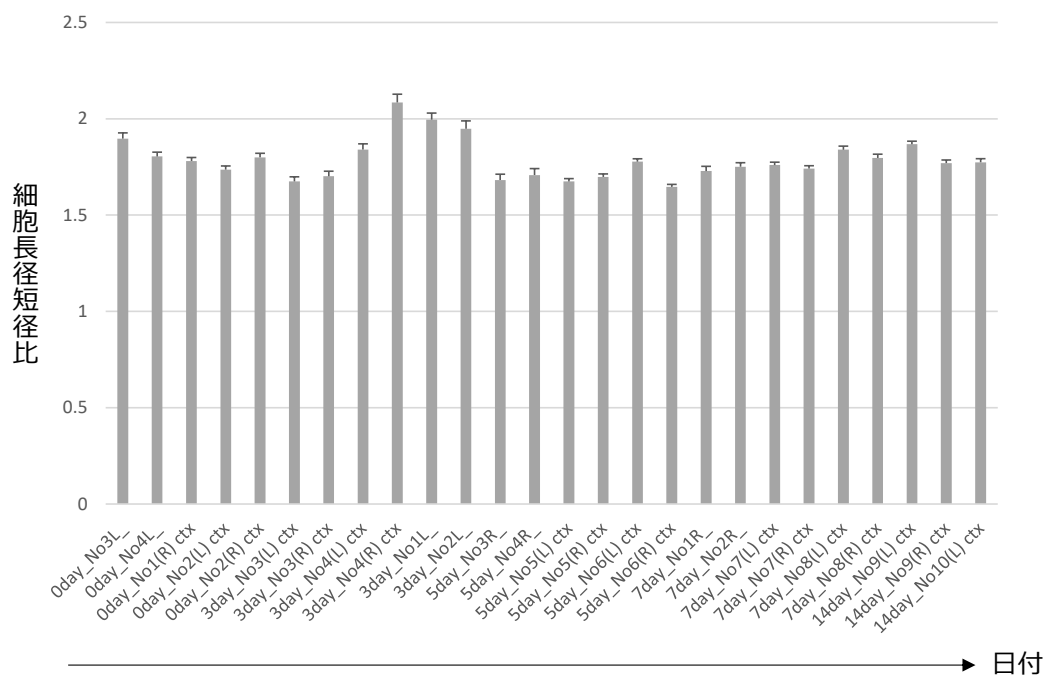


図 2.17 WSI に含まれる細胞の feret 比 (長径短径比) の平均

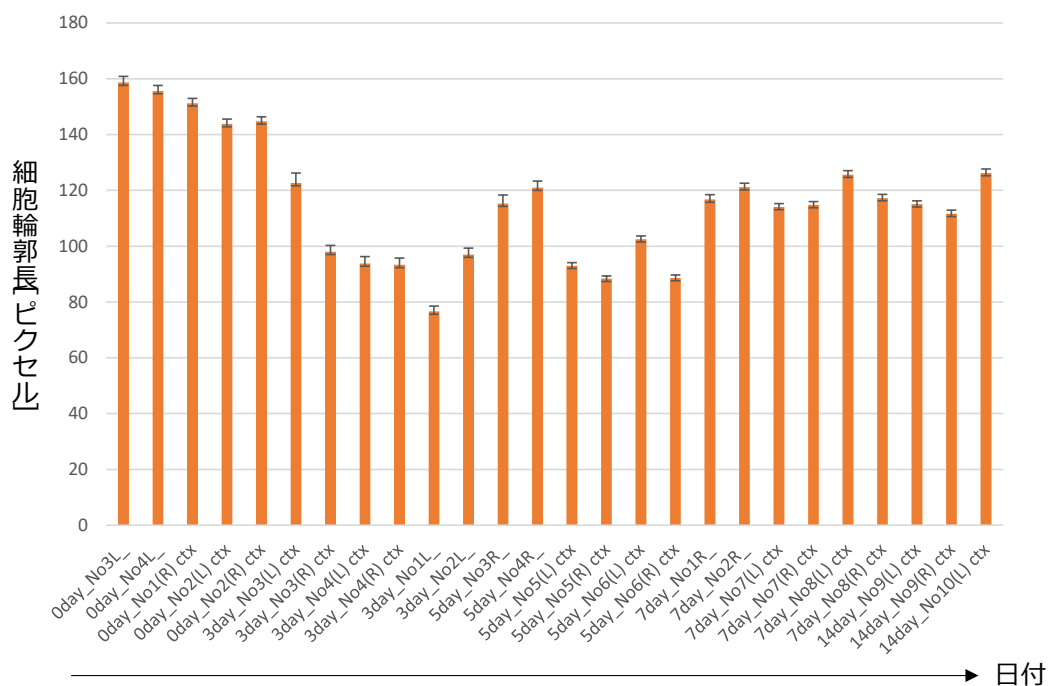


図 2.18 WSI に含まれる細胞の輪郭長平均

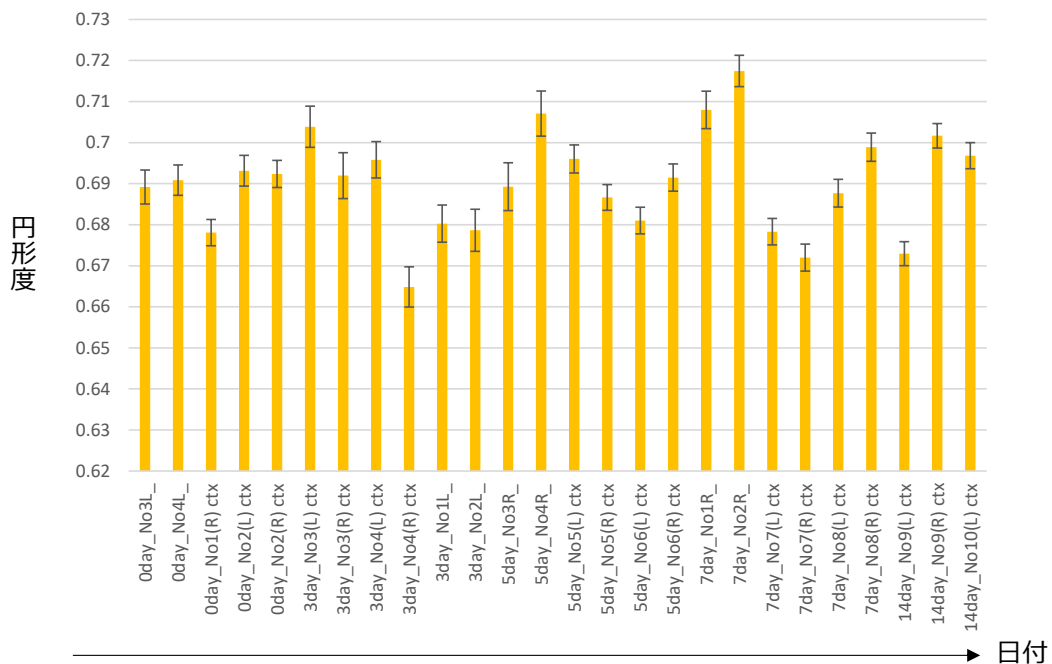


図 2.19 WSI に含まれる細胞の円形度平均

図 2.16, 2.17, 2.18, 2.19(エラーバーは 95% 信頼区間) で示されるフェレット径, 細胞輪郭長, 円形度は細胞の形態を表す明示的な特徴量であり [57], Myosoft [10] が提供するデフォルト値では, 円形度 0.3 以下, Feret 比 4.0 以上である場合筋線維ではないという閾値としている. このように明示的な特徴量を用いた線形カーネルのサポートベクター分類 [58] によって, 各細胞を安定または再生完了 (赤), 損傷 (青), 再生前期 (黄色), 再生後期 (オレンジ) にクラス分けする予備実験を行った結果を図 2.20に示す.

専門家が与えた真値とは大きく離れた推論結果となり, 従来の明示的な特徴量では回復段階の分類が難しい結果が得られた. サポートベクター分類が失敗した原因として, 再生過程に現れる細胞は明示的な特徴量が近い値を示していることが挙げられる. 図 2.21のように壊死したゴーストファイバー, 筋分化を始めた筋幹細胞, 定常状態の筋線維はいずれも円形度, 細胞輪郭長, 面積といった特徴量が近いことが判る. ゴーストファイバーはもやのような輪郭をし, 筋幹細胞は細胞内にヒビが入っている. こういった「もや」や「ヒビ」といった人間の知識を明示的な特徴量で表現することは難しい. 再生過程筋組織の細胞のように特徴が明示的ではない場合に, 暗黙的な多次元表現が強力な武器となる.

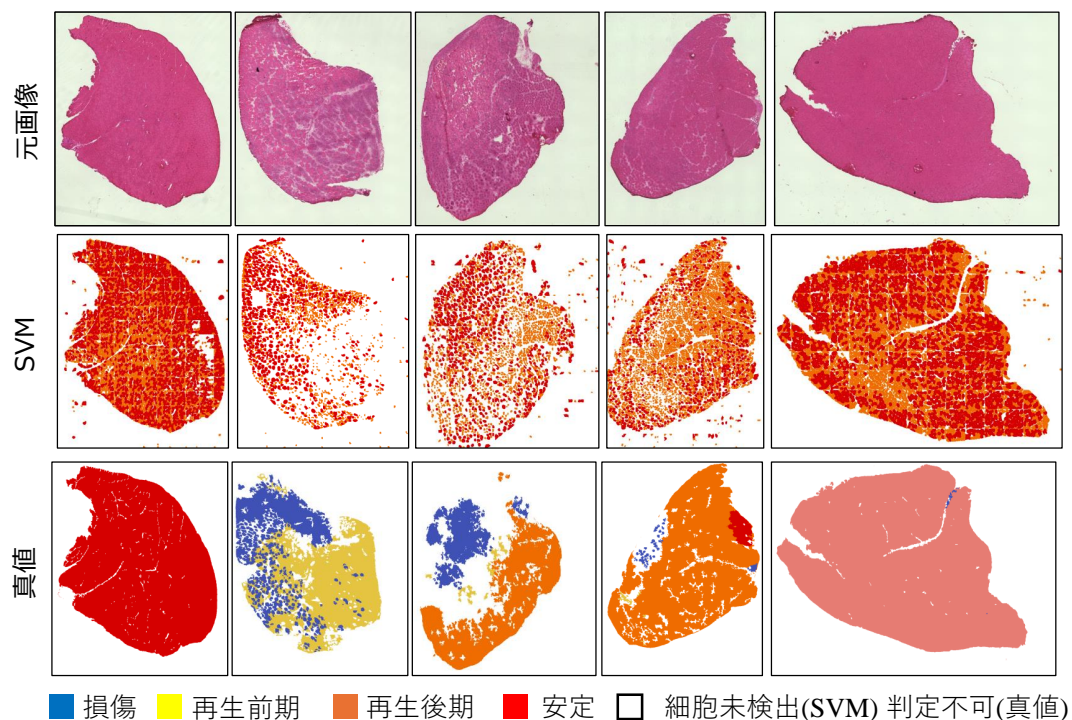


図 2.20 回復過程の細胞の明示的特徴量によるサポートベクター分類結果

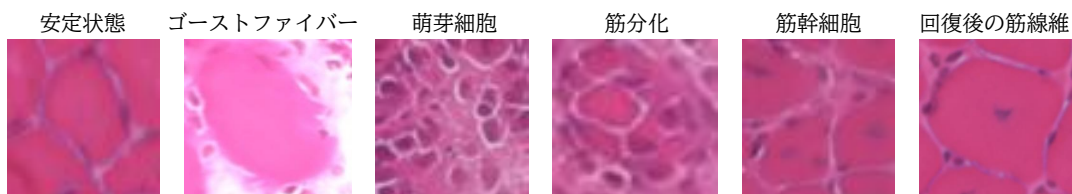


図 2.21 再生過程の細胞特徴

2.6 結言

本章では筋組織 WSI 解析において、HE 染色および蛍光染色の観察目的の違いに言及しながら、再生過程筋組織の WSI データ作成方法について述べた。蛍光染色は観察対象となる細胞ごとに適切な抗体が異なるが、筋細胞の回復には多種多様な細胞が関わっており、1つの抗体蛍光染色画像だけでは、各細胞の回復段階を評価できない。一方 HE 染色は一つの染色だけで多様な細胞を可視化するため回復段階の評価をすることができる。各細胞領域の回復段階を4段階に分けるクラス分類課題と見なし、のセグメンテーションから明示的な特徴量を計算し、従来法で再生段階を分類できるか試した。結果として再生段

階の筋細胞は明示的な特徴量だけでは表現できないことが判明し、暗黙的特徴量を用いた機械学習手法の必要性が示唆された。機械学習手法では各細胞領域に再生段階の教師を作成コストを回避するために、WSI データ作成時の個体ごとに持つ毒素注入からの日付を利用した弱教師あり学習に臨む。

第 3 章 筋組織再生評価に向けた暗黙的特徴量によるクラス比率学習

3.1 緒言

2.2.2節にある通り、損傷した筋組織には回復機能が備わっていることが知られているが、再生は筋組織全体で均一に進行するわけではない。再生レベルは組織の領域ごとに異なっており [59, 60]、図 3.1 のように解析する場所によっては誤った解釈を導く可能性がある。局所的な筋線維に注目するのではなく筋組織全体をくまなく解析することが求められるが、広い視野にわたった観察は労力が大きく、客観的かつ定量的なコンピュータによる画像解析の自動化を狙う研究が進められてきた [6–13, 46–50]。

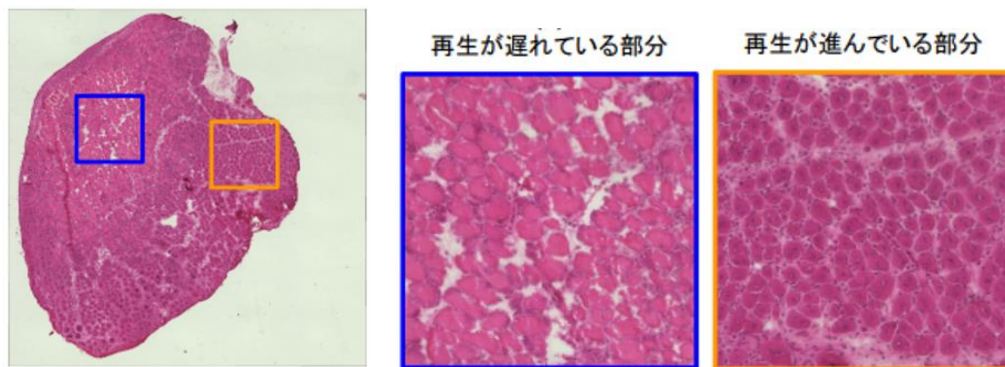


図 3.1 筋組織再生の局所性

予備実験の図 2.20 が示すように回復期の細胞分類にサポートベクターマシンを使用し Myosoft [10] で手動選択した特徴がリカバリ中の HE 染色画像には適用できないことを確認した。特に回復期間の 3 日目から 8 日目にかけてのミオシン重鎖のように明示的な特徴が存在しない場合手動での明示的な特徴選択は適用できないことが判った。そこで機械学習における暗黙的な特徴を用いることで [53, 61–63]、手動で選択された特徴よりも多次元で強力な表現力を持ち、幅広いドメインに適応することを可能とする。

本章では、筋組織 WSI の回復過程を HE 染色画像から外観検査的に行うことのできるソフトウェア MyoRegenTrack を開発する [14]。これは様々な機械学習、画像処理技術を統合したソフトウェアである。当該ソフトウェアは HE 染色された筋組織の顕微鏡画像である Whole Slide Image (WSI) を入力として筋組織全体の再生状態を評価することに

成功した。また CTX と Glycerol を用いた筋組織再生の評価実験では、MyoRegenTrack の導いた結論は先行研究の Glycerol に再生障害を見られる知見 [9, 64–68] と一致した。

3.1.1 機械学習手法による推論法

一般的にクラス分類を行うとき推論モデルは特徴量抽出器 (画像の [ピクセル数*channel 数] 次元から数百～数次元への圧縮:BackBone) とクラス推定器 (特徴量からクラスを予測:Head) の段階に分かれている [69, 70]。近年これらの段階を特別意識せずに End-to-End と呼ばれる学習法によって、入力データと教師のみを与えてモデル内部を完全にブラックボックス化して学習する手法も有名であり [71]、大規模なデータセットと計算器によって学習されており [5, 53]、事実上 accuracy needs money な状態になっている。End-to-end は中間層の次元圧縮最適化に着目しないで済むというメリットがある一方、その汎化性能を得るには大規模データセットが必要とされる。ひるがえって限られたデータセットの場合、特徴抽出の挙動にも目を配る必要があり、End-to-end では精度が出づらいうという一面もある。さてこれまでの筋線維の状態分類を試みる画像解析では、HE 染色向けではないものの Myosoft [10]、MuscleJ [6] や Open-CSAM [13] が免疫染色された筋組織のために開発されており、セグメンテーションした筋線維から面積と円形度などの特徴量を取得する。Open-CSAM [13] は CTX によって損傷させてからの経過日数が Day8 から Day28 を対象とした免疫染色画像の CSA 計算を行っているが、ユーザは経過日数やマウスの年齢・病気状態に応じて大きさの閾値 (Size threshold) と円形度閾値 (Circularity threshold) をチューニングしたり、偽の線維 (false fibers) や未検出線維をマニュアルで追加する必要があったりと、人手による閾値の設定や補助作業はユーザーの能力や撮影条件の影響を受けてしまう。かといって Myosoft [10] のようにフェレット比 (Feret aspect ratio) や最小フェレット径 (Minimum feret) など形状を評価できる解析項目の特徴量を増やした場合は、特徴量選択に開発者のセンスが含まれてしまうことでドメインギャップに弱くなってしまう傾向にあり、Day3 から Day8 に見られる回復途中の胚性ミオシン重鎖 (embryonic/neonatal myosin heavy chain) のように状態を示す明示的な特徴量が判らない場合、マニュアルでの特徴量選択は適用できない。実際我々も当初は線維の形状を表す特徴量 [10] を参考に選択してサポートベクター分類ができないか挑戦したがその分類性能は高いものではなかった (図 2.20)。一方機械学習的手法を用いて網羅的に数百次元の特徴量を取得しクラス推定を行う場合、次元圧縮された空間上において次元数に応じた情報表現力を獲得することができるため幅広いドメインに適応することが

できる．本研究では特徴量抽出器 (Backbone) には教師ラベルのない画像からの特徴量抽出として既に drug target と gene family classification の課題において CellProfiler [63] を上回る精度を発揮した報告のある self-distillation with no labels(DINO) [53] を用いることとした．得られた特徴量を Full Connection と ReLU [72], Softmax 関数から成るニューラルネットワークに通し (Head), クラス推定器の学習に試みる．

3.1.2 日付を弱教師としたクラス比率学習

クラス分類問題を解くときに教師ラベルを作成してモデル学習する教師ありの手法 [73, 74] は強力だが，今回のように再生 Process 中の筋線維画像ではアノテーション自体のコストが大きく，さらにそのアノテーションが正確であるという保証をつけることは困難である．そこで我々は唯一得られる事前情報である時系列情報 (CTX を注入してからの経過日数) を弱教師として活用し分類に試みた．限られた GPU メモリ上で扱うために筋組織画像はセグメンテーションされた線維ごとに切り取った．切り取り画像 (clip image) は線維の数だけ存在し，それぞれの状態について 2.7のように損傷，再生前期，再生後期，再生完了または損傷前を安定期の 4 つの段階といった細胞クラスを定義し，個々の細胞画像のクラスラベル作成はできなくても，筋組織画像の概観観察から whole slide image (WSI) における各クラス比率を出すことは難しくない．たとえば Day0 は CTX 注入前であるからすべての線維が安定的であり，Day3 はゴーストファイバーや萌芽細胞が多い．Day が経つにつれて筋幹細胞の数は増えていき，反対に萌芽細胞の数は減っていくであろう．図 3.2に細胞形態の日付に応じた分布の変化の一例を示す．各線は日付に対応し，色は細胞クラスに対応する．横軸は一つ一つの細胞の回復度合い (Recovery Score) を示しておりその確率密度分布となっている．Day3 は回復度合いの小さい細胞群が多い分布であり，日を追うごとにその分布が右に移動していることを示している．この分布の移動を事前情報として持ったうえで，WSI に非常にラフなアノテーションを行い，日付情報と WSI のクラス比率を紐づけることを行う．

こうしたクラス比率から学習する手段として最も一般的な方法は疑似ラベルである．日付ラベルに関連付けられた筋線維の状態のクラス比率があるため，確率的なクラスを切り取り画像に付与することで既存の教師あり学習に落とし込む．疑似ラベル [75] はバニラな手法であるがノイズを生み出してしまう可能性が残り必ずしも良い精度を期待できない．そこで個々のインスタンスにラベルはなくとも，複数のインスタンスで構成されたグループ (本論文 [76–78] では先行研究に習いバグと呼称する) のクラス割合が事前に与

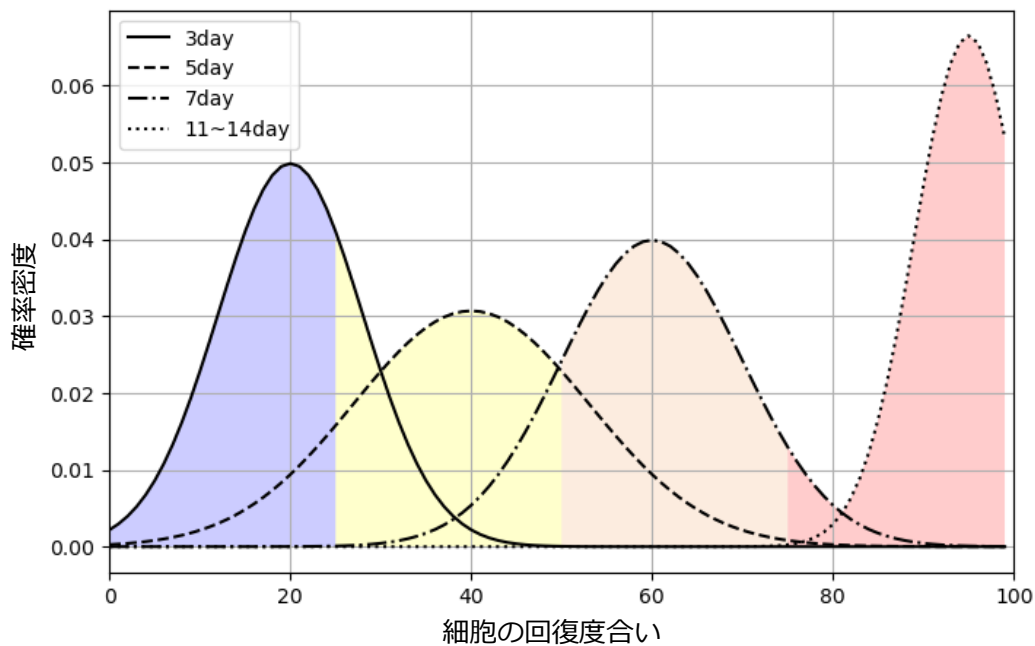


図 3.2 日付ごとの細胞状態分布の正規予測

えられているという課題設定にすることで学習を行う weekly supervised な手法であるクラス比率学習 (Learning from label proportions: LLP) [76] に我々は注目した。LLP では図 3.3のように各インスタンスにクラスラベルはないが、バッグと呼ばれるインスタンス群にはクラス比率が与えられている問題設定における学習方法である。

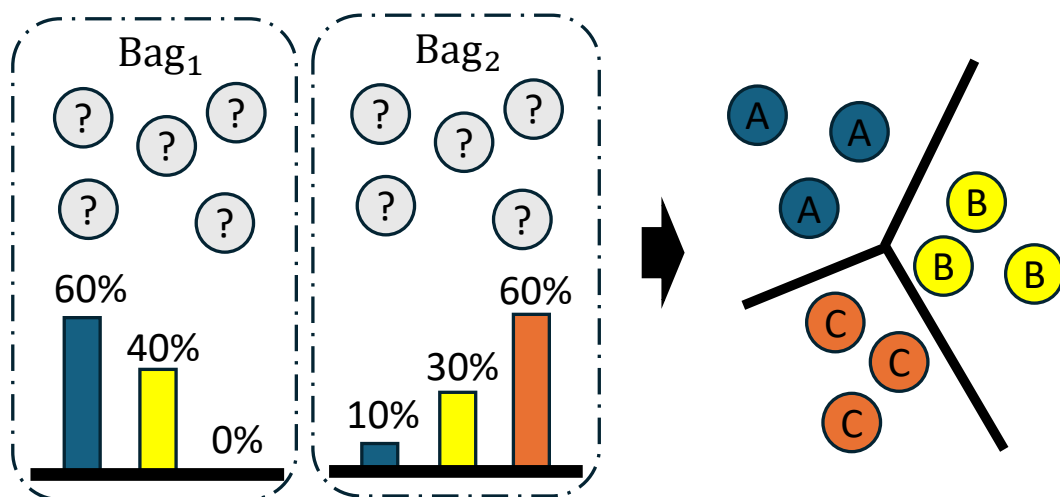


図 3.3 クラス比率学習

すなわち機械学習の損失を計算するときにある日付の画像から切り取って得た複数のイ

インスタンス群の予測結果を用意し、日付ラベルに紐づけられたクラス割合と比較することでモデルの勾配を更新する。LLP はプライバシーの観点から個々のインスタンスラベルを生成することが困難な医療データベースを用いた学習においても強い力を発揮し、妊娠率を向上させるための胚選択 (embryo selection) において胚着床予測データ (data from embryo implantation prediction) から実際に着床した割合を使って個々の事例での推定に成功している [77]。医療画像を入力とする場合でも LLP は使われており特に大規模画像の数千万ピクセルから数億ピクセルの WSI を扱うときにも力を発揮する。ほとんどの場合において WSI をピクセルレベルで様々な組織の状態にアノテーションすることは専門家にとっては負担が大きく実施されるケースは少ない。腫瘍の懐死 2 値判定に LLP を使用した報告 [78] では、Fuzzy Proportions と WSI をクリップした画像群セットを用いて LLP を行いその精度を確認している。本研究は組織画像の日付ラベルと線維状態クラスの割合を関連付けさせた LLP の学習手法を取り入れることで線維のクラス分類を可能とする推論器の構築を狙う。

3.2 提案手法

3.3.2節、3.3.3節で用いた学習データ (train data) は表 2.2の「学習用データ」、テストデータ (test data) は「アノテーション済み検証用データ」である。3.3.4節で用いたデータは表 2.2の「比較実験用データ」である。

学習時には png 形式で行う。これは tiff 形式であるとデータサイズが大きくなりすぎてしまい学習および推論時に GPU 負荷が大きくなることが理由である。ただし負荷を小さくするために jpg 形式にすることには注意が必要で、目視では png 形式と jpg 形式には大きな差はないが、openCV などプログラム上で扱う際には同じ画像であってもデータの中身が異なる。本論文では初期データは tiff 形式、モデル学習を行う際は png 形式、それ以外のときに jpg 形式を採用した。

3.2.1 提案手法の概観

図 3.4に HE 染色の WSI を解析する提案ソフトウェア MyoRegenTrack の概観を示す。WSI はセグメンテーションモデルの Cellpose と組織領域を検出する Edge ソフト (Python) に並行に入力され、2 つの出力画像の共通領域のみをセグメンテーション結果として扱う。これはセグメンテーションモデルが染色されていない組織の背景に細胞がいると判定する誤検出に対応するためである。得られたセグメンテーションデータから各

細胞画像の特徴量を DINO [53] によって取得し，クラス分類器をかけることで細胞の回復程度を 4 段階に分類する．その後各クラスに応じて細胞部への色塗りを行うことで MyoRegenTrack の出力結果とする．また本ソフトウェアの使用法については動画を Github 上 (<https://github.com/RyuAmakaze/MyoRegenTrack>) に公開している．

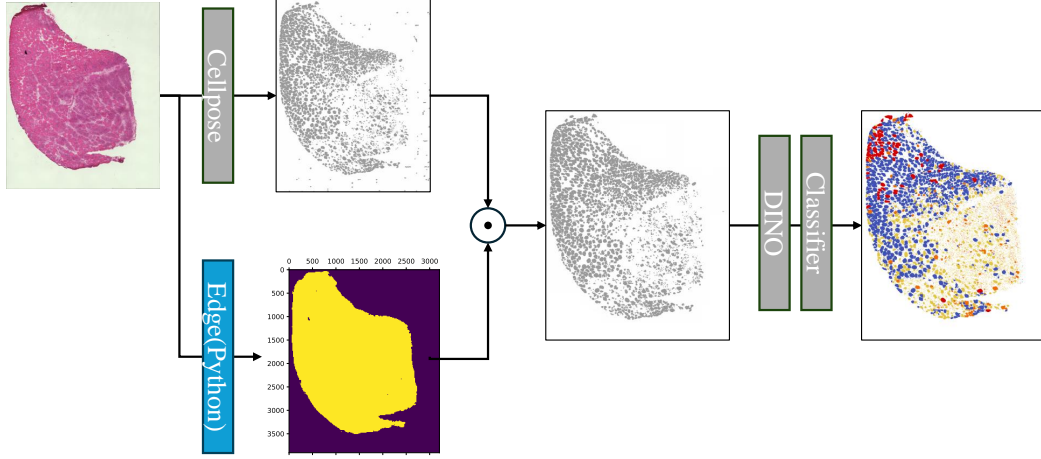


図 3.4 提案ソフトウェアの概観

図 3.5は 1 枚の Whole Slide Image から各細胞のクラス分類を行うまでの推論パイプラインである．1250 万画素の画像を格子状に $L \times L$ のピクセルに切り分けてから Cellpose [52] による細胞セグメンテーションを行う．ここの分割サイズ L は分類精度に重要ではない．あくまで数千万画素をそのまま Cellpose に適用することができないため計算がしやすいように $L \times L$ のピクセルサイズへカットするだけである．セグメンテーションを行うと各細胞の輪郭と中心座標が手に入るため，細胞を中心とした $l \times l$ ピクセル正方面像を切り取る．この l ピクセル正方面像を当論文では線維画像あるいは細胞画像とし，Vision Transformer ベースの特徴抽出器である DINO [53] を経て D 次元の特徴量インスタンス $\mathbf{x} \in \mathbb{R}^D$ を手に入れる．また D は圧縮された次元数である．この特徴量抽出器は通例バックボーン (Backbone) と呼称される．続いて特徴量インスタンス $\mathbf{x} \in \mathbb{R}^D$ をクラス分類器に通すことで $Class : \mathbb{C} = \{1, \dots, k, \dots, K\} (ex : red, blue, yellow, orange)$ を得る．クラス分類器は Backbone に対してヘッド (Head) と呼ばれる．

特徴量インスタンス $\mathbf{x} \in \mathbb{R}^D$ から $Class : \mathbb{C} = \{1, \dots, k, \dots, K\}$ (例 : *red, blue, yellow, orange*) を予測する推論器のアーキテクチャについて述べる．図 3.9に示す通り 3 層のニューラルネットワークを用いて学習を行った．入力 $\mathbf{x} \in \mathbb{R}^D$ であり，ニューラルネットワークの出力はクラス数 K についてのベクトル \mathbb{R}_+^K であるため，図 3.9を関数 \mathcal{F} と見なした場合，

$$\mathcal{F} : \mathbb{R}^D \rightarrow \mathbb{R}_+^K \quad \left(\sum_{k=1}^K \mathcal{F}(\mathbf{x})_k = 1 \right) \quad (3.1)$$

と書き下すことができる．

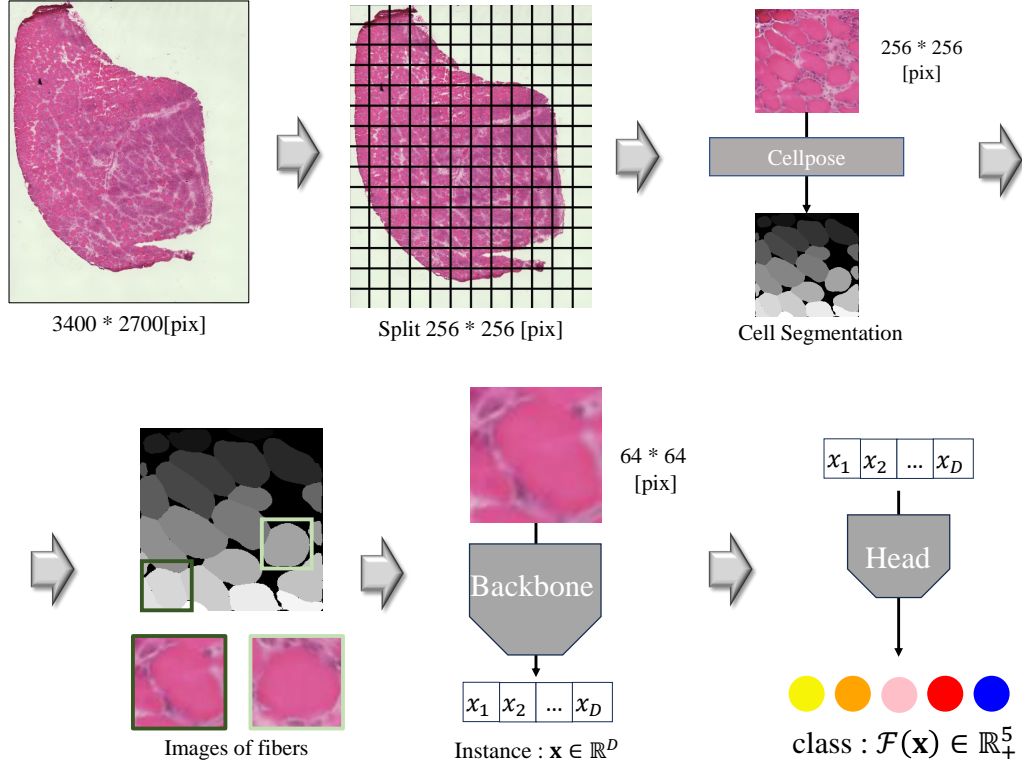


図 3.5 クラス推論のパイプライン

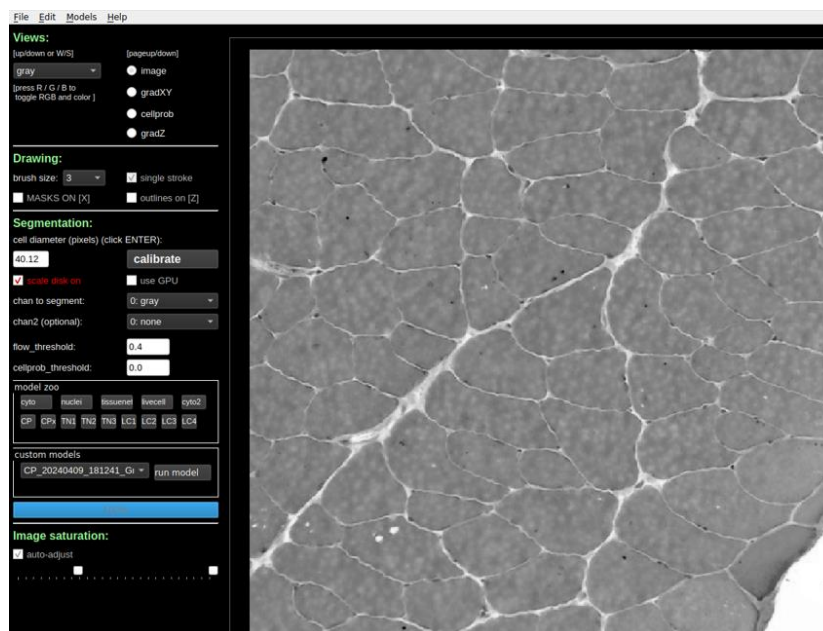
ただし $\mathcal{F}(\mathbf{x})_k$ はクラス分類器 (head) が出力するクラス k の確信度である．

WSI から $L \times L$ [pixel] の画像に分けた後に Cellpose [52] で推論された細胞セグメンテーションを得る (表 3.5 参照)．cell を中心とした $l \times l$ [pixel] の画像にしてから特徴量抽出器によりインスタンス特徴量 (\mathbf{x}) $\in \mathbb{R}^D$ を得て，分類モデルによって $\mathcal{F}(\mathbf{x}) \in [0, 1]^K$ ($\|\mathcal{F}(\mathbf{x})\|_1 = 1$)，ただし $\mathcal{F}(\mathbf{x})$ はクラス分類モデルが出力するクラス確信度である．ただし $\|\cdot\|_1$ は L1 ノルムを意味する．

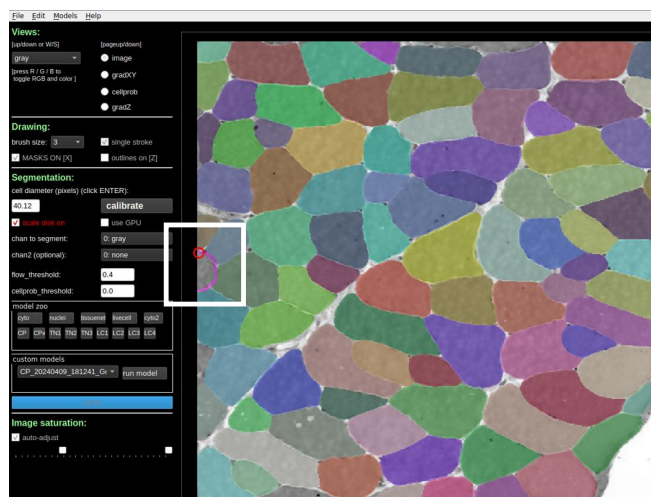
3.2.2 Cellpose のファインチューニング

本節では領域検出 (セグメンテーション) のための Cellpose [52] モデルのファインチューニングの方法について述べる．HE 染色画像の細胞セグメンテーションの先行研究 [8] では，回復過程における筋組織画像に現れる安定状態な筋線維以外の細胞について

脆弱であるため、これに対応するために Cellpose モデルをファインチューニングする必要がある。



Cellpose GUI



既存モデルでの推論例



アノテーション例

図 3.6 Cellpose の GUI を用いたアノテーション

表図 2.2の学習用データデータの Day0 から Day7 の 1250 万画素の画像を 256×256 [pixels] の画像に切ってから各日付 4 枚ずつ計 20 枚 256 [pixel] 正方画像の細胞輪郭アノテーションをマニュアルで行った。なお図 3.6のように細胞輪郭アノテーションコストの削減のために既存モデルで推論を行い、セグメンテーションに誤りがある部

分に修正を行う形で作業は進められた。アノテーションソフトについては Cellpose が提供する GUIを用いて行った。

3.2.3 日付ごとのクラス比率計算

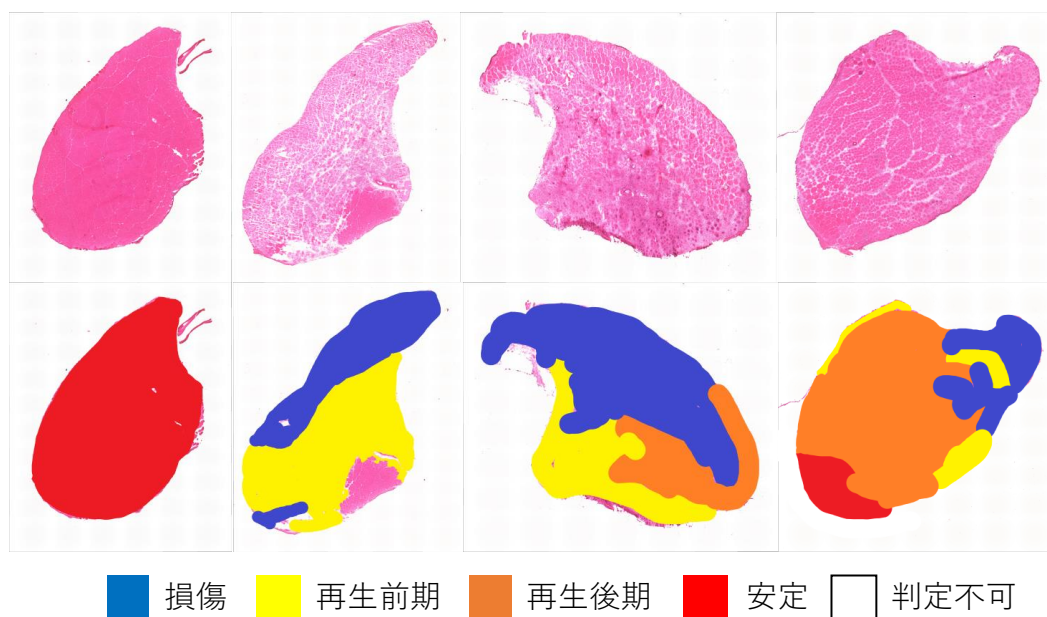


図 3.7 ラフアノテーション

壊死した筋線維から発生する damaged-myofiber-derived factors (DMDFs) によってサテライト細胞の発生が促進され、サテライト細胞が筋分化の過程を経て線維は回復する [45]. この過程は図 2.7のように損傷, 再生前期, 再生後期, 再生完了または損傷前を安定期の 4 つの段階に分けることができる. これを分類課題にするためにクラス C : *red*(安定期), *blue*(損傷), *yellow*(再生前期), *orange*(再生後期) を定義する.

表 2.2にある学習用画像すべてを図 3.7のように手作業でラフなアノテーションを行った. 粗さの指標として, Windows にプリインストールされているペイントアプリケーション (バージョン 11.2404.45.0) で, 少なくとも 100 ピクセルの円形ペンでアノテーションを行っている. これらのアノテーションはクラス比率を決定するためのものであり, 分類モデルを学習するための各繊維の正確な教師としては機能しない. なお Day14 はすべてが回復しきっていると見なして Day0 の画像群に含める. よって学習時の日付ラベルは $day \in \{day0, day3, day5, day7\}$ となる. 得られた色塗り画像から各色のピクセル数をカウントし, 日付ごとに集計および正規化を行い, 日付と各クラス比率の対応関係

p_{day} を導いた.

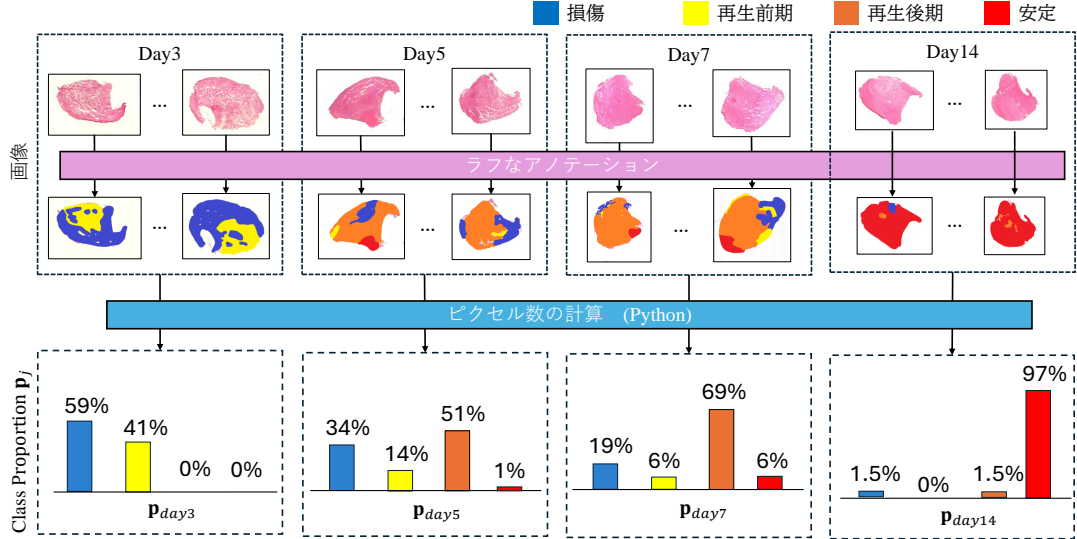


図 3.8 クラス比率の計算方法

3.2.4 推論器 (head) のアーキテクチャ設計

D 次元の特徴量インスタンスはニューラルネット 1 層目に全結合 (Full connection [18]) される (FC1). FC1 層から出力された値は活性化関数 ReLU [72] を通じて次の FC2 層に結合される. このとき特徴量は D' 次元に圧縮される. ニューラルネットワークの全結合 (Full connection) 層同士の間はすべて活性化関数 Rectified Linear Unit (ReLU) [72] $f: \mathbb{R}^D \rightarrow \mathbb{R}^D$ で繋げている. ReLU とは下記に示すように 0 以下の値はすべて 0 とする関数である.

$$\begin{cases} f(h) = h & h > 0; \\ f(h) = 0 & h \leq 0 \end{cases} \quad (3.2)$$

3 つ目の Full connection 層 (FC3) から出力されたベクトルは値の総和を 1 とするために SoftMax 関数で正規化する.

$$\text{SoftMax}(\mathbf{h})_i = \frac{\exp(h_i)}{\sum_{j=1}^K \exp(h_j)} \quad (3.3)$$

SoftMax 関数で得られた出力を各クラスの推論確信度 (confidence score) とした. クラスを決定する際は確信度の最も高いものを選ぶ.

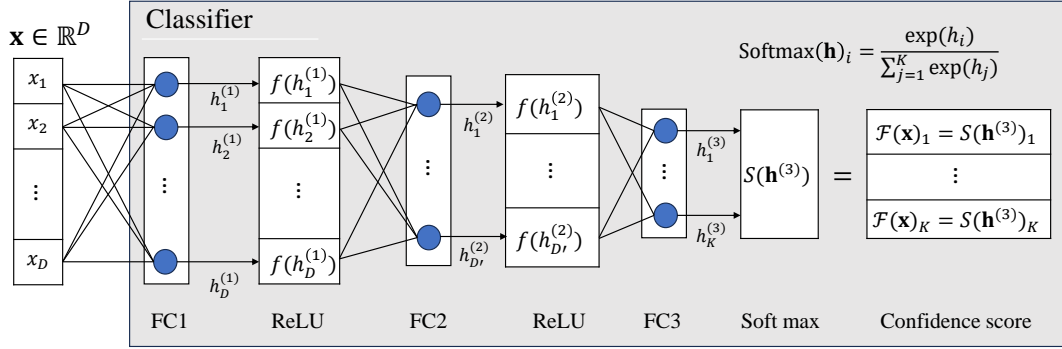


図 3.9 クラス分類器のアーキテクチャ

3.2.5 クラス比率学習

Learning from Label Proportion (LLP) は事前情報から推定された日付ごとの各クラスの Proportion を利用した弱教師あり学習である。Proportion loss \mathcal{L}_{prop} は予測した $\hat{\mathbf{p}}_{day}$ と図 3.10 で得た \mathbf{p}_{day} の間で $D_{KL} : \mathbb{R}^K \cdot \mathbb{R}^K \rightarrow \mathbb{R}_+$ を計算する。

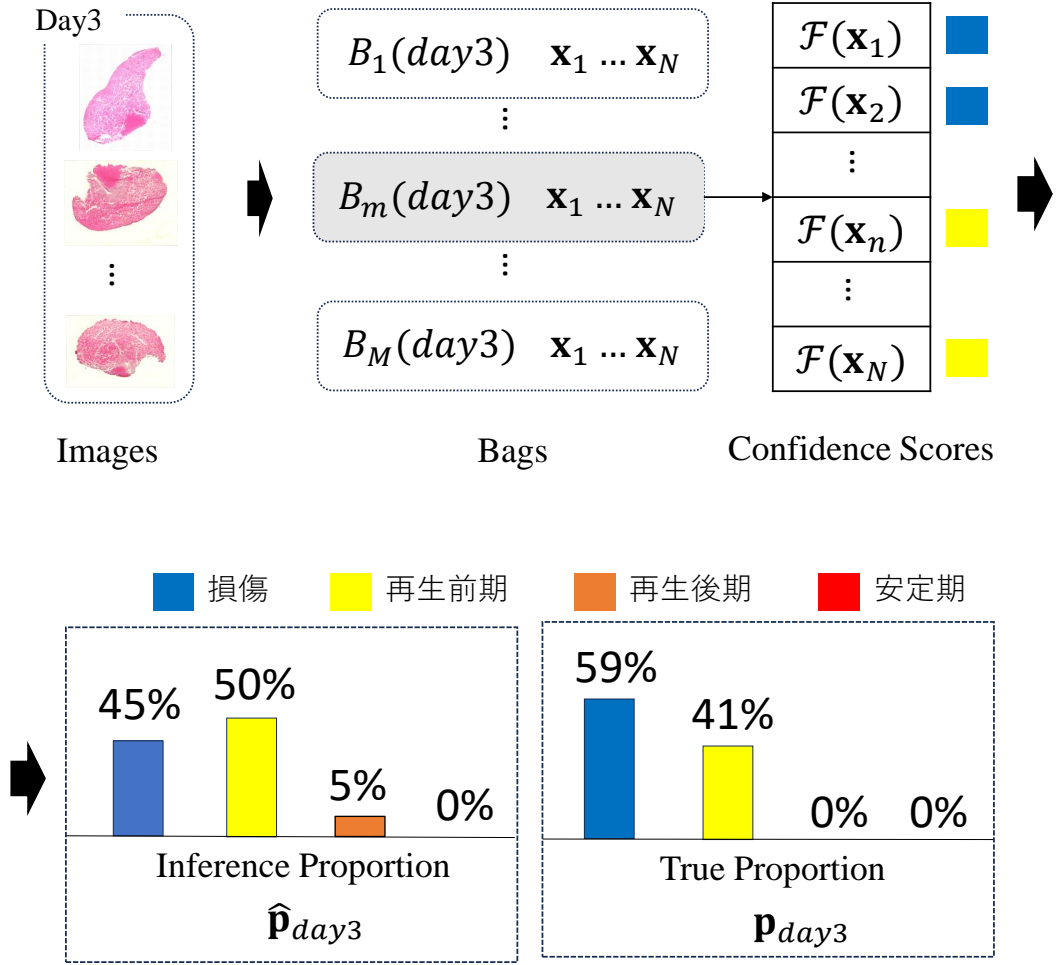
$$\mathcal{L}_{prop}(\theta) = D_{KL}(\hat{\mathbf{p}}_{day} \parallel \mathbf{p}_{day}) = \sum_{k=1}^K \hat{p}_k \log \left(\frac{\hat{p}_k}{p_k} \right) \quad (3.4)$$

まず各線維画像は Whole Slide Image (4K ピクセルの画像) が持つ日付ラベル $d \in \{day0, day3, day5, day7\}$ を持つ。線維画像は Backbone に通されることでインスタンス $\mathbf{x} \in \mathbb{R}^D$ (D : 圧縮次元数) となる。同じ日付ラベルを持つ N 個のインスタンス $\mathbf{x} \in \mathbb{R}^D$ を集めることで Bag $\mathbb{B} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ を作成する。Bag は日付ラベルに応じて真のクラス比率 \mathbf{p}_d を持つ。

損失は Bag ごとに計算されるためインスタンスから予測クラス比率 $\hat{\mathbf{p}}$ を求める。Bag $\mathbb{B} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ に含まれるインスタンスをクラス分類器 (head) $\mathcal{F} : \mathbb{R}^D \rightarrow \mathbb{R}_+^K$ に通すことで各インスタンスのクラス確信度ベクトル $\{\mathcal{F}(\mathbf{x}_1), \mathcal{F}(\mathbf{x}_2), \dots, \mathcal{F}(\mathbf{x}_N)\}$ を得る。 $\hat{\mathbf{p}}$ はインスタンスのクラス予測確信度ベクトル $\mathcal{F}(x)$ を bag に含まれるインスタンスで平均を取ったものとなる。すなわち Bag に含まれるインスタンス数 N (bag size) に対して

$$\hat{\mathbf{p}} = \frac{1}{N} \sum_{n=1}^N \mathcal{F}(\mathbf{x}_n) \quad (3.5)$$

となる。Class : $\mathbb{C} = \{1, \dots, k, \dots, K\}$ ($ex : red, blue, yellow, orange$) それぞれの値は $\hat{\mathbf{p}}_{day} = [\hat{p}_1, \dots, \hat{p}_k, \dots, \hat{p}_K]$ において,



$$\mathcal{L}_{prop} = D_{KL}(\hat{\mathbf{p}}_{day3} || \mathbf{p}_{day3})$$

図 3.10 クラス比率学習

$$\hat{p}_k = \frac{1}{N} \sum_{\mathbf{x} \in \mathbb{B}} \mathcal{F}(\mathbf{x})_k \quad (3.6)$$

ただし

$$\sum_{k=1}^K \hat{p}_k = 1 \quad (3.7)$$

3.2.6 リカバリースコアの導入

クラス分類による Whole Slide Image への色塗り表示は定性的な結果である一方定量性に欠ける．よって推論したクラスの分布に応じて回復度を定量化する手法リカバリース

コアを導入する．リカバリースコアとは推論クラス比率 $\hat{\mathbf{p}}$ と各クラスの回復への寄与を示す重み $\boldsymbol{\omega}$ によって以下のように計算される．

$$\text{RecoveryScore} = \hat{\mathbf{p}}_{day} \cdot \boldsymbol{\omega} = \sum_{k=1}^K \hat{p}_k \cdot \omega_k \quad (3.8)$$

ただし

$$\sum_{k=1}^K \hat{p}_k = 1, \quad 0 < \omega_k < 1 \quad (3.9)$$

重み $\boldsymbol{\omega}$ の計算方法について述べる．日付に対する組織の RecoveryScore を sigmoid 関数

$$\sigma(x) = \frac{1}{1 + e^{-a(x-d)}} \quad (3.10)$$

ただし gain $a > 0$, 回復の変曲点を示す日付 $d \in \mathbb{N}$, e はネイピア数でモデル化する．gain a はシグモイド関数の勾配の大きさを示し回復度の速度を反映しており，変曲点 $d[day]$ は回復が大きく進展する日付に対応している．図 3.8で個別の WSI ごとに得た真のクラス比率 $\mathbf{p}_{day} \in \mathbb{R}^{1 \times K}$, 回復スコア $\sigma(x) \in \mathbb{R}$, 重み $\boldsymbol{\omega} \in \mathbb{R}^{K \times 1}$ を用いて目的関数である最少二乗和 $(\sigma(x) - \mathbf{p}\boldsymbol{\omega})^2$ を最小化する．最小二乗Numpy v1.20.3の `linalg.lstsq($\mathbf{p}_{day}, \sigma(day)$)` から $\boldsymbol{\omega}$ を導出した．

3.3 実験結果と議論

3.3.1 実験手法

3.3.2節「セグメンテーションモデルのファインチューニング」では cardiotoxin(CTX) を注入したマウスの筋組織の回復経過画像 Day0, 3,5,7,11,14 を用意し, Day3 や 5 に見られる Myoblast や Myotube を正しくセグメンテーションできるようにファインチューニングによるモデルの改良を行った. ファインチューニング時の Cellpose のバージョンは 2.0.3, 学習開始時点のモデル (initial model) は cyto2, 画像チャンネル (channel) は [0,0] つまり細胞は白黒 (gray), 核はなし (none) を選択し他のパラメータについてはデフォルトを使用した. Cellpose に入力する際の WSI から切り分けた正方画像サイズ $L \times L$ は, $L = 256$ である. 画像を 512 正方に切り分ける際に細胞が一つもない画像についてこの時点でファインチューニング用のデータから除去している. この結果が先行研究 [8] およびベースライン [52] の精度を超えていることを確認する.

3.3.3節「筋線維のクラス分類」では DINO による特徴量抽出及び LLP 手法で得られた Classifier モデルによるクラス推論を行い, 各クラスを色付けて専門家によるアノテーションとの比較並びに交差検証を行い提案手法の精度を確かめる. DINO に入力する際に 512 正方画像から細胞単位を切り分ける際の画像サイズ $l \times l$ は, $l = 64$ である. DINO のパラメータについてはパッチサイズ (patch size) は 8, avgpool patchtokens True, checkpoint key は student である.

3.3.4節「CTX とグリセロールの回復経過解析」では既に知られている Glycerol の回復障害現象を本提案のソフトウェアで自動解析できることを確かめ Classification による詳細な解析の有用性を示した. リカバリースコアについて, 今回のデータセットはシグモイド関数のパラメータを $a = 0.65, d = 6[\text{day}]$ と設定し, 得られた値は $\omega = [0.230(\text{青}), 1(\text{赤}), 0.756(\text{オレンジ}), 0.367(\text{黄色})]$ である.

3.3.5節「免疫染色と手動アノテーションの比較」では手動アノテーションによる回復段階の分類及び MyoRegenTrack 推論の妥当性を, 抗体を用いた免疫染色の結果を用いて示す.

3.3.2 再生過程筋組織のセグメンテーション

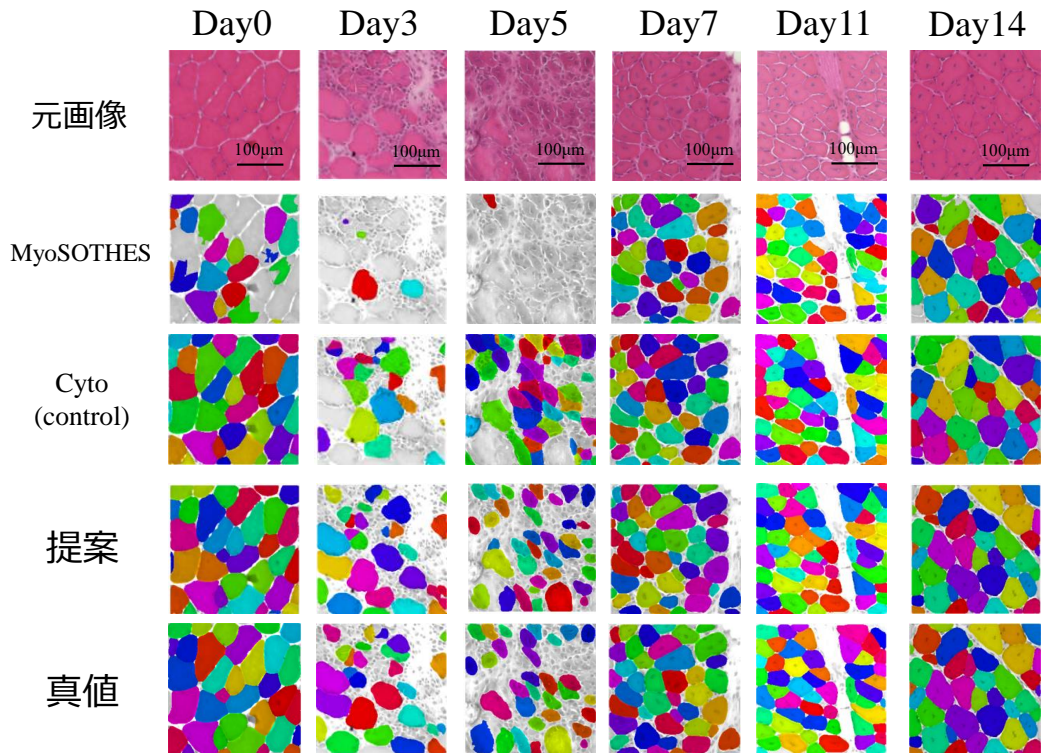


図 3.11 256 ピクセル正方画像のセグメンテーション結果例 (文献 [14] の Fig.1(a) 再掲)

Day0 は CTX を Injection する前を意味しており，以降は CTX を Injection してからの経過日数を示す．すべての画像は Cellpose に入力されるときに Whole slide image(約 4K*4K[pix]) から 256*256[pix] の正方画像へ格子状に均等にカットされ，その結果の例は図 3.11 に示される．MyoSOTHEs [8]，cyto(Cellpose [52] が提供する事前学習モデル)，提案 (ファインチューニングを行ったモデル) の 3 つのモデルと手作業によるセグメンテーション真値を比べた結果を CTX 注入からの経過日数別に示す．また切り分けた各 256 正方画像を統合し直して元の Whole slide image の通りに並べなおした結果が図 3.13 である．

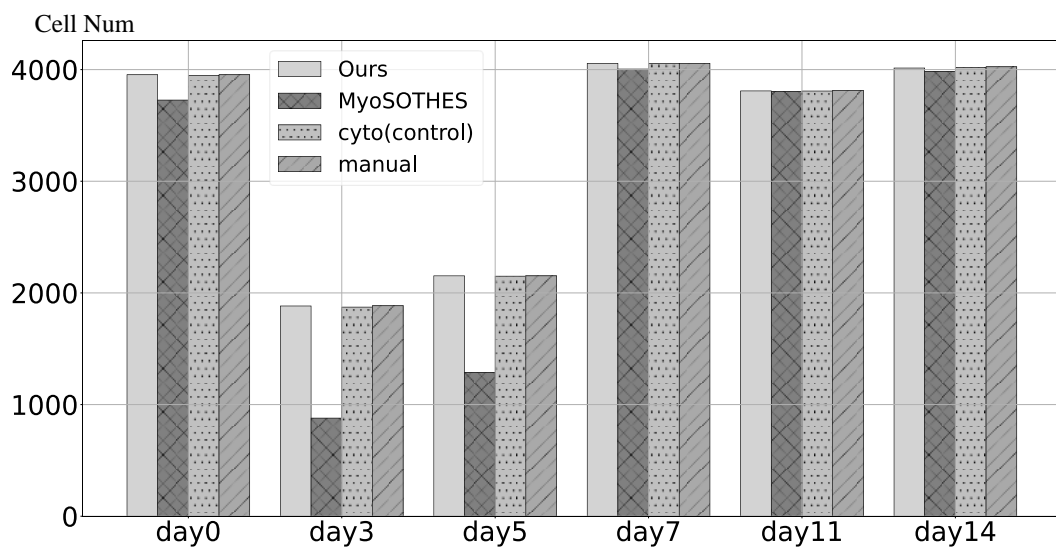


図 3.12 検出した細胞数 (文献 [14] の Fig.1(c) 再掲)

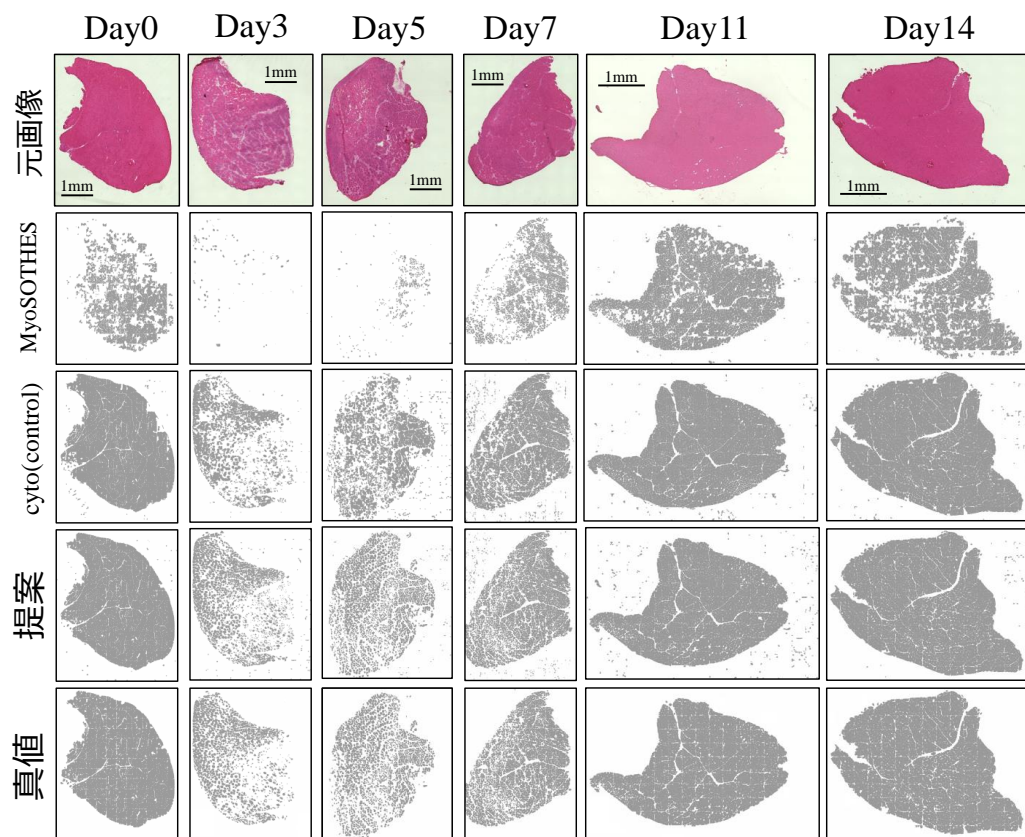


図 3.13 WSI に統合後のセグメンテーション結果 (文献 [14] の Fig.1(b) 再掲)

図 3.11 3.13の定性結果から、ファインチューニングがセグメンテーションタスク効果的

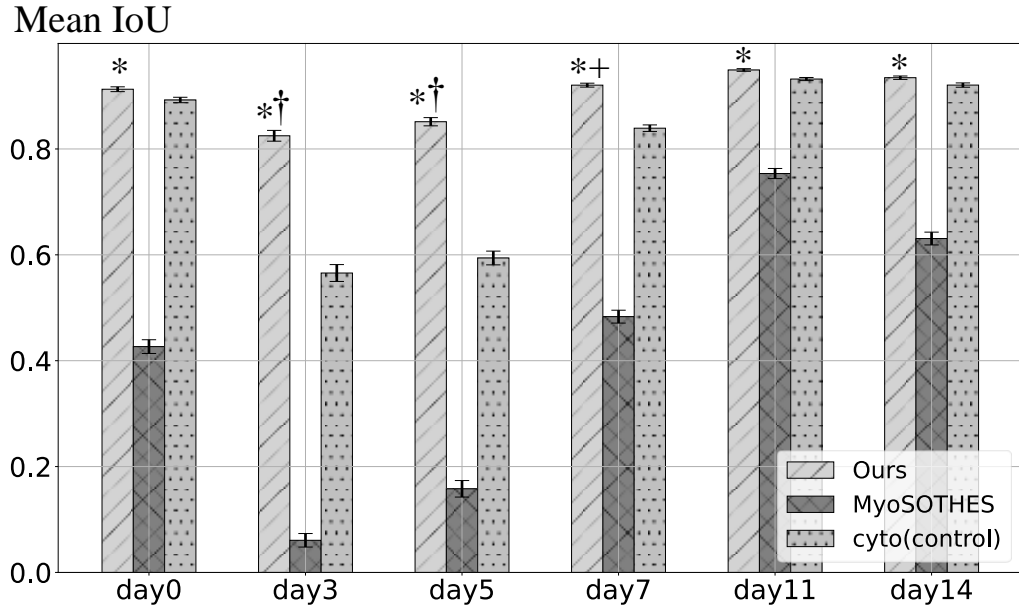


図 3.14 セグメンテーション平均 IoU(文献 [14] の Fig.1(d) 再掲)

であることが判る．特に Day3 および Day5 では先行研究では捉え切れなかった細胞を提案法では捉えられている．

細胞セグメンテーションによって検出した細胞数を図 3.12, Manual を真値として推論したモデルの性能を測る指標として真値と予測の Overlap の平均値である Mean IoU(Intersection over Union) を図 3.14(エラーバーは 95% 信頼区間. *: Mann-Whitney U 検定 (両側検定) $p < 10^{-27}$, †: Cliff's $d > 0.474$ [79]), F1-score を図 3.15(*: Mann-Whitney U 検定 (両側検定) $p < 10^{-11}$, +: Cliff's $d > 0.330$, †: Cliff's $d > 0.474$.) に示す．

Mean IoU は 256 ピクセル正方形画像ごとに推論 Object と最も IoU の高い手作業のマスクとの重なる面積を求めて平均値を取った．サンプルサイズは図 3.12の各日付の細胞数である．F1-score の計算は MyoSOTHEs [8] にならい予測と真値の IoU の各閾値で Precision と recall を計算してから求め, IoU thresh=0.7 のときの値を示している．このときの IoU 閾値は図 3.16を参考に各日付の差が顕著になる値として採用している．各検定については Cyto [52] の結果を Control とした．サンプルサイズは WSI から 256 ピクセル正方形画像を切り取った枚数であり 100 から 124 である．

図 3.14, 図 3.15の定量結果から提案手法が統計的優位であることが確認できる．

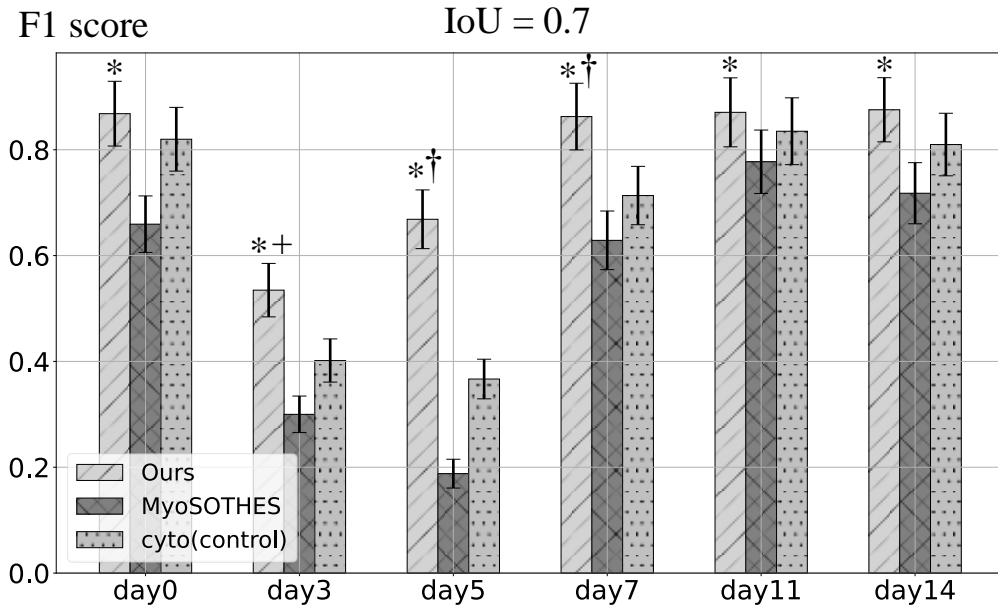


図 3.15 セグメンテーション F1-スコア (文献 [14] の Fig.1(e) 再掲)

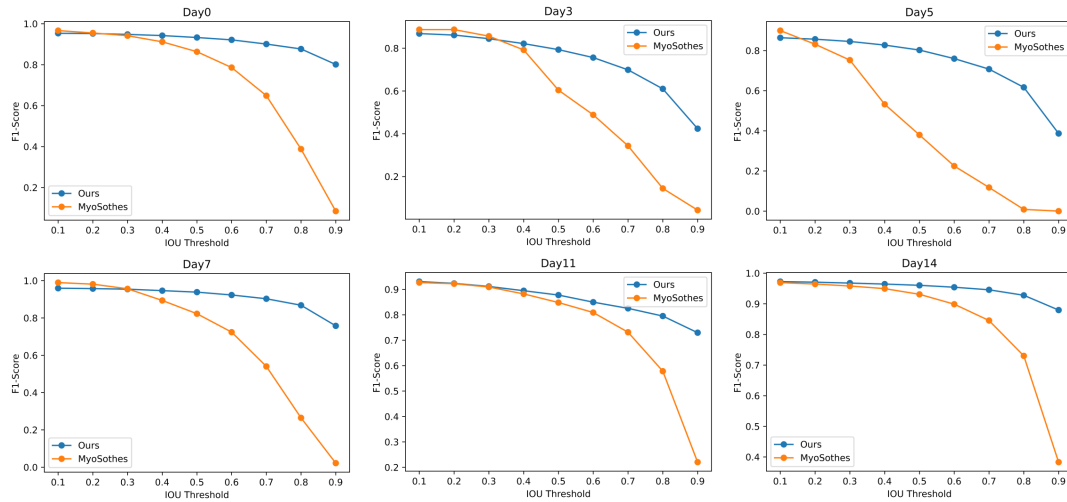


図 3.16 日付ごとの IoU と F1-score

なおコルモゴロフ・スミルノフ検定では正規性が確認できなかったため、ノンパラメトリック検定を行って p 値と効果量を導出した。Mann-Whitney 検定による p 値, Cliff のデルタ [79] を用いた効果量, およびそれらの信頼区間は, Python ライブラリ `scipy.stats` (version 1.7.3) を用いて計算した。グラフは Python ライブラリ `matplotlib.pyplot` (version 3.5.1) を用いてプロットした。

3.3.3 再生過程の筋細胞クラス分類

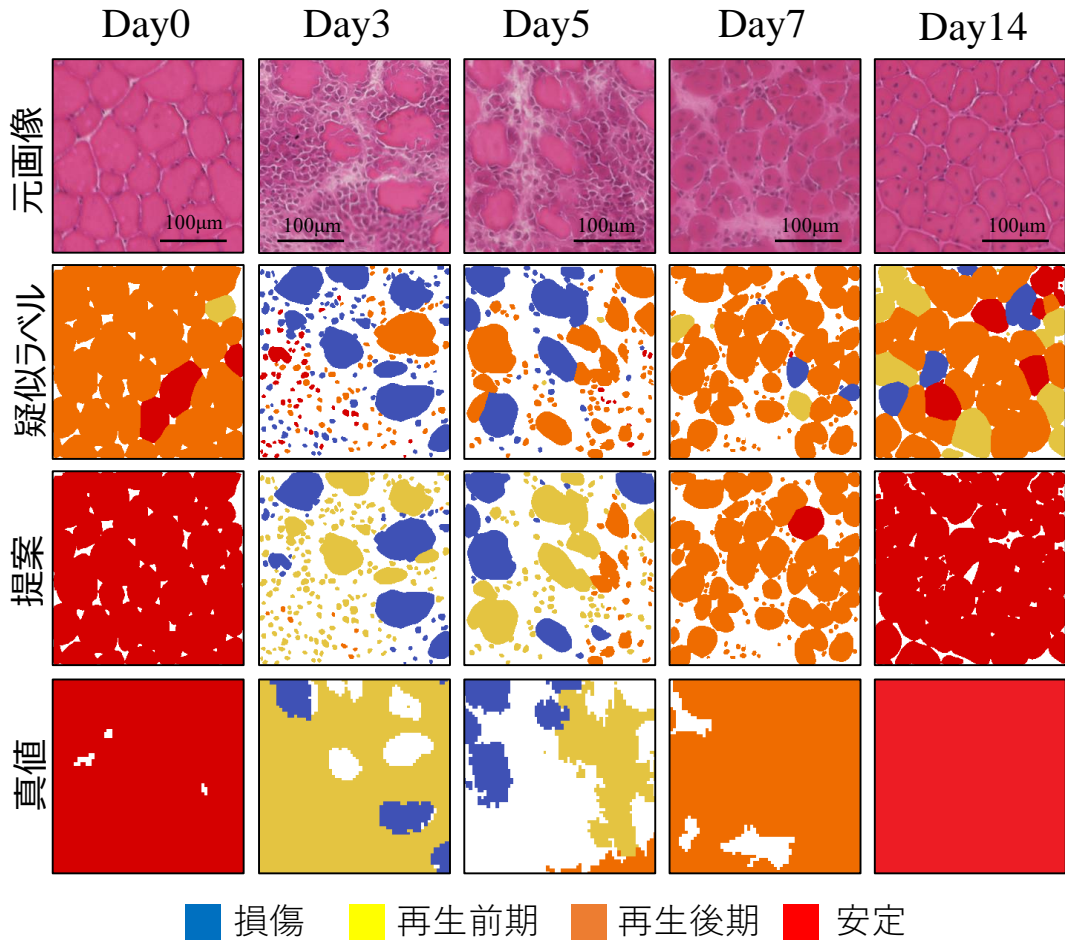


図 3.17 クラス分類結果例 (文献 [14] の Fig.2(d) 再掲)

cardiotoxin(CTX) を注入した回復過程のマウス筋組織の whole slide image(WSI)Day0, 3,5,7,14 を用いて、筋組織再生における筋線維の状態を 4 段階 (損傷前または再生完了 (赤), 損傷 (青), 再生前期 (黄色), 再生後期 (オレンジ), ただし White は判断のつかない箇所) に分けた専門家によるアノテーションと比較することで提案手法の精度を検証した. セグメンテーションモデルに入力されるときは, WSI(about 4K*4K[pix]) から 256*256[pix] の正方画像へ格子状に均等に切り分けられる. 256 ピクセル正方画像における各細胞ごとに切り分けた 64 ピクセル正方画像ごとにクラス分類を行い, 256 ピクセル正方画像の色塗りを達成する. 図 3.17には 256 ピクセル正方画像におけるクラス分類結果の例を示している. セグメンテーションに用いた Cellpose は

Diameter=5 または None(default) の設定でセグメンテーションタスクを行い、検出した Cell ごとに 64pix 正方画像に切り取って特徴量抽出とクラス推論を行う。Diameter=5 では小さな細胞を検出しており第 0 層の分類結果に用い、Diameter=None(default) の設定では筋線維程の大きさの細胞を検出し第 1 層の分類結果として用いる。出力結果は第 0 層の上に第 1 層を重ねることで得られる。すなわち色が競合したときは第 1 層の結果を優先する。

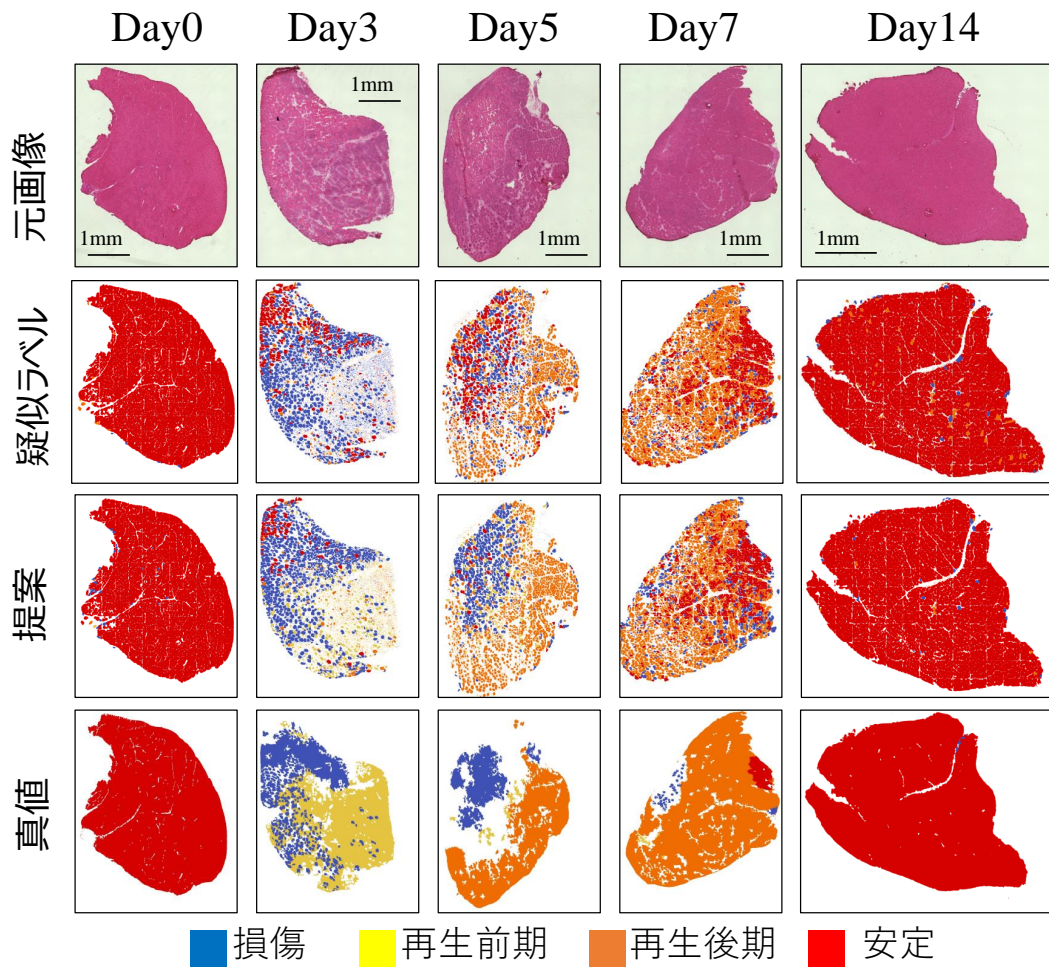


図 3.18 WSI に統合後のクラス分類結果 (文献 [14] の Fig.2(a) 再掲)

撮影した Width:2877~4606, Height:2720~4355, 1250 万画素の Whole Slide Image に対するクラス分類結果を図 3.18に示す。これは図 3.17のような各 256 正方画像結果を WSI に並べ直すことで得られる。

Cellpose で検出したオブジェクトごとのクラス推論結果を混合行列の図 3.19に、各クラスごとに Recall, Precision, F1-score の結果を図 3.20 3.21 3.22に示す。

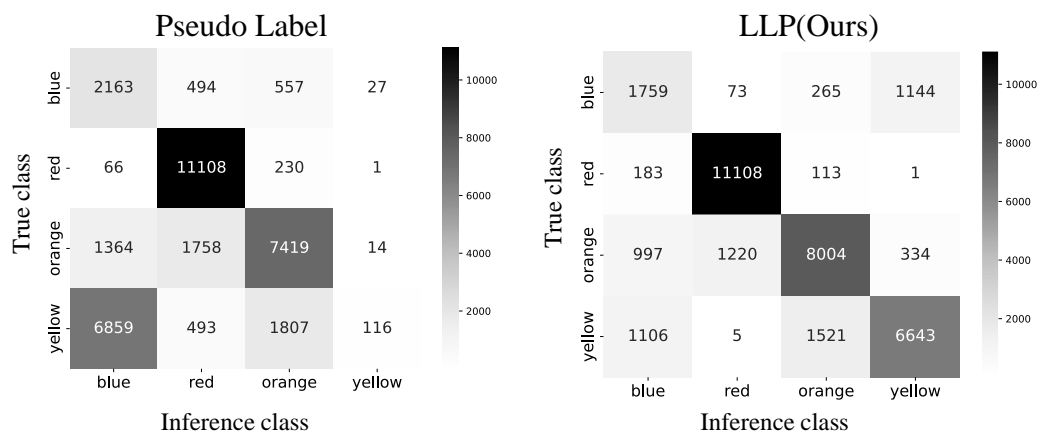


図 3.19 クラス分類の混合行列結果 (文献 [14] の Fig.2(b) 再掲)

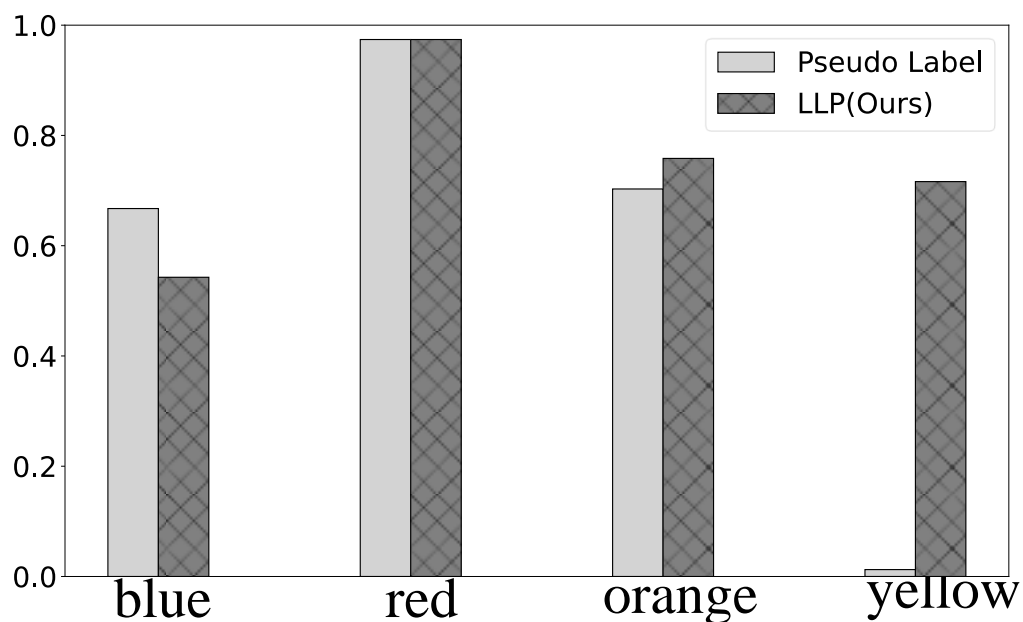


図 3.20 クラス分類の recall(文献 [14] の Fig.2(c) 再掲)

また学習データと Test データを入れ替える 3 分割交差検証を行い汎化性能を確認した。このときの各テストデータを推論したときのクラス分布と真のクラス分布とのカルバック・ライブラー情報量を表 3.1にまとめる。また WSI 画像での日付ごとの推論結果を図 3.23に示す。定量データである図 3.20, 3.21, 3.22ではほとんどのスコアについて疑似ラベル [80,81] より優位な値を確認した。Recall の青クラスについては疑似ラベル法に軍配が上がるが、混合行列の結果図 3.19からも判る通り、青を過剰検出しているためであることに注意する。

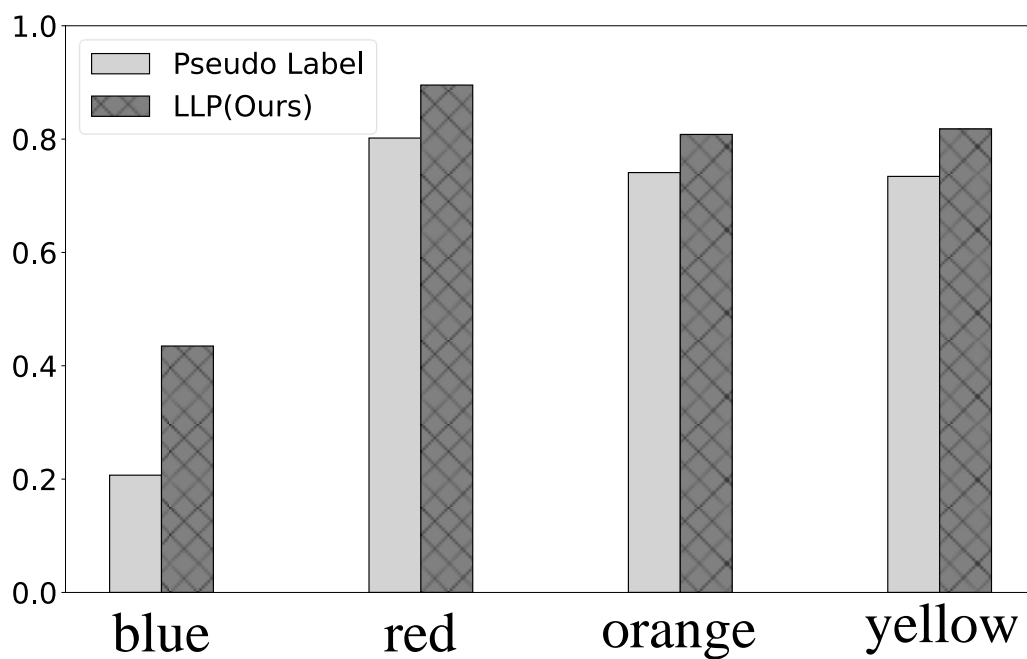


図 3.21 クラス分類の precision(文献 [14] の Fig.2(c) 再掲)

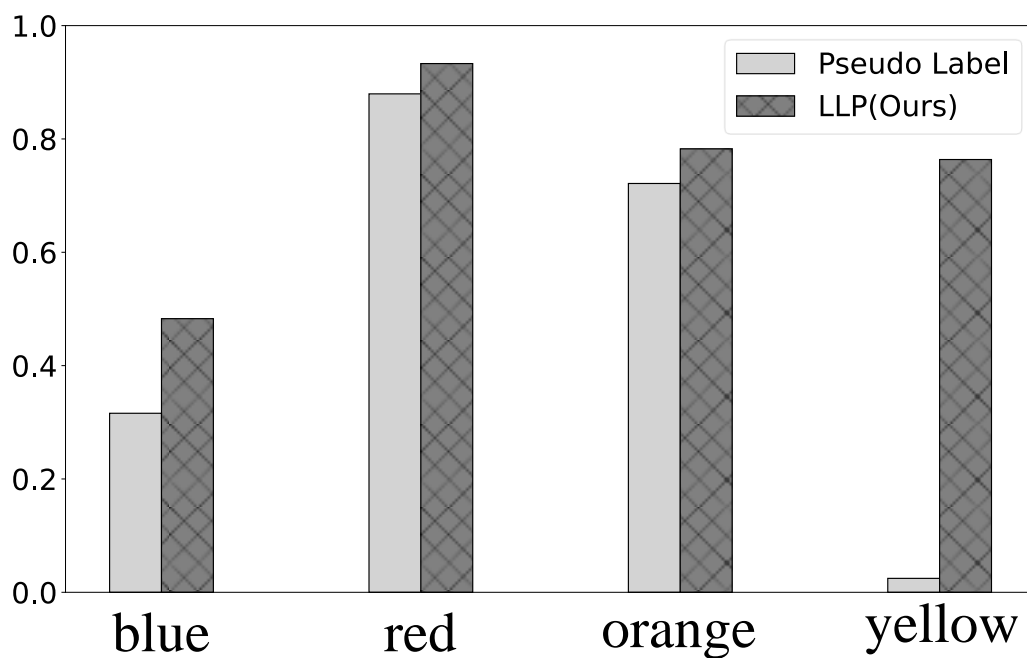


図 3.22 クラス分類の F1-スコア (文献 [14] の Fig.2(c) 再掲)

定性結果図 3.18においても疑似ラベル法では Day3,5 で真値と異なる色塗りがなされていることが判る。また限られたデータの中で分割検証を行い汎化性能についても確認する

ことができた。

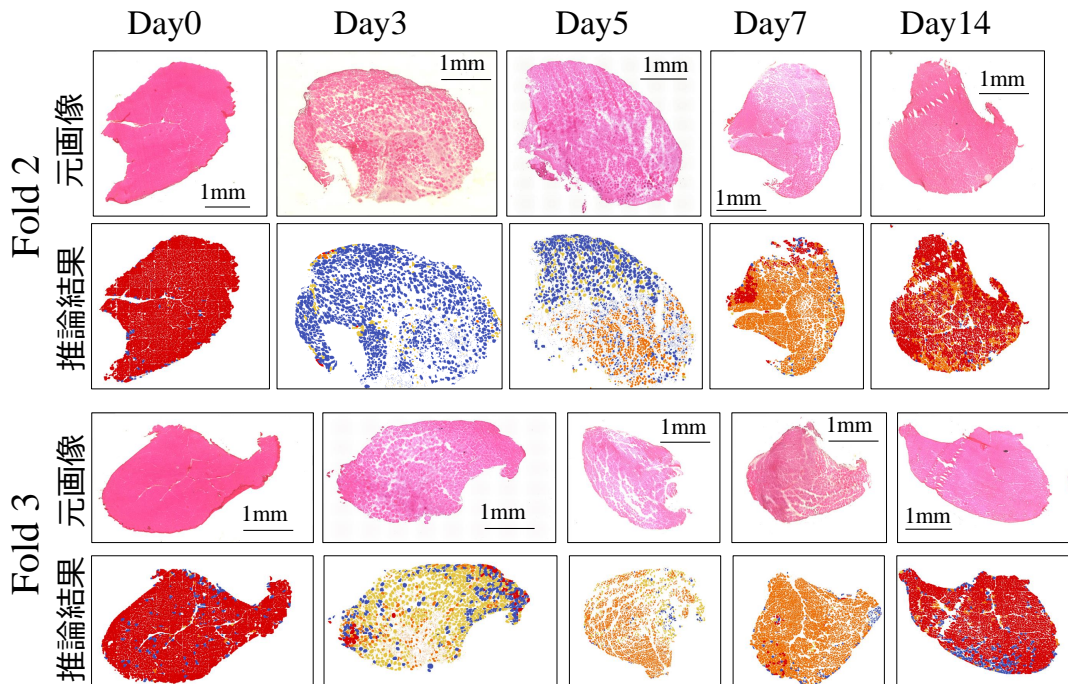


図 3.23 3 分割交差検証 (文献 [14] の Fig.2(e) 再掲)

表 3.1 3 分割検証による分類の汎化性能評価

$D_{KL}(\hat{\mathbf{p}} \mathbf{p})$	Day0	Day3	Day5	Day7	Day14
Fold1	0.021	0.385	0.208	0.074	0.016
Fold2	0.058	0.103	0.274	0.031	0.185
Fold3	0.063	0.246	0.466	0.105	0.174
mean	0.047	0.245	0.316	0.070	0.125

3.3.4 CTX とグリセロールの回復経過解析

これまで glycerol と CTX の 2 種類の注入による組織の壊死と再生には有意差があることが報告されており [9, 64–68], glycerol では脂肪細胞による再生阻害が知られている。事前の実験として、図 3.24 の通り健常なマウスにグリセロールを注入してから 7 日目と 14 日目の筋組織の HE 染色及び免疫染色の結果図 3.24 を示す。

7 日目の白がコラーゲン染色、赤がペリリピン (Perilipin) 染色、緑がラミニン染色、青が DAPI 染色の結果である。ペリリピンは脂肪生成 (adipogenesis), コラーゲンは線維

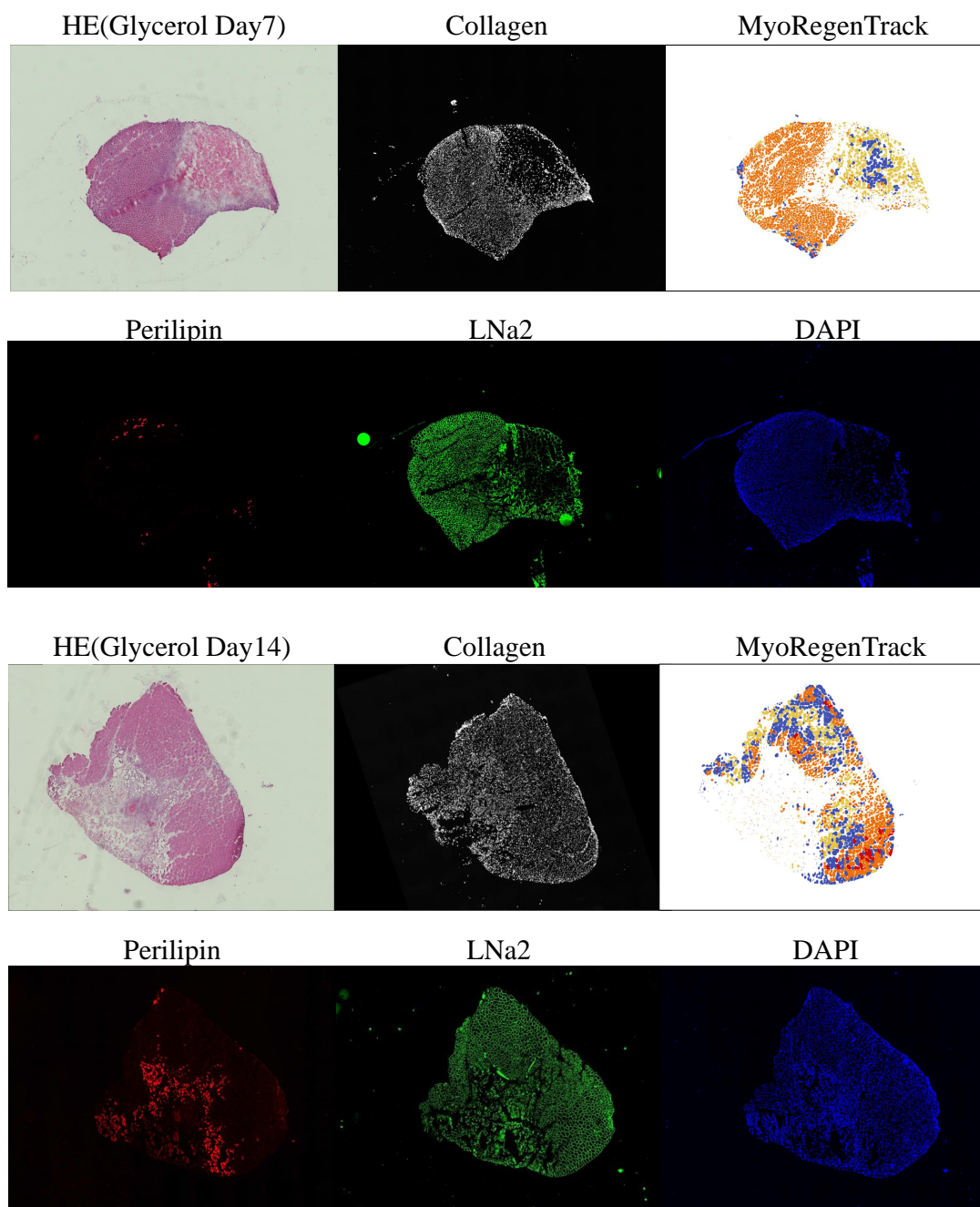


図 3.24 グリセロールの回復傾向例

化 (fibrosis) の発生を示している。なおラミニン染色は細胞膜，DAPI は細胞核を染色する。図 3.24よりペリリピンとコラーゲン染色が Day7 から Day14 にかけて広がっていることが判る。このことからグリセロールの回復傾向は日にちが経過したからといって組織全体として回復するわけではなく，部分的に回復が進行しその他の領域については炎症 (inflammation) が起こることが判る。すなわち日数が経過したからといって回復程度が

向上するとは限らないのである．CTX は日数が経過するに従って回復することが知られており，一般に 7 日目に比べて 14 日目は組織全体が回復する．

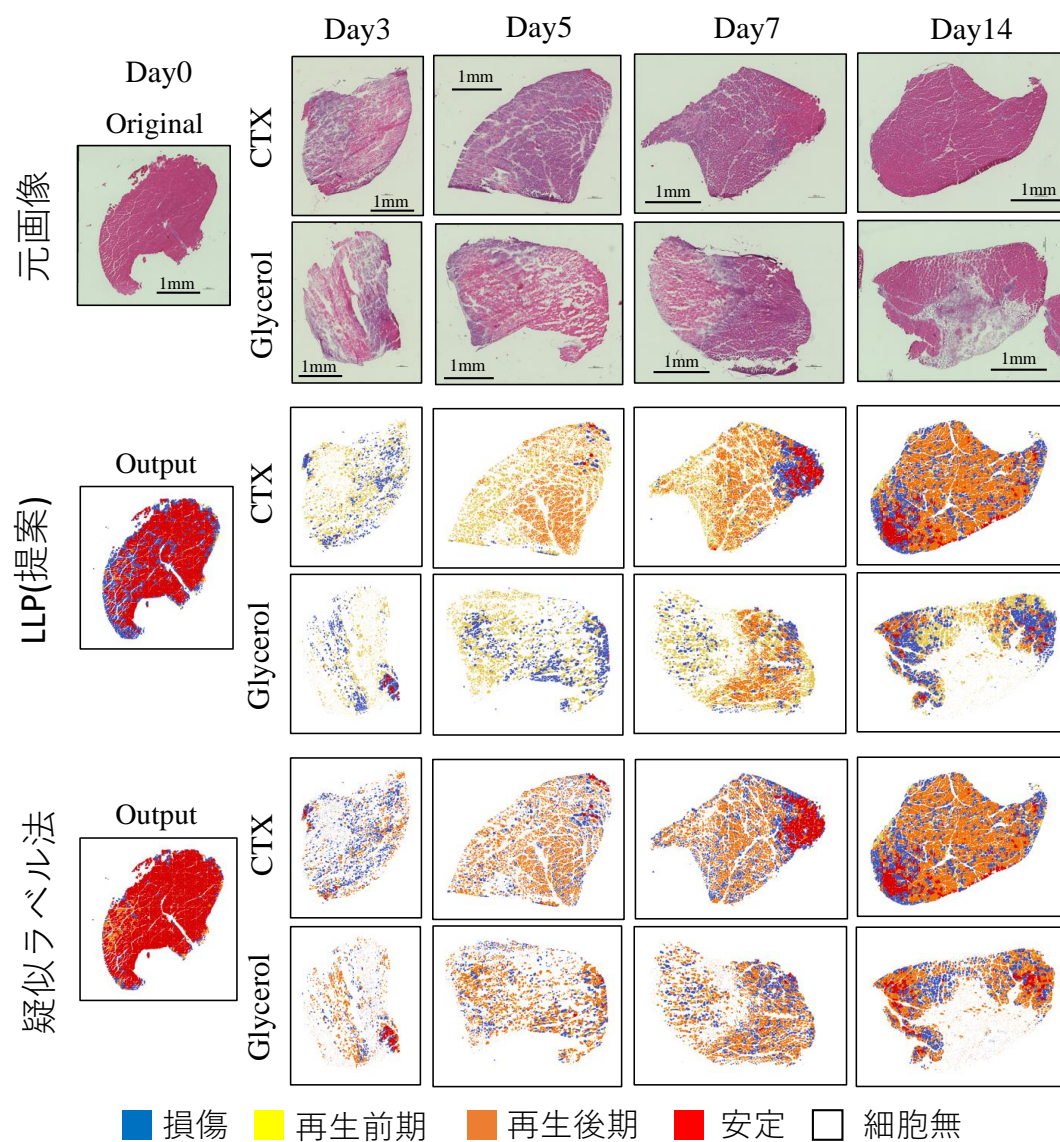


図 3.25 CTX とグリセロールの回復傾向の解析 (文献 [14] の Fig.3(a) 改変)

そこで本研究で作成したソフトウェア MyoRegenTrack を用いて glycerol または cardiotoxin(CTX) を注入した回復経過のマウスの筋組織画像の解析を行う．図 3.25 の HE 染色の WSI の解析結果についてクラス比率学習 (Learning from Label Proportion: LLP) を用いた提案手法と，同じ弱教師を疑似ラベル法で学習した手法と比べる．提案手法では日付ごとの CTX とグリセロールの回復傾向を捉えているのに対し，疑似ラベル法では推論クラスに偏りがあることが判る．

またクラス分類から定量的に回復の程度を測る指標として従来法 (cross-sectional area) と Recovery Score を計算し、2 次元上に散布図 3.25 および薬剤投与からの経過日数に対する折れ線グラフ 3.27 を示す。CSA は Recovery Score と比較するために正規化を行う。

$$\text{Cell Area Rate} = \frac{\text{細胞面積}}{\text{染色面積}} \quad (3.11)$$

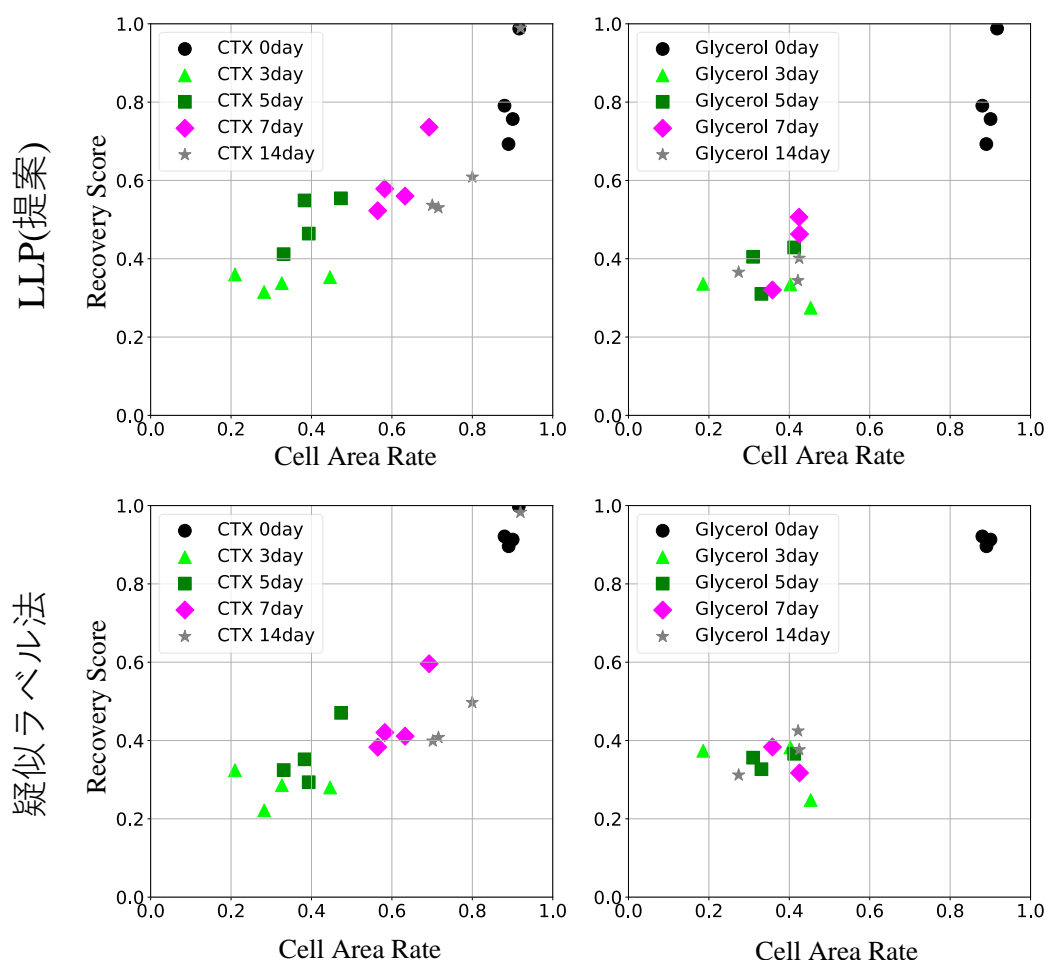


図 3.26 リカバリースコアと細胞領域割合の 2 次元散布図 (文献 [14] の Fig.3(b) 改変)

図 3.27 の結果ではセグメンテーションのみから得られる組織の細胞面積割合 (Cell Area Rate) とクラス分類によるリカバリースコア (提案法) の日付経過を観察した。提案手法ではリカバリースコアによって Day3 と Day5 の回復傾向が CTX と Glycerol 間で異なることが図 3.27 の灰色の薄枠部から判る一方、線維セグメンテーションの情報のみから得られる Cell Area Rate では区別が付けられないことが判る。

このことから Cell Area Rate と併用してリカバリースコアを用いることはより詳細な細胞状態を知ることに関わり、回復傾向の評価に有効であることが判る。また疑似ラベル法については Glycerol の方が Day3 において回復傾向が高いという誤った結果が得られた。これは図 3.25の結果からも判る通り疑似ラベル法はオレンジや赤を推論する傾向が強く、Day3に見られる損傷状態や再生前期の細胞特徴を捉え切れていないことが原因である。従って Glycerol と CTX の比較実験においても、疑似ラベル法に比べて LLP 法が優れていることが判った。ただし LLP 法においても Day0 の推論時に損傷クラスを推論しているなど課題が残る。

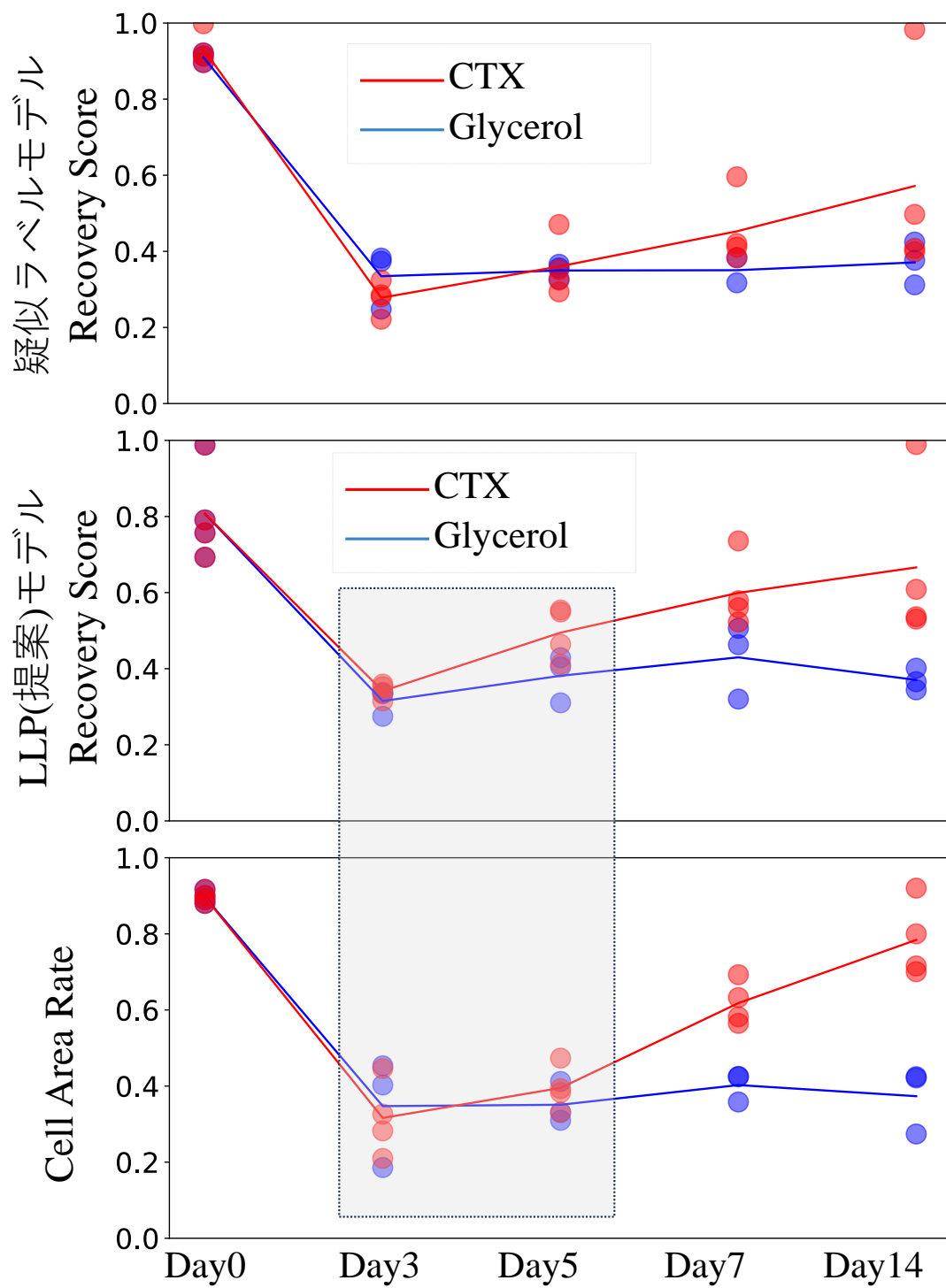


図 3.27 リカバリースコアと細胞領域割合の経過日数ごとの値 (文献 [14] の Fig.3(c) 改変)

3.3.5 免疫染色と手動アノテーションの比較

この手動アノテーションと提案したモデルを組み込んだオープンソフトウェアである MyoRegenTrack の妥当性は免疫染色によって得られたプロテインマーカー図 3.28 3.29 3.30と比較することで確認する.

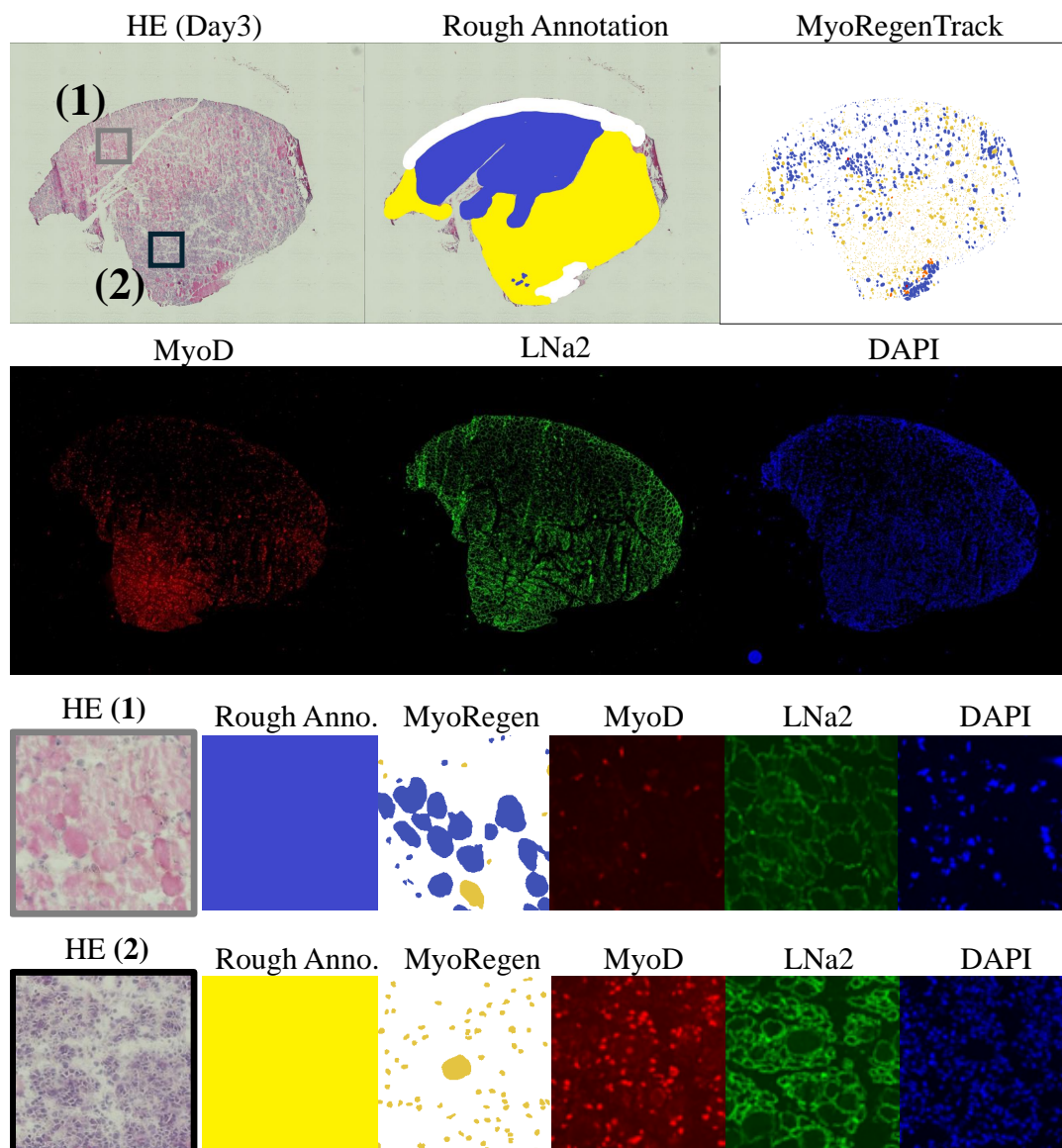


図 3.28 CTX 注入 3 日後の筋組織連続切片画像群 (文献 [14] の Fig4.(a) 改変)

図 3.28は CTX 注入 3 日後の TA 連続切片の免疫染色画像である．赤は MyoD の発現を示しており，緑はラミニン染色による線維膜，青は DAPI 染色による細胞核を表す．ラフアノテーションおよび MyoRegenTrack の結果は HE 染色画像のみから出力されてお

り、安定または再生完了（赤）、損傷（青）、再生前期（黄色）、再生後期（オレンジ）を示している。青い領域は MyoD の発現が少なく、細胞膜も崩壊直前であるということがラミネン染色の画像からも判断できる。MyoD の発現は再生初期に見られ小さな核が DAPI 染色の結果からも判り、萌芽細胞 (Myoblast) の数が多く再生初期の特徴を捉えている。

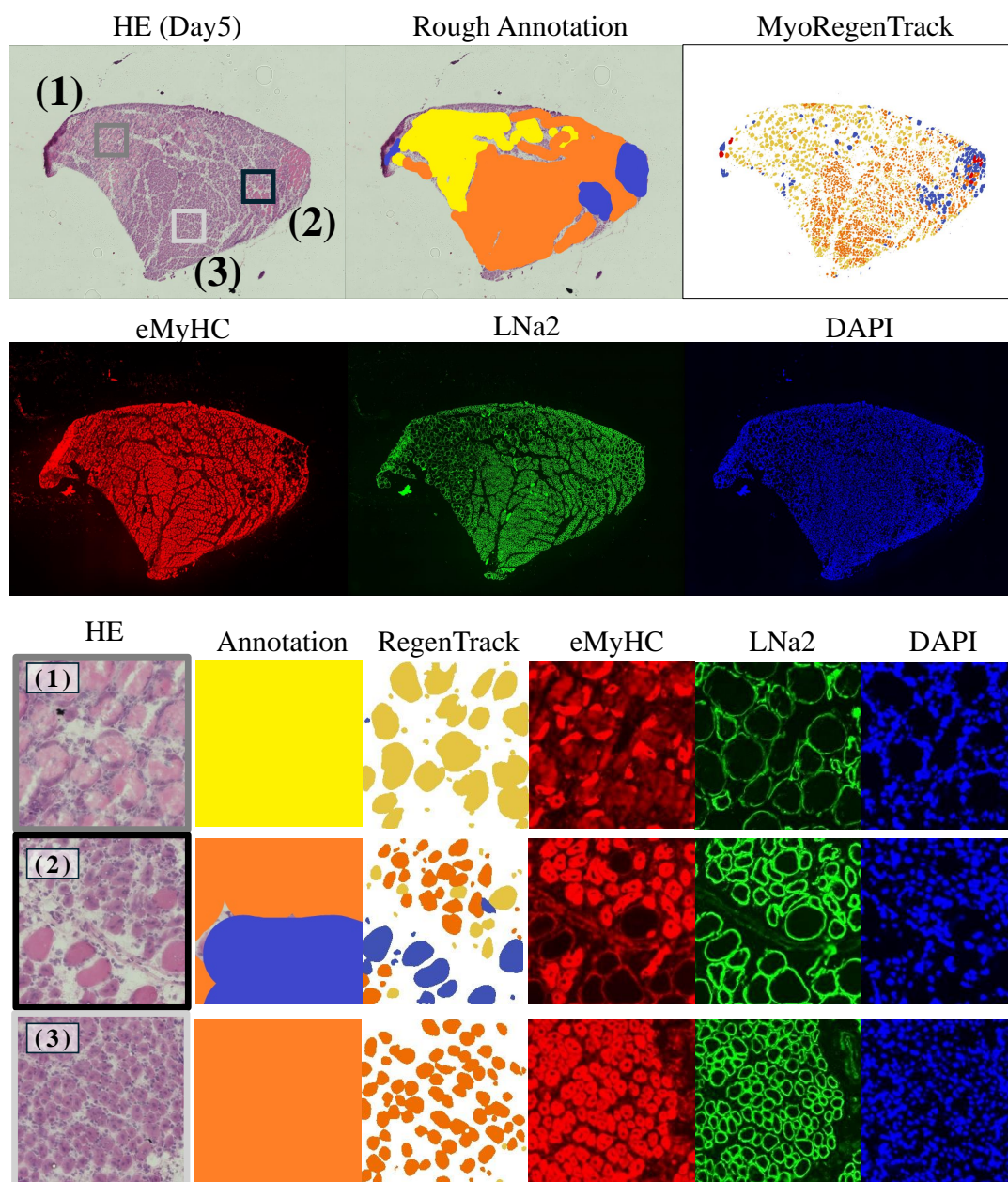


図 3.29 CTX 注入 5 日後の筋組織連続切片画像群 (文献 [14] の Fig.4(b) 改変)

図 3.29は CTX 注入 5 日後の TA 連続切片の免疫染色画像である。5 日目に入ると組織のいたるところで再生が進んでいることが HE 染色の結果から判る。また免疫染色につい

ては赤が胚性ミオシン重鎖の強弱を表している。緑は引き続きラミニン染色による線維の膜、青は DAPI 染色による核を表す。線維膜がはっきりとしておりかつ eMyHC の赤色が強いということは筋管細胞が成長していることを示しており、再生後期の特徴を示している。

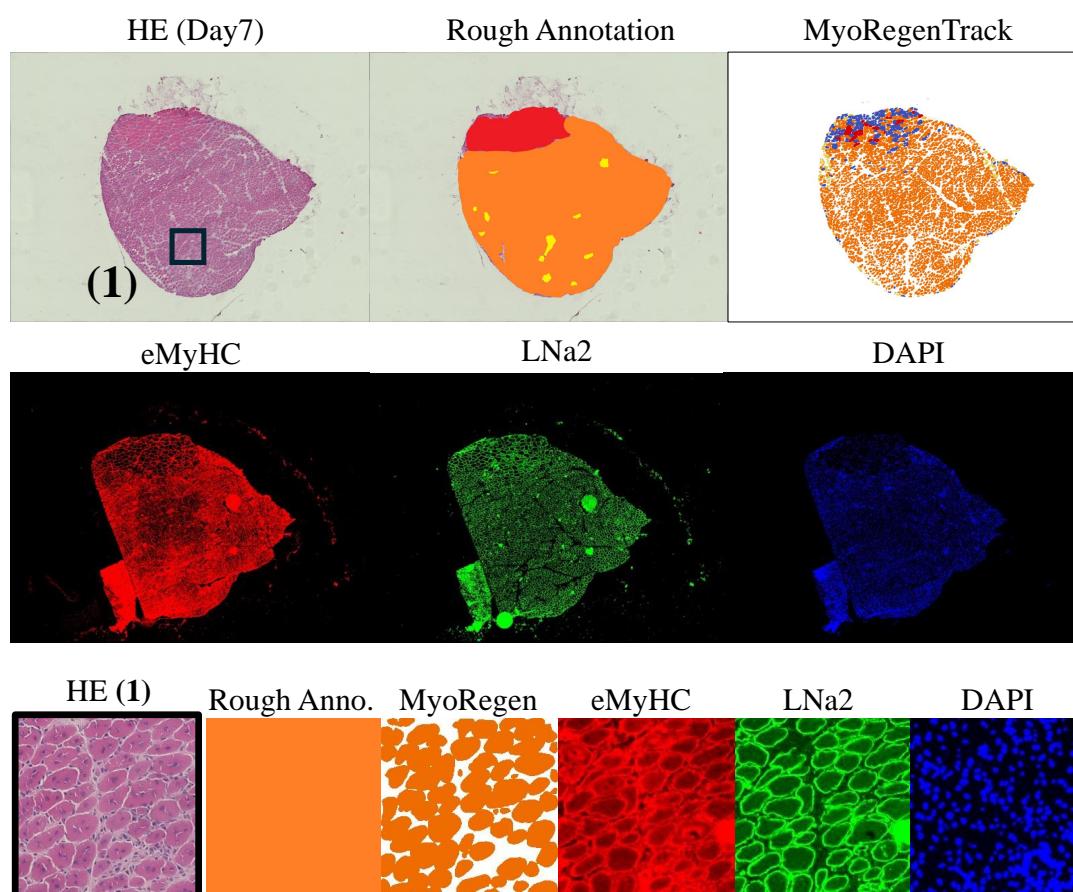


図 3.30 CTX 注入 7 日後の筋組織連続切片画像群 (文献 [14] の Fig.4(c) 改変)

図 3.30は CTX 注入 7 日後の TA 連続切片の免疫染色画像である。7 日目では 5 日目に比べて回復が進んでいることが判る。その証拠に組織の多くの領域で eMyHC の赤色が見られ、線維間に隙間がありつつもラミニン染色による細胞膜から線維面積が大きくなっていることと判る。この結果はラフなアノテーションとも一致している。

上記の免疫染色と HE 染色へのラフなアノテーションとの比較から、HE 染色画像のみから人間の目であれば再生期の判断を下すことが可能であることが判る。よって専門家によるクラスアノテーションの妥当性が認められるとの結論で本節を閉じる。

3.4 議論

3.3.2節の結果では再生過程ドメインに適応したファインチューニングモデルの性能について確かめている．提案手法は萌芽細胞 (Myoblast) や筋幹細胞 (Myotube) を手作業に近い精度で検出できている一方，表 3.13-Ours-Day7,11,14 の結果を見れば，染色部でない画像部分でも細胞が検出されてしまっていることが伺える．これはモデルが萌芽細胞などを検出する為に過剰検出をする状態になっていることを示しており，染色部のエッジ検出スクリプトを加えることで空白地帯の検出については無視できる一方，モデルの性能及び信頼性という観点では疑問符が残ってしまう．学習データに空白地帯を用意し細胞がないという情報を明示的にモデルに学習させることで解決が見込める．

3.3.3節の結果では回復過程の細胞とその周辺情報から回復期を分類するモデル評価を行った．これまで組織の回復過程を解析するとき，図 3.18のように Whole Slide Image(WSI) を用いた解析では組織全体を概観することには適している一方細部の各線維の状態が判りづらく，図 3.17のような Clip 画像であれば各線維の状態は見やすいが組織全体が把握できない．このジレンマの一つの解決策はすべての Clip 画像を掲載することであるが紙幅の観点から好まれない傾向にあり，組織全体の線維の回復状態を定量的に解析する手法が望まれていた．クラス分類は一つの解決策であり実際筋組織を全体にわたって俯瞰する解析によって各領域のおおよその傾向を把握することができる．今回用いた手法はセグメンテーション＋クラス分類であるため図 3.17(Day3,5) に見られる Cell が検出できない領域を塗ることはできない．

3.3.4節では分類性能を確認したモデルを解析ソフトウェア MyoRegenTrack とし，その実用性について既にこれまでの薬学研究で良く知られた CTX とグリセロールの回復傾向の解析に用いた．図 3.25ではクラス分類を加えた解析により Day5 の CTX と Glycerol 間に回復経過の差異があることが求められたが，一方で提案手法の汎化性能については課題も見つかっている．これまでの筋線維クラス分類は明示的 (explicit) な線維の特徴量 (Area・円形度・Feret 比) と見分けたい線維タイプの相関が事前知識としてあった．しかし明示的な特徴量のみでは識別しづらい萌芽細胞や筋幹細胞のように暗黙的 (implicit) な特徴量を持つ線維を分類するときは機械学習的手法を用いた多次元特徴量による表現力が必要である．ただし暗黙的 (implicit) である故に AI 全般に言える説明可能性の問題は残っており，分類器の結果は必ずしも人間の目と同等である保証はなく，実環境で

想定外のドメインギャップが生じた際に精度の担保や改善を行うことは難しい。実際、図 3.25-Day0 では青色を示す損傷領域は存在しないはずであるが、動物処理における凍結処理に不足があったために線維に気泡のような空白が生まれてしまい、本ソフトウェアでは無傷な線維を損傷状態と誤判定してしまうことが見られた。凍結処理に大きく失敗した例については今回解析対象から外しており、図 3.31のように MyoRegenTrack が無傷の線維を損傷と判定してしまうことが判る。

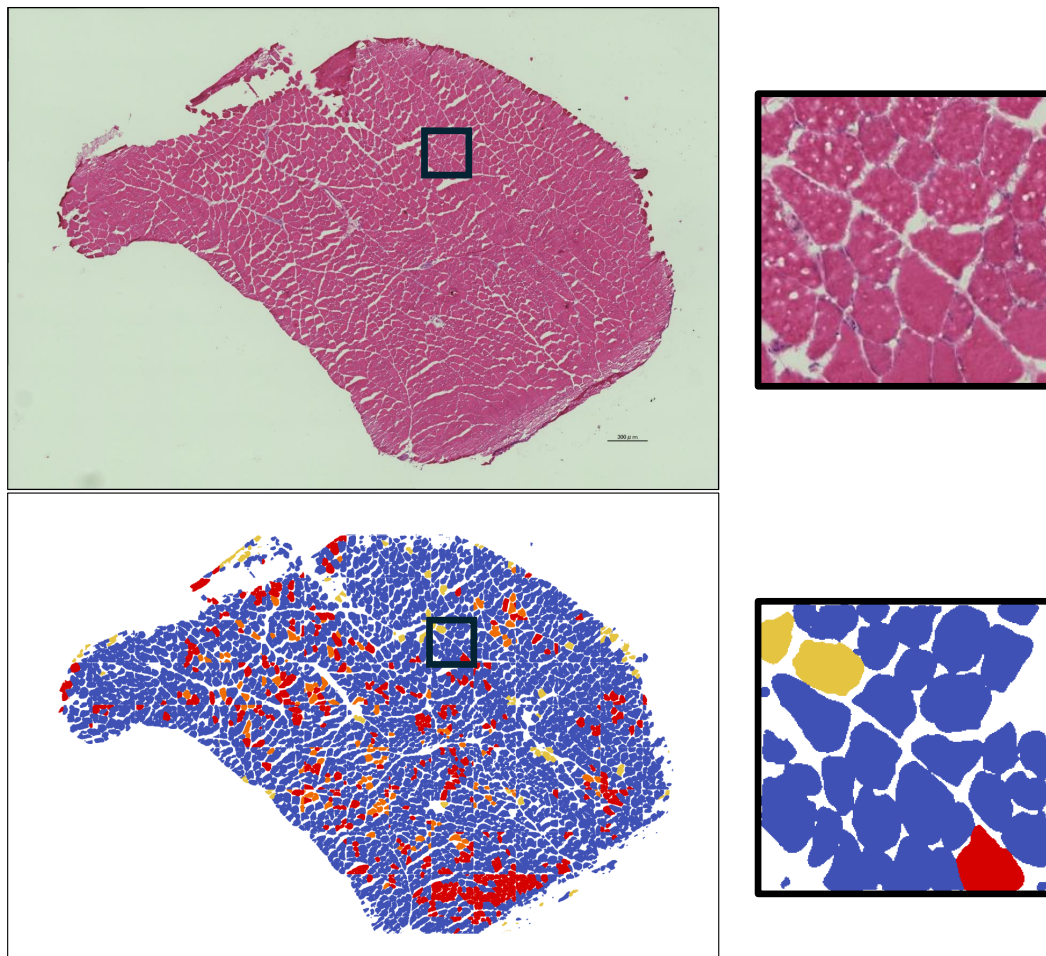


図 3.31 凍結処理に失敗した画像の解析例

幅広いドメインに対応できる頑健性を得るには学習データセットのドメイン幅を広げることが手っ取り早いですが、動物処理や光学条件などあらゆる条件に対応したデータセットを構築するためには多大な労力が必要である。または各種手続きや環境条件に依存しない細胞の分類の根幹となる特徴を抽出できるように、特徴量抽出器 (backbone) のファインチューニングを行うということが考えられる。ただし LLP では特徴量空間での損失計算

が出来ず、筋細胞に合わせた特徴量抽出器を得られているか不確かである。本ソフトウェアについては今回多様な入力ドメインの汎用性について担保することはできず、入力画像は限定的なドメインに留まる中で、組織解析への活用が期待される。

3.5 結言

我々は回復中の筋組織にあるセグメンテーションと再生段階の分類を行う MyoRegenTrack を開発した。セグメンテーションは事前学習モデルをファインチューニングすることで得られその頑健性は先行研究のモデルと比較することで確かめられた。分類モデルの学習には日付ラベルと紐づくクラス比率による Learning from Label Proportion(LLP)を用いることで、専門家と遜色のない精度を担保することが判った。ただし推論パイプラインそのものの問題として、細胞セグメンテーションされていない領域を分類し色を塗ることはできない。オブジェクトの検出と領域検出は従来別々のモデルで推論することが一般的であり、より適切な色塗りを行う場合は領域クラスを追加することで達成できる。開発した MyoRegenTrack を回復過程の異なることで知られる CTX とグリセロールの注入経過観察実験における解析で用いて、回復経過の違いについて先行研究と一致した結果が得られ有効性を確認した。しかし画像を用意する際に検体処理での凍結不良など、学習データにないドメインには対応できない。今後はより汎化性能を高めるためにデータセットの増加や新しいデータ拡張法、そして特徴量抽出器の検討をする必要がある。

第 4 章 類似比率学習による筋細胞へのドメイン適応

4.1 緒言

筋組織の損傷再生過程の評価は加齢や疾患に関連する筋力低下のメカニズムを解明するための生物学的研究 [39, 82, 83] における実験操作の効果量を測定する上で基本的な解析である。筋組織の損傷と再生を誘導するコブラ毒 (cardiotoxin; CTX) [43] をマウスの下腿筋に注射すると、筋線維の壊死が発生すると同時に筋組織が持つ再生機能が働く。筋組織の回復速度は局所性を示しているため [59, 60]、組織回復を評価するためには特定の領域ごとに細胞の状態を評価する必要がある一方 [13]、筋組織画像は組織全体を大量のピクセルによって画像化された WSI であり、無数の細胞と筋線維を含むため労力を要する。こういった WSI の各領域をコンピュータで解析・評価する方法は複数インスタンス学習 (Multiple instance learning: MIL) [25] と呼ばれている。

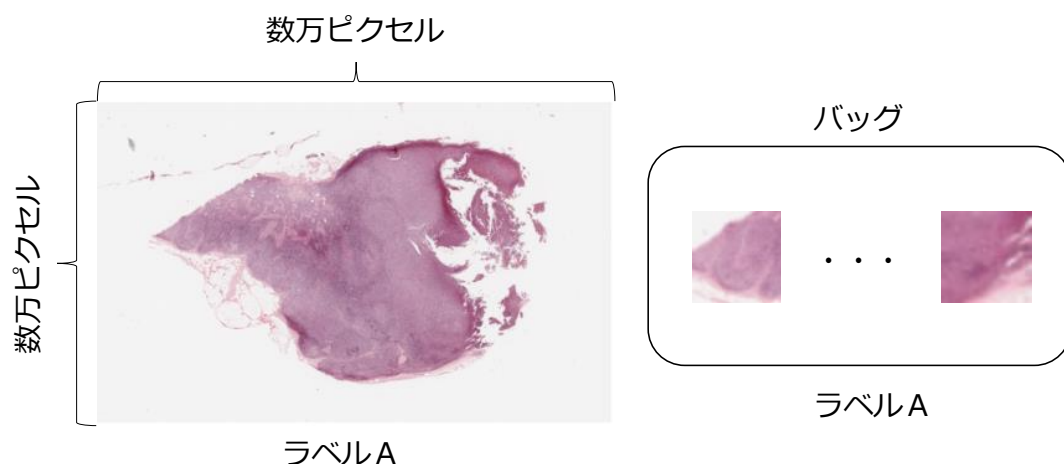


図 4.1 複数インスタンス学習 (Multiple instance learning: MIL) の例 [15]

たとえば図 4.1 の癌腫瘍例 [15] のように数ギガピクセルの画像は直接コンピュータ上で扱うことができない。よって各領域のアノテーションは難しいが画像一枚には悪性腫瘍かどうかといったラベルが与えられているため、画像を各領域に切り分けて集めたバッグと呼ばれるインスタンス群ごとに学習し、予測器を作ることを目指す。WSI の大きな画像サイズがもたらす GPU メモリの制限を回避し、WSI を切り分けた各領域においてクラス分類することで画像解析の説明可能性を向上することができる。このように WSI の MIL ではしばしば弱教師あり学習の手法を用いる。

クラス比率からモデル学習する LLP [27, 76, 78, 84–92] は、WSI における MIL の文脈において主要な弱教師あり学習の 1 つである。これはクラス比率の弱教師を用意することが比較的容易であることが理由であり、WSI における LLP 適用例として腫瘍の壊死の二値 [78] または三分類 [1] や、関心領域のスクリーニング [93] が挙げられる。しかし LLP は現在、本研究で扱う再生過程の筋組織を評価する際に 2 つの課題に直面している。

- 特徴量抽出器の更新が不可

LLP はバグ内の全インスタンスを予測した後に損失関数を計算するため、通常事前学習済みのモデルを特徴抽出器 (backbone) として使用し学習時はそのパラメータを固定し [94]、分類層 (head) のみを更新する [1, 78](図 4.2 参照)。DINO [53] のような大規模な事前学習モデルは一般的な対象物で訓練されているため、医療用 WSI タスクの特徴抽出には適さない可能性が残る。

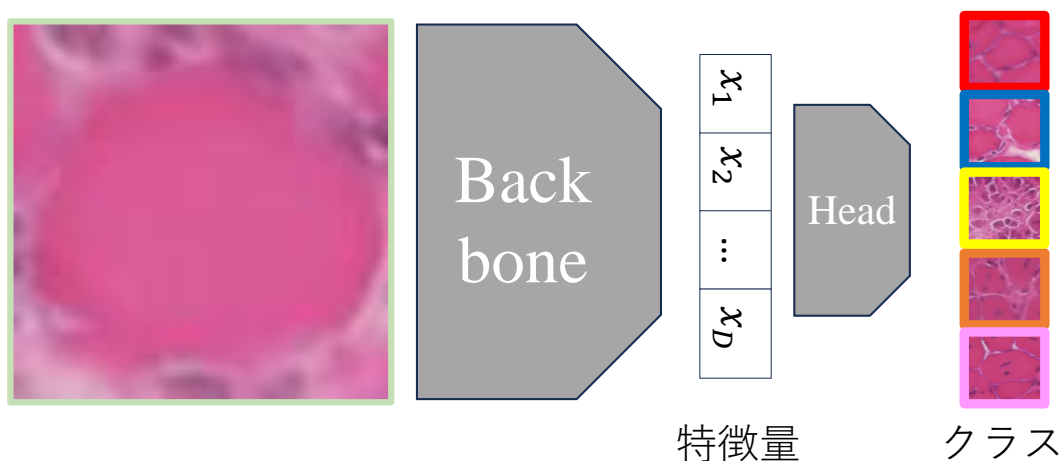


図 4.2 クラス分類のパイプライン

- クラス類似情報の喪失

再生過程筋組織の細胞は CTX 筋注から時間の経過とともに形態が徐々に変化し [59]、各段階間に生物由来の形態変化の順序が存在するが、LLP ではクラスを名義尺度として扱い、順序情報が損なわれる。具体的には再生する細胞には次の段階が含まれる。無傷の筋線維、ゴーストファイバー (CTX 注射後の基底膜が残存している細胞)、萌芽細胞 (サテライト細胞が育った細胞)、筋管細胞 (筋芽細胞が筋分化したもの)、および再生した筋線維。これらは再生時間順序に応じた類似性を持つ。クラスを順序尺度として扱うことで、再生下の細胞類似性を損失関数に反映させる。

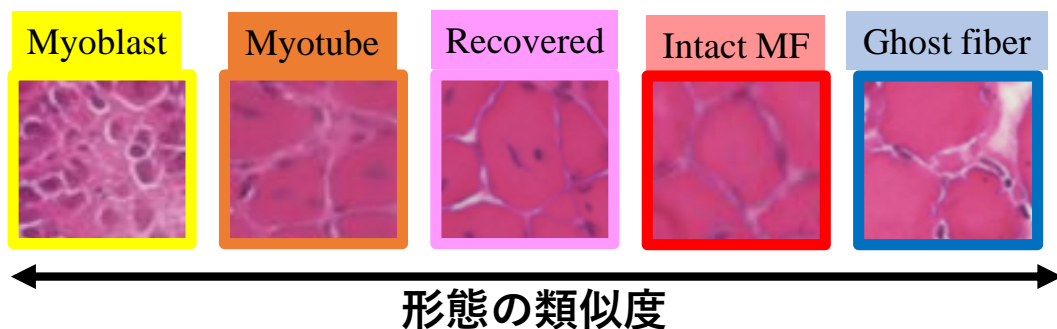


図 4.3 クラスの類似性

大規模な事前学習モデル [53] を用いて抽出した特徴量空間において，学習データおよびテストデータの特徴点分布を確認すると，データ間に距離が生じることがある．いずれも同じ再生過程の筋組織画像であるため，本来データの分布は再生過程分類に必要な特徴量差であるべきだが，図 4.4からも判る通り，例えば明るさといったドメインギャップが特徴抽出に悪影響を及ぼす可能性がある．本章では筋細胞再生分類に必要な特徴抽出を行えるように，抽出器のファインチューニングを試みる．

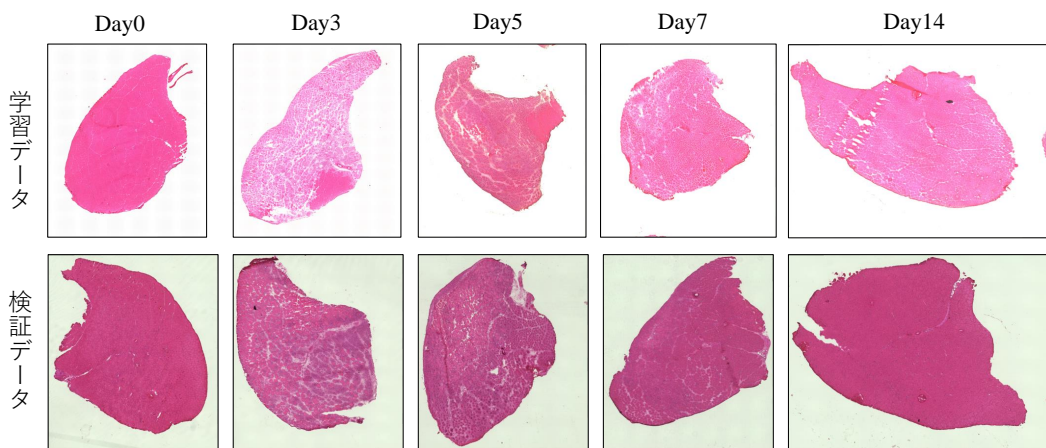


図 4.4 検証データと学習データのドメインギャップ

そこで LLP パラダイムの下でクラスの順序を考慮しながら特徴抽出器を更新する順序尺度類似比率学習 (Ordinal Scale Learning from Similarity Proportion; OSLSP [95]) を考案し，上記 1,2 両方の問題に同時に解決を試みた．

4.2 関連研究

WSI の課題を解決する手法について表 4.1に示す。クラス尺度は最終的にモデルが出力する推論結果の尺度を示す。更新目標は当該手段で更新するモデルアーキテクチャの層を示しており、Backbone はデータから特徴量抽出をするレイヤであり、Head は抽出された特徴量からクラス推定を行うレイヤになる。Loss は学習に用いる損失関数を構成する情報を示す。これまでの研究は bag 単位の学習を行う場合はクラス比率を利用した損失関数を設計し、インスタンス単位の学習を行う場合はクラス比率から疑似ラベルを各インスタンスに生成することを行っていた。インスタンス単位の学習では特徴量抽出器の更新を行える点がメリットである反面、疑似ラベルは誤ったラベル生成に繋がるノイズの可能性を指摘されていた [26]。またこれまでの bag 単位の学習では腫瘍の 2 値判定 [26, 80, 96] または 3 値判定 [1], 病理画像の注目領域の抽出 [93] などいずれも名義尺度を出力結果として扱っており、順序尺度を持つ筋組織の線維形態変化を対象とはして来なかった。WSI を対象とはしていないが Jeronimo's [77] では既に特徴量化されている患者データと着床した割合を使って Embryo Implantation Prediction を行う。DLLP [27] は LLP と DeepLearning を組み合わせた初期研究である。LLP-VAT [89] では従来の LLP ロスに加えて、個々のインスタンス画像に軽微なノイズを加えてもクラス予測に一貫性が担保されるような一貫性損失 (consistency loss) を提案している。SIM-LLP [85] では従来の LLP ロスに加えて、特徴ベクトルの異なる予測を罰するペアワイズ類似性ベースの損失が追加されている。いずれの LLP 手法についても特徴量抽出には事前学習したエンコーダを用いており、head 部の学習に焦点を当てていた。提案手法である Ordinal Scale Learning from Similarity Proportion(OSLSP) は対照学習 (Contrastive learning [30]) のようにインスタンス同士の類似度を損失関数に反映することで backbone の更新を行う。またその際にクラス類似度に注意することでクラスの順序尺度を反映することを可能とする。IIB-MIL [26] は Backbone の更新のためにクラス比率から疑似ラベルを生成し、損失関数に加えることで Backbone と Head の更新を行っている。ただし疑似ラベルが抱えるノイズの問題を解決したわけではない。なおバグを用いた学習については 付録 A 章に付録を載せた。

表 4.1 WSI の複数インスタンス分類の手段一覧.

手法名	クラス尺度	更新目標	Loss
Pseudo [80]	名義尺度	Head	Pseudo Class + Class Prop
LLP [97]	名義尺度	Head	Class Prop
MIL [96]	バイナリ	Head	Class Prop
IIB-MIL [26]	バイナリ	BackBone +Head	Pseudo Class + Class Prop
OS-LSP(Ours)	順序尺度	BackBone +Head	Sim Prop + Class Prop

4.3 類似比率学習 (Learning from Similarity Proportion) の提案

4.3.1 細胞形態の類似度

図 4.5に示す通り, 筋線維はコブラ毒 (cardiotoxin: CTX) 注入から崩壊が始まりゴーストファイバー (ghost fiber), 萌芽細胞 (myoblast), 筋幹細胞 (myotube) を経て回復し

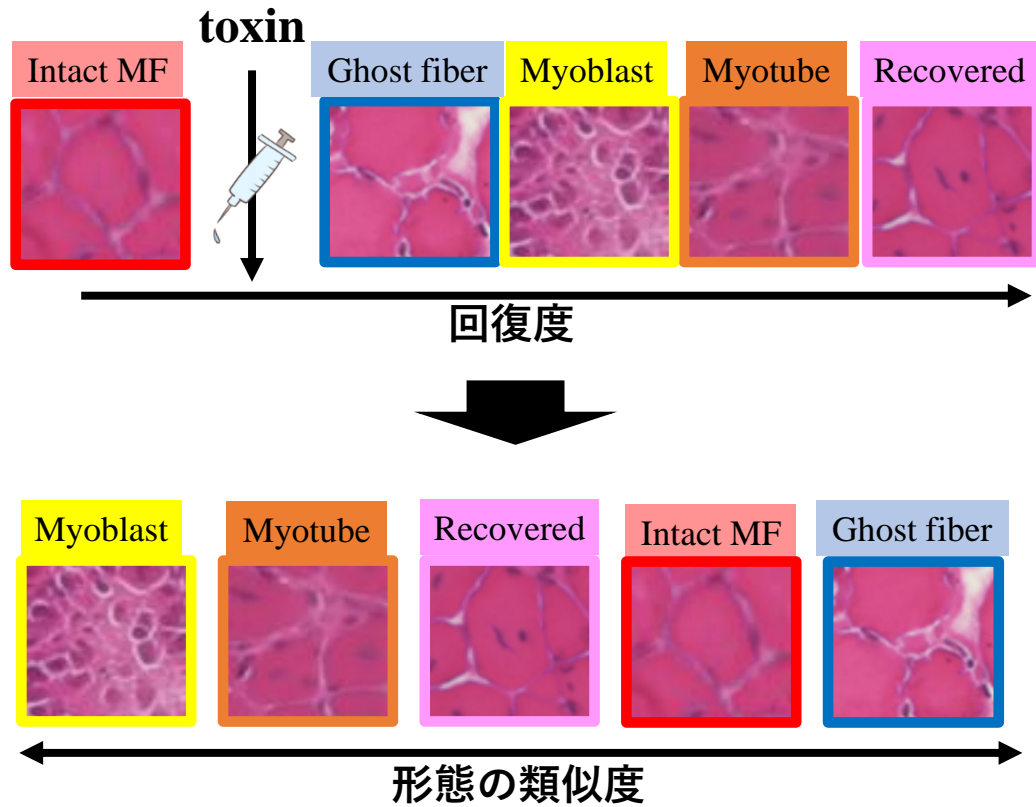


図 4.5 線維の時間経過による形態変化と類似度

た筋線維 (recovered myofiber) となる, といった時系列に沿った形態変化と再生現象を起こす. 無傷の筋線維 (Intact MF) は CTX で損傷を受ける前の安定状態を示している. ゴーストファイバーは CTX によって壊死が誘発された筋線維の基底膜が残った状態である. 萌芽細胞は damaged-myofiber-derived factors (DMDFs) [45] などによって活性化したサテライト細胞が増殖すること形成される. 細胞分裂を停止した隣接し合う萌芽細胞が筋分化することで多核の筋幹細胞となる. 筋幹細胞が成長することで線維の隙間が埋め尽くされ回復した筋線維となる. 萌芽細胞はその名の通り筋線維再生の始まりと見なすことができ, ゴーストファイバーは筋線維が壊死する段階であるため, クラス間類似度は萌芽細胞 (黄色) から見るとゴーストファイバー (青色) が最も遠い位置にある. よって図 4.5 のような形態の類似度関係を得る.

4.3.2 類似比率損失の概要

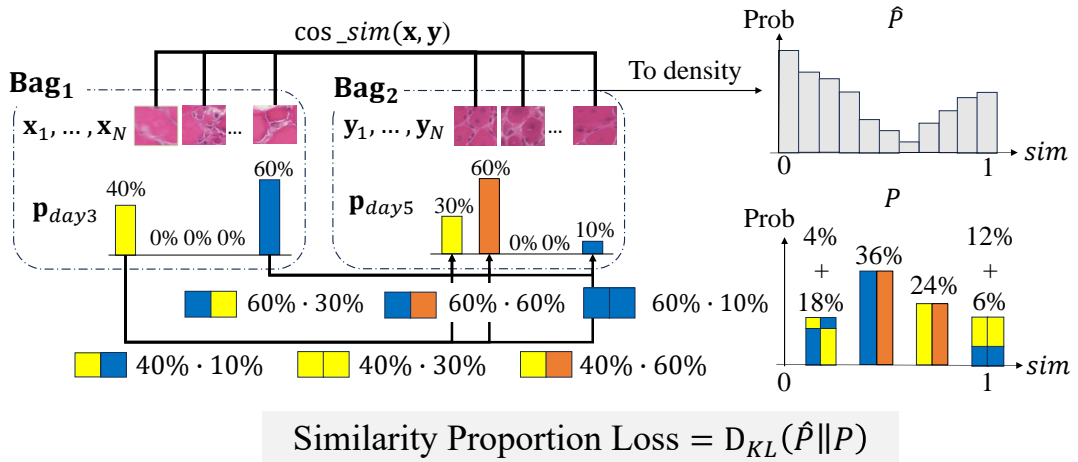


図 4.6 順序尺度類似比率学習 (OSLSP) の概要

標準的な LLP のバッグ作成方法 [90] に従い, 同じ日付ラベルのインスタンス $(\mathbf{x}) \in \mathbb{R}^D$ をバッグサイズ N 個でまとめる. 各バッグはクラス比率 $\mathbf{p}_d \in [0, 1]^K$ ($|\mathbf{p}_j|_1 = 1$) と N 個のインスタンス $\mathbf{x} \in \mathbb{R}^D$ から構成される. 図 4.6 に示す通り, 類似性比率損失 (similarity proportion loss) の計算方法は真の確率密度関数 (PDF) P と予測された PDF \hat{P} の間の Kullback–Leibler 情報量を取ることで計算される.

$$\mathcal{L}_{SimProp} = D_{KL}(\hat{P} || P) = \sum_{i=1}^{bins} \hat{P}(i) \log \left(\frac{\hat{P}(i)}{P(i)} \right) \quad (4.1)$$

P は 2 つのバッグ間のクラス比率の組み合わせから計算され, \hat{P} は 2 つのバッグのイ

ンスタンス特徴間の類似性の密度分布によって導かれる。

4.3.3 インスタンス類似度による予測密度分布

図 4.6の例では Day3 と Day5 の WSI から切り取られた線維画像によってバッグが 2 つ用意されている。Bag1 には Day3 の線維画像を特徴量抽出器に通すことで得られるインスタンス群 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} (\mathbf{x} \in \mathbb{R}^D)$ と図 3.8の方法で計算された 3 日目のクラス比率 \mathbf{p}_{day3} が与えられている。Bag2 にも同様にインスタンス群 $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\} (\mathbf{y} \in \mathbb{R}^D)$ と 5 日目のクラス比率 \mathbf{p}_{day5} を有する。similarity proportion loss を計算するとき、予測される類似度の密度関数 \hat{P} はインスタンス群から計算し、真の確率密度関数 (PDF) P はクラス比率 $\mathbf{p}_{day3}, \mathbf{p}_{day5}$ から計算される。類似度密度関数の予測 \hat{P} の導出について述べる。2 つのバッグから選ばれたインスタンス $\mathbf{x}_n, \mathbf{y}_n (n \in \{1, 2, \dots, N\})$ の類似度はコサイン類似度 $\text{CosSim}(\mathbf{x}_n, \mathbf{y}_n): \mathbb{R}^D \times \mathbb{R}^D \rightarrow [0, 1] \in \mathbb{R}$ を用いて計算する。

$$\text{CosSim}(\mathbf{x}_n, \mathbf{y}_n) = \frac{1}{2} \left(\frac{\mathbf{x}_n \cdot \mathbf{y}_n}{\|\mathbf{x}_n\| \cdot \|\mathbf{y}_n\|} + 1 \right) \in [0, 1] \quad (4.2)$$

得られた各インスタンスの比較データ $\text{CosSim}(\mathbf{x}_n, \mathbf{y}_n)$ をヒストグラム化する。max, min, bins はヒストグラム化する際のパラメータであり、データの最大最小値の設定およびヒストグラムのビン (bins) の個数を指す。コサイン類似度の値は 0 から 1 までなので今回の場合 max=1, min=0 である。ビンの個数は類似度の組み合わせ数に相当するとして、クラス数 K のうち 2 つのクラスを選ぶただし同じクラスの比較を含めるとし、

$$\text{bins} = \text{異なるクラスの組み合わせ数} + \text{同じクラスの組み合わせ数} \quad (4.3)$$

$$\text{bins} = {}_K C_2 + K = \frac{K^2 + K}{2} \quad (4.4)$$

で求めた。またパラメータからヒストグラムの幅 Δ

$$\Delta = \frac{\text{max} - \text{min}}{\text{bins}} \quad (4.5)$$

を計算できる。するとヒストグラムの i 番目の bin の値 $\text{hist}(i)$ は以下のように求まる。

$$\text{hist}(i) = \sum_{n=1}^N \mathbf{I}(\text{CosSim}(\mathbf{x}_n, \mathbf{y}_n) \in [i \times \Delta, (i+1) \times \Delta]) \quad (4.6)$$

ただし \mathbf{I} は範囲内にある値の場合にカウントアップを行う指示関数 (indicator function) である。すなわち範囲 $[i \times \Delta, (i+1) \times \Delta]$ に含まれるコサイン類似度が得られた場合に計

数され、インスタンスを比較した類似度のヒストグラムを図 4.7の通り導く。

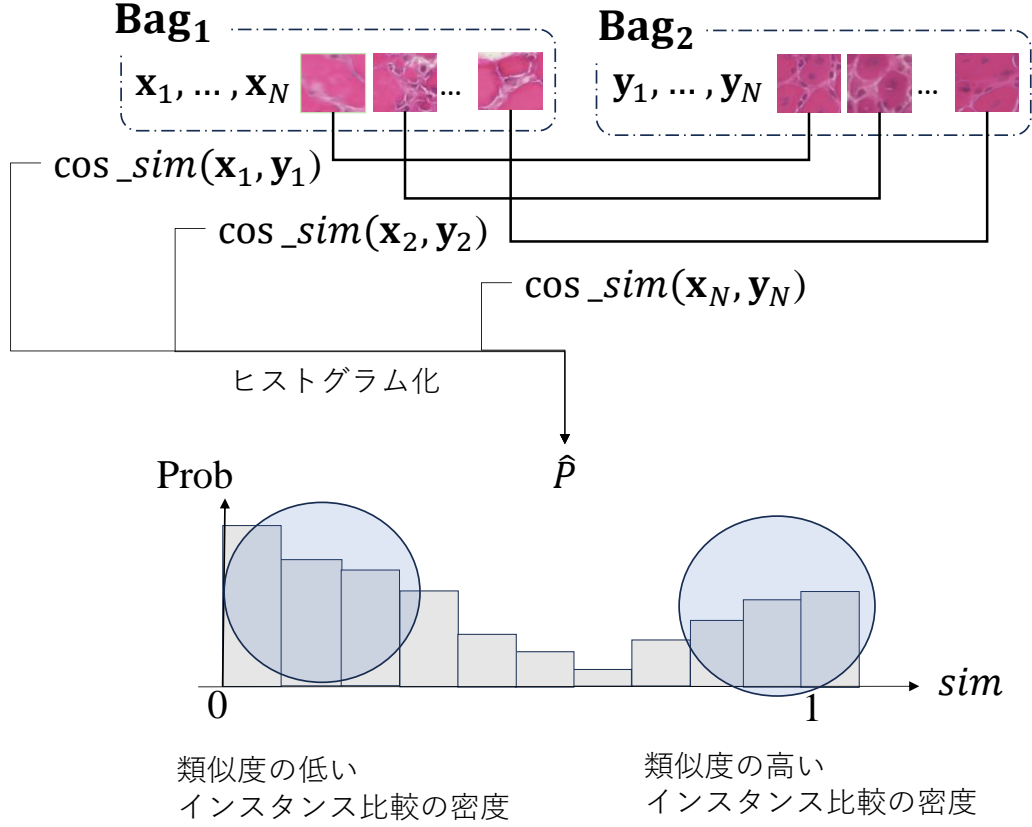


図 4.7 インスタンス類似度の確率密度

しかし指示関数 I を損失関数の実装時に用いた場合、微分不可能な計算プロセスになり Loss からモデルの勾配 (grad) へ誤差逆伝播 (Backpropagation) ができない。本研究では当該問題の解決に inter-atomic distances の離散から連続への変換方法 [98] を参考にし、 $\text{bins} = \frac{K^2+K}{2}, \sigma = 0.1$ のガウス展開法 (Gaussian expansion) 式.4.7によって式.4.6の近似計算をする。

ガウス展開法ではデータ点 i ごとにガウス分布の密度関数の重ね合わせを行い bin の幅で積分することでヒストグラム計算を近似する。ただし μ_i は $i \in \{1, 2, \dots, \text{bins}\}$ 番目のビンの類似度平均値であることに注意する。よってビン番号 i における予測類似度密度 $\hat{P}(i)$ は、

$$\text{hist}(i) \cong \hat{P}(i) = \sum_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\text{CosSim}(\mathbf{x}_n, \mathbf{y}_n) - \mu_i)^2}{2\sigma^2}\right) \Delta \quad (4.7)$$

と計算される。なおガウス展開法と torch が提供するヒストグラム関数を比べると、

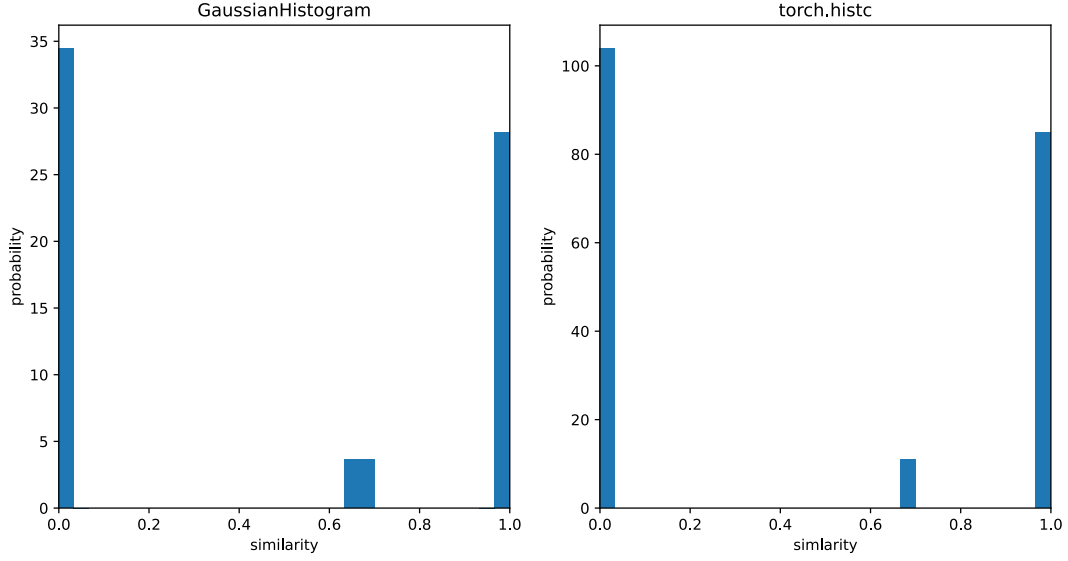


図 4.8 ガウス展開法と torch.hist の比較

図 4.8のようにガウス展開法ヒストグラムと torch ヒストグラムの分布が近く、近似が成功していることが判る。

4.3.4 クラス比率の組み合わせによる真密度分布

2つのバッグが持つクラス比率の組み合わせから P 計算する。クラス間の類似度を損失関数に反映させるために、図 4.5から、赤: 無傷な筋線維 (intact myofiber), 青: ゴーストファイバー (Ghost fiber), 黄: 萌芽細胞 (Myoblast), オレンジ: 筋幹細胞 (Myotube), ピンク: 回復した筋線維 (Recovered Myofiber) の順序を持つクラス $C = \{1, \dots, k, \dots, K\}$ を定義し、クラス間の類似度 $\text{sim}(k, k')$

$$\text{sim}(k, k') = 1 - \frac{|k' - k|}{K - 1} \in [0, 1] \quad (4.8)$$

と定義する。同じクラスを比較した場合 ($k = k'$)、その類似度 $\text{sim}(k, k') = 1$ となる。ただしクラス数 $K > 2$ であることに注意する。2つの異なる Bag に与えられた各クラス比率 (class proportion) $\mathbf{p}, \mathbf{p}' \in [0, 1]^K$ を用いて、真の PDF P は下記のように計算される。

$$\begin{cases} P(\text{sim}(k, k')) = p_k * p'_{k'} & \text{if } k = k'; \\ P(\text{sim}(k, k')) = p_k * p'_{k'} + p_{k'} * p'_k & \text{otherwise.} \end{cases} \quad (4.9)$$

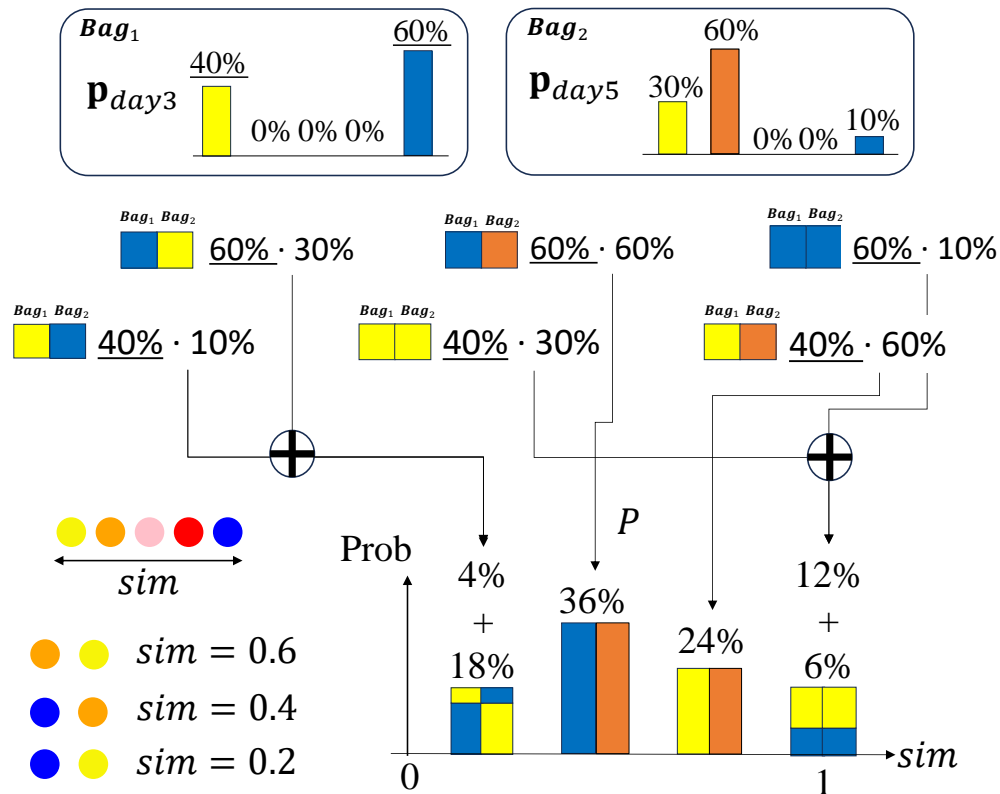


図 4.9 バッグのクラス比率から導出された類似度確率密度

図 4.9 の例では Bag_1 は黄色 40%、青色 60% のクラス比率であり、 Bag_2 は黄色 30%、オレンジ色 60%、青色 10% のクラス比率を持つ。2 つの Bag を比べると同じクラスが比べられる確率は Bag_1 の黄色の割合 40% と Bag_2 の黄色の割合 30% をかけあわせた 12% と Bag_1 の青色 60% と Bag_2 の青色 10% をかけあわせた 6% であるため、類似度 1 の割合は $12\% + 6\% = 18\%$ となる。黄色 (萌芽細胞) とオレンジ (筋幹細胞) は類似度が高く、 Bag_1 の黄色 40% と Bag_2 のオレンジ色 60% を乗算し、24% の密度を得る。以下同様に次の類似度である青とオレンジは 36%、最も類似度が小さい黄色と青色の確率密度は $40\% \cdot 10\% + 60\% \cdot 30\% = 4\% + 18\% = 22\%$ となる。

なおクラス分類器 (head) の学習は 3.2.4 節と等しく活性化関数 ReLU を挟んだ 3 層 MLP (Multilayer perceptron) を用いた。

4.4 再生過程の細胞クラス分類実験

4.4.1 実験手法

順序尺度類似比率学習 (Ordinal Scale Learning from Similarity Proportion: OSLSP) の評価には、表 2.3から CTX 注入の学習用データ、アノテーション済み検証用データの計 31 枚の Whole Slide Image(WSI) を使用した。うち各日付から 1 枚ずつ WSI を選択して計 5 枚の WSI に専門家による 5 分類のアノテーション: 赤: 無傷な筋線維 (intact myofiber), 青: ゴーストファイバー (Ghost fiber), 黄: 萌芽細胞 (Myoblast), オレンジ: 筋幹細胞 (Myotube), ピンク: 回復した筋線維 (Recovered Myofiber) を施しテストデータ (アノテーション済み検証用データ) とした。

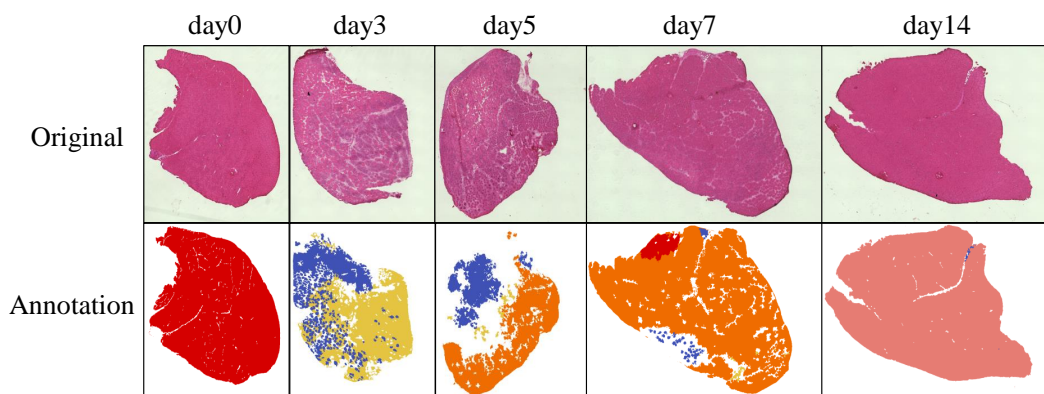


図 4.10 専門家による 5 クラス分類のアノテーション

残りの 26 枚は学習用データとした。データ拡張は 2.5.1 節と同じく回転、鏡像反転、ランダムな光学変換 RandomBrightness ($p=0.5$), RandomContrast ($p=0.5$), RandomGamma ($p=0.5$) を Albumentations v1.3.1 で適用した。これらは WSI から Cellpose に入力する際の切り分けるサイズ $L = 256$ とし、 256×256 ピクセルにカットした後に行った。またデータセット内の比率を正しく反映するために学習データとして 256×256 ピクセルの画像から DINO に入力する際のサイズ $l = 64$ として、ランダムに 64×64 ピクセルの画像をクリップした。

Backbone には ViT-B/8 モデル^{*7} を使用しておりパッチサイズは 8 である。check point は student を採用し、パッチトークンの平均プーリングを有効にした。その他のパ

^{*7} Self-Supervised Vision Transformers with DINO. <https://github.com/facebookresearch/dino>, last accessed on June 29, 2024.

ラメータはデフォルトのままにしている。

OSLSP においてはモデルの最後のブロックのみをファインチューニングした。これはすべての層を更新する場合、DINO が事前学習したパラメータを放棄することに等しく、一般物体で学習したオブジェクトの特徴抽出力を一定程度残すために初期層のパラメータを残す目的がある。DINO のファインチューニングにおいては公式の学習パイプラインに従い事前学習済みの ViT-B/8 モデルからトレーニングを開始した。特徴抽出器を準備した後日付のクラス比を使用した LLP による分類モデル (head) のトレーニングでは ReLU [72] 活性化関数を備えた 3 層のパーセプトロン [18] を使用した。テスト時の推論には Cellpose [52] モデルを用いて Cyto モデルを基にトレーニングデータでファインチューニングしたセグメンテーションモデルに基づき細胞の中心に合わせてテストデータから 64×64 ピクセルの画像をクリップした。実験結果では OSLSP と事前学習された DINO モデル [53] およびトレーニングデータでファインチューニングされた DINO モデル [53] を比較した。

4.4.2 クラス分類実験結果

表 4.2の通り手動の専門家によるアノテーションを真値として順序尺度類似比率学習 (Ordinal Scale Learning from Similarity Proportion: OSLSP) とベースラインを比較した。多クラス分類における定量結果指標では、各クラスのサンプル数の不均衡に影響を受けない指標として micro 平均による accuracy, また過剰検出か否かのバランスを見る指標として macro 平均による precision, recall, F 値, クラス順序を考慮した Root mean square error(RMSE) を用意した。micro 平均では全クラスの True Positive(TP), False Positive(FP), False Negative(FN) を先に計算し, macro 平均では各クラスごとに指標を計算してからクラス間平均を取る。各値の計算方法は以下の通りである。なお指標に関する具体例を踏まえた説明は付録 C に記載する。

$$\text{Micro Accuracy} = \frac{\sum_{k=1}^K \text{TP}_k}{\text{Total Number of Samples}}$$

なお多クラス分類の場合 Total Number of Samples は $\sum_{k=1}^K (\text{TP}_k + \text{FP}_k)$ でもある。

$$\text{Macro Recall} = \frac{1}{K} \sum_{k=1}^K \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k}$$

$$\text{Macro Precision} = \frac{1}{K} \sum_{k=1}^K \frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k}$$

$$\text{Macro F1-score} = 2 \times \frac{\text{Macro Precision} \times \text{Macro Recall}}{\text{Macro Precision} + \text{Macro Recall}}$$

RMSE は図 4.5に示した「青, 赤, ピンク, オレンジ, 黄」のクラス順における誤った予測の重みを反映している. なお表中太文字は最もスコアの高い手法を意味する. 推論クラスを \hat{y} , 真値を y とし, インスタンスの数 N に対して,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

となる. 表 4.2の Accuracy, Recall, Precision, F1-score, (RMSE) 欄の横に付随する矢印 (\uparrow や \downarrow) は, 上矢印 (\uparrow) は当該項目が高ければ, 下矢印 (\downarrow) は低ければ良いモデル性能であるということを意味する.

表 4.2 異なる特徴量抽出器を用いたときのクラス分類器 (head) の推論結果

Method	Accuracy [%] \uparrow	Recall \uparrow	Precision \uparrow	F1-score \uparrow	RMSE \downarrow
DINO Pre-trained [53]	44.442	0.436	0.332	0.377	2.431
DINO Fine-tuned [53]	20.967	0.251	0.191	0.217	1.76
OSLSP (Ours)	46.005	0.492	0.375	0.425	2.152

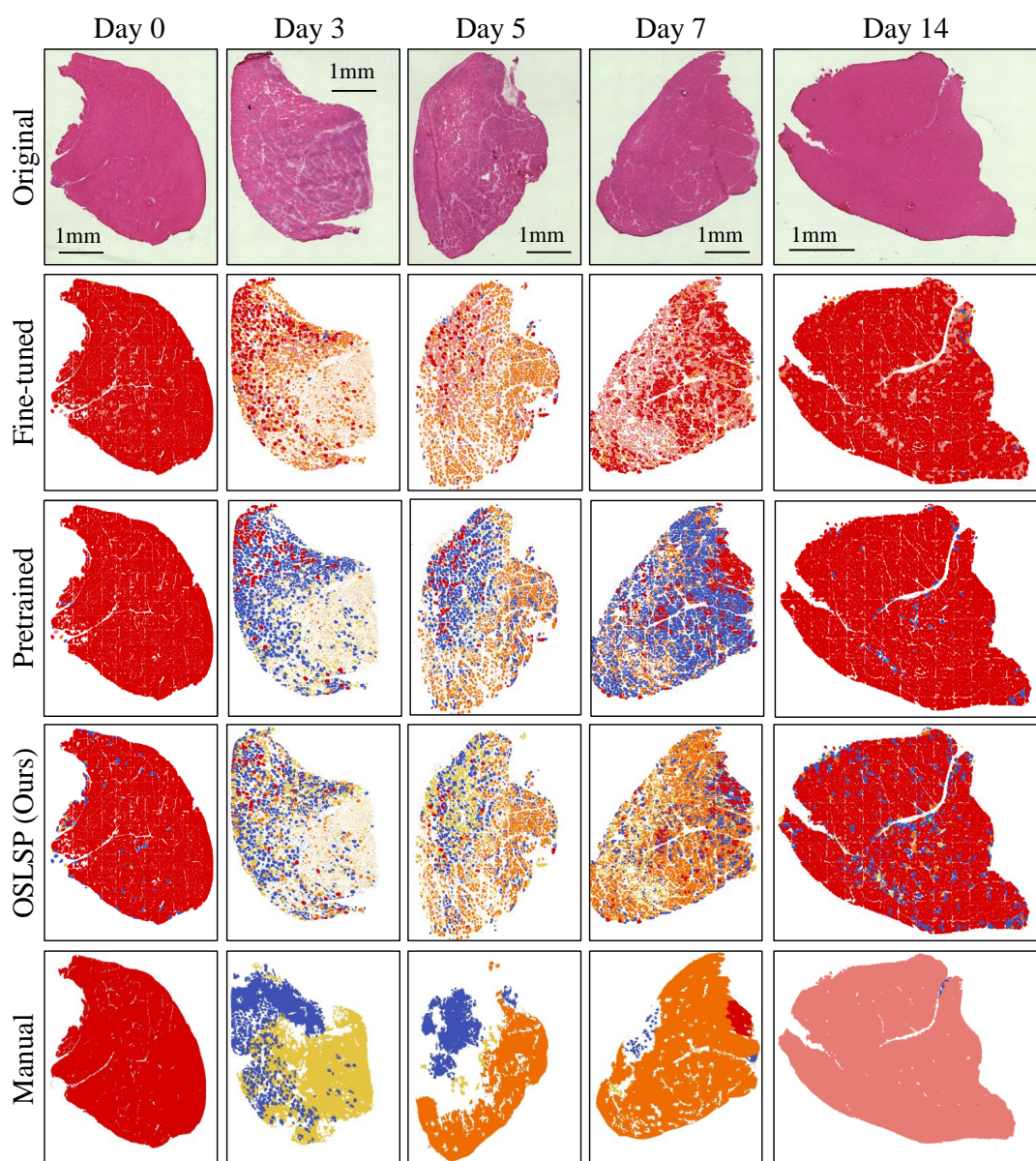


図 4.11 クラス分類による WSI への色塗り結果

図 4.11には OSLSP およびベースラインの視覚的な定性結果を示す．これらは特徴量抽出器 (backbone) が異なる一方で，クラス分類器 (head) はどれも同じアーキテクチャ及び同じ学習 (LLP) をしている．Original は元の HE 染色画像，Fine-tuned は DINO の学習パイプラインと学習用の細胞画像を用いて特徴量抽出器をファインチューニングした．Pretrained は DINO が提供する ImageNet で事前学習したモデルである．OSLSP は提案する類似比率損失 (similarity proportion loss) によって特徴量抽出器をファインチューニングしている．なお図 4.11の Manual の空白部分は専門家でも細胞クラスを判

別できなかった領域を示している．表 4.2の定量的評価を行う際、この領域の細胞は計算から除外している．

表 4.2における DINO Fine-tuned [53] の RMSE が最も良い値を示しているが、図 4.11の結果から判るように赤やオレンジ、ピンクといった類似性が近く、青や黄色といった極端な形態を持つ細胞であるという推論をモデルが行っていない挙動である．これは特徴量抽出器が適切に次元圧縮を行えていない結果である．その他の定量結果については提案手法が他手法を上回っている．

4.5 考察

4.5.1 検証/学習データのドメインギャップ

検証データと学習データの間には顕微鏡、レンズの光学条件、検体の薄さ、凍結処理の精度、染色処理といったドメインギャップの影響を撮影画像は受ける可能性がある．実際検証データと学習データを見比べた図 4.4を見ても、マウスの筋組織画像である点については一致しているものの、画像には目で見ても判る差が存在する．特徴量抽出器は筋組織内の細胞特徴とは別のドメインギャップに紛らわされないように、上手に細胞形態の特徴を抽出する必要がある．

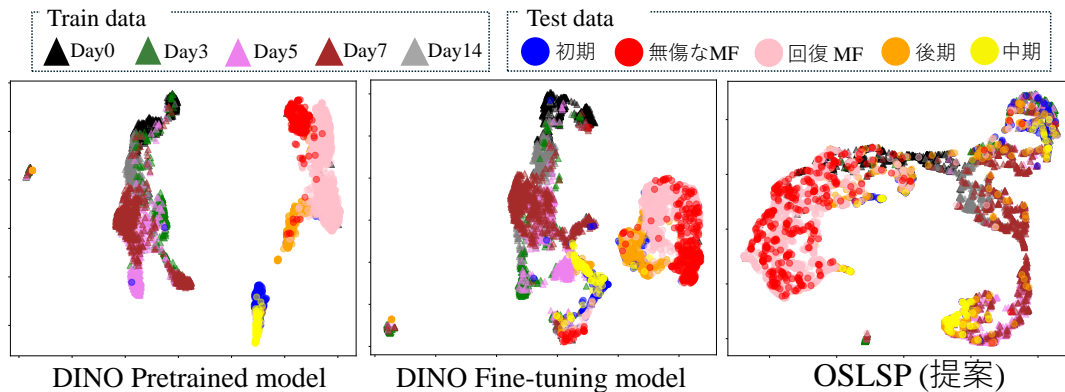


図 4.12 抽出器別の特徴量空間

そこで DINO から出力された 768 次元の特徴量を 2 次元に射影した図 4.12では検証データ (test data) と学習データ (train data) 間の画像のドメインギャップを図示している．図 4.12では△マークは日付ラベルを持つ学習用細胞画像、○マークはクラスアノテーションがある検証用細胞画像である．事前学習モデル (Pretrain model) では学習データの特徴点○と検証データの特徴点△が遠い位置に分布していることが判り、特徴抽出した

時点で既に細胞形態とは無関係に近い特徴量を抽出していることを意味する。特徴点の距離が大きい場合に推論器 (head) はいくつかの特徴量を分類に当たって無視する (重みを 0 に近づける) 必要があり、適した次元圧縮が行えていないことになる。DINO によるファインチューニングには学習データの細胞画像を用いており、実際特徴量空間においても△と○が事前学習モデルよりも近くに分布していることが判る。なおこれらの手法はいずれも教師なし学習を用いた特徴量抽出および 2 次元可視化であるため、教師を使わないクラスタリングによる分類可能性について確認することができる。○のテストデータについて、事前学習モデルでは無傷な筋線維 (赤) と回復した筋線維 (ピンク)、再生後期 (オレンジ) が近いクラスターを形成し、初期 (青色) および中期 (黄色) が近い位置に分布している。仮にクラスタリング手法を採用した場合テストデータは 2 群に分けられ、後から各クラスがどのクラス群 (再生初期・中期または後期・安定期) に属するかを付与することになる。ただしクラスタリングはある一点のデータに対してではなく他のデータとの比較の中で行うため、例えばある日付の WSI のみを推論しようという場合、他のデータ点がないため性能が劣化するおそれがあることに注意しなければならない。提案手法である順序尺度類似比率学習 (Ordinal Scale Learning from Similarity Proportion: OSLSP) は学習データ Day0 の黒△が、検証データの赤○やピンク○といった特徴点と近い位置に分布している。Day0 はすべて無傷な筋線維であり、赤○やピンク○もまた筋線維であることから、細胞の特徴を上手に抽出していることが判る。

また OSLSP によって学習した特徴量抽出器を用いたインスタンスをランダムな順番でコサイン類似度で比較しその分布を確かめた図 4.13を見ると、Day0 の画像同士を比べた場合は高い類似度の多い分布となっている。Day0 と Day3 のインスタンス群を比べた場合は Day0 に比べると類似度の傾向は低くなっている。また同じ日付のインスタンス群を比べた場合は比較的高い類似度分布になる傾向は変わらないが、Day7 は類似度分布がどの日付のインスタンス群とも一樣になる傾向にあり、クラス比率の図 3.8からも判るように様々な細胞クラスが存在していることが反映されている。

これらの結果から図 4.12の DINO Fine-tuning model が示す通り DINO のファインチューニング手法に基づいた教師なし学習によって train-test 間のドメインギャップを抑えていることが判る。また OSLSP によるモデル学習は DINO のファインチューニング以上に細胞の特徴を順序的に学習できていると言える。その結果、表 4.2に示す通り OSLSP は RMSE と精度の両方において改善が見られ、RMSE の結果からも OSLSP がクラスの順序をより正確に学習していることを示しています。一方、ファインチューニン

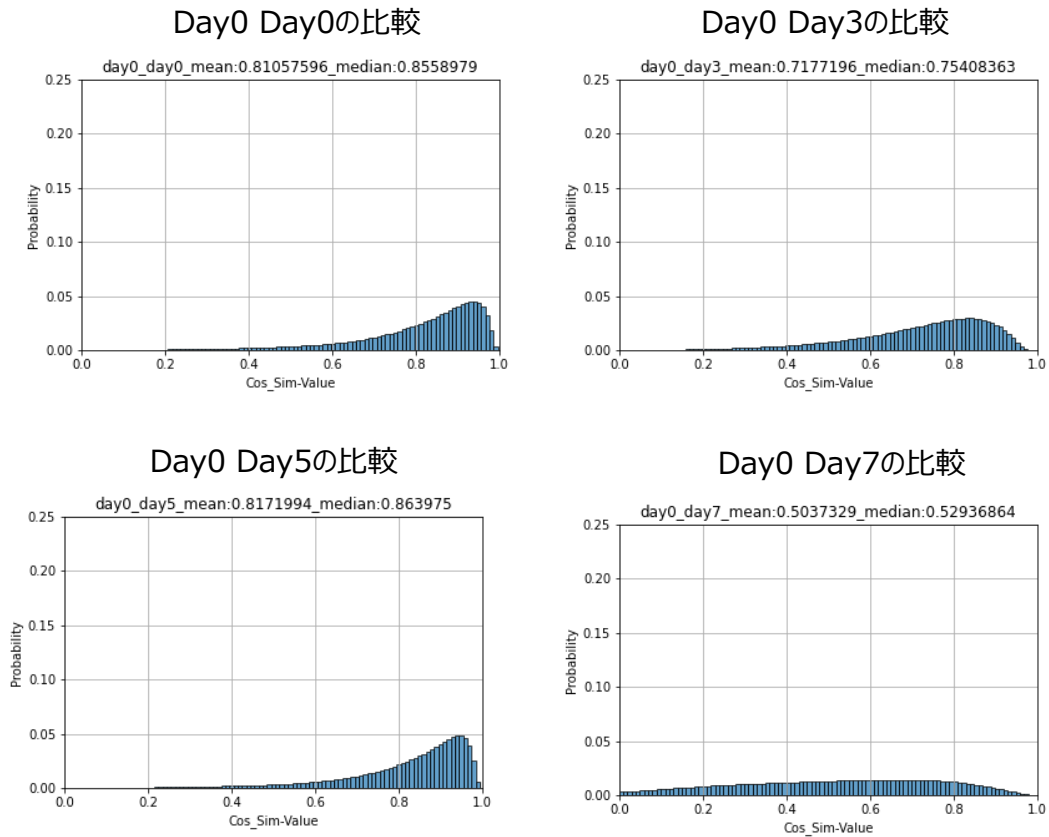


図 4.13 Day0 と他の日付のインスタンス群コサイン類似度分布

グモデルは最も低い RMSE を示しているものの、図 4.11 無傷の MF (筋繊維)、回復した MF、筋管細胞 (myotubes) の推論に収束しており、精度に大きな悪影響を与えている。つまり UMAP 上で学習データ (train data) とテストデータ (test data) 間のインスタンス特徴量が近づくことが必ずしも精度向上に繋がるわけではなく、その結果は表 4.2 にも表れていることに注意したい。ファインチューニングの結果が事前学習モデルよりも悪いということは、大規模データセットで学習した利点を消していることに等しい。すなわち ImageNet などの一般画像に含まれるオブジェクトを認識する力がモデルアーキテクチャの低レイヤで得られているにも関わらず、細胞画像を用いたファインチューニングによって低レイヤの汎用的オブジェクト認識用の次元圧縮が過学習気味になった可能性がある。ただしこれは DINO のファインチューニング手法にも依存する話であるため、他の手法についても調査が必要である。ひるがえって提案手法 OSLSP では順序尺度に注意を払いながらクラス比率を用いた特徴抽出器の更新を行うことが効果的に働いたことを示している。ただ今回の更新ではモデルの最終層の更新に留め、中間層 (middle layer) や注意機構

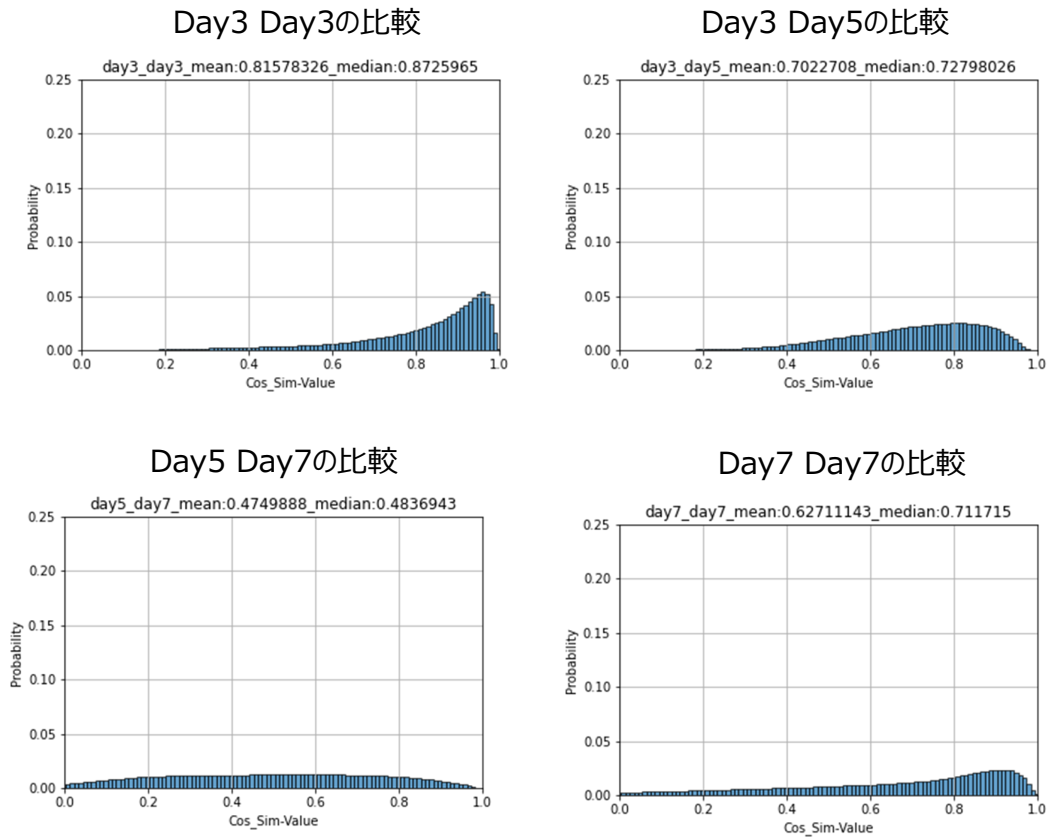


図 4.14 Day3,5,7 と他の日付のインスタンス群コサイン類似度分布

層 (attention layer) の更新は行わなかった。すべての層の重みを学習し直すことは大規模データでの事前学習による利点を消してしまうことになるため避けるべきである。また過学習の判断として用いられる学習時のクラス分類器 (head) のテスト損失 (test loss) と学習損失 (train loss) の推移についても図 4.15の通り観察した。精度が拮抗している事前学習モデルと OSLSP を比較したところ train loss の収束は等しい反面、テスト側では収束傾向が見られなかった。この結果からもクラス分類器 (head) 側でも適切な特徴量がなければ過学習気味になってしまうことが判る。

4.5.2 実用に応じた類似度定義

従来の LLP がクラスを名義尺度として扱うのに対し OSLSP はクラス間の類似性の順序尺度性を考慮した。無傷の筋線維 (MF) とゴーストファイバーは、CTX 注射の直前および直後の時間関係であることから形態が似ている、といった時間順序を用いた類似性の定義を行った。この類似比率損失のアプローチは、学習中の推定エラーに異なる重みを割

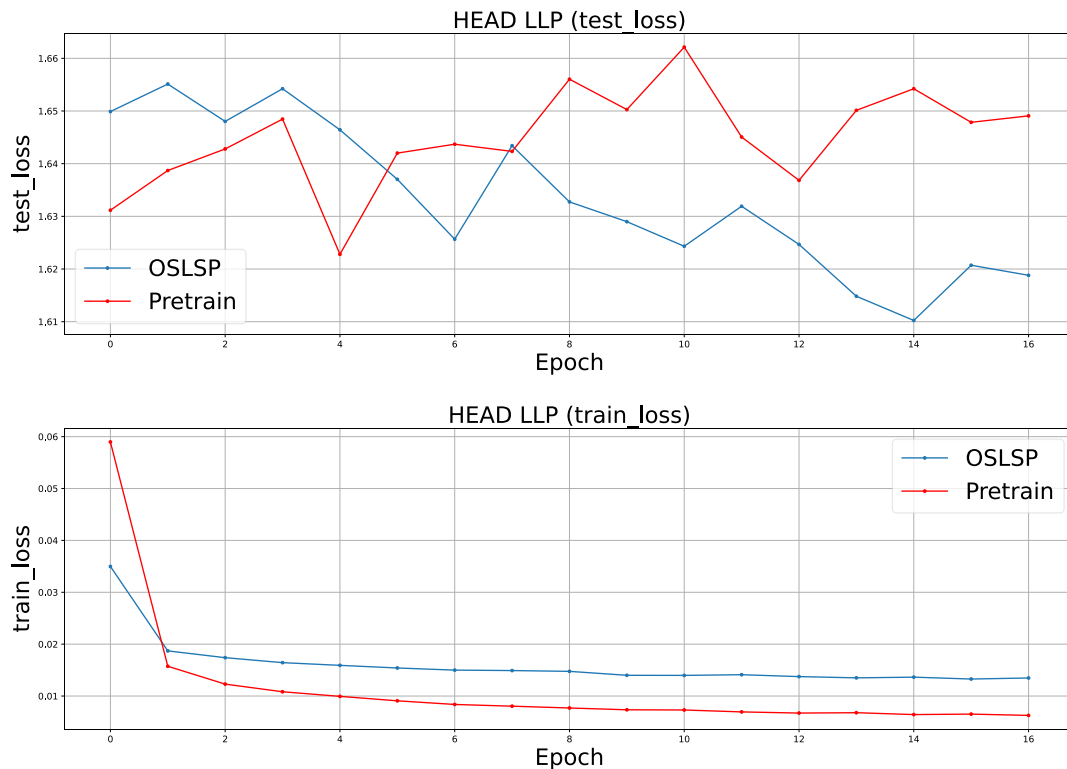


図 4.15 クラス分類器の LLP における test,train 損失

り当てる教師あり学習でのクラス重み付け適用に近い発想である [99]. 反面本手法の制約 (limitation) の 1 つは図 4.11 の Day14 に示されるように、無傷の筋線維 (赤) とゴーストファイバー (青) が事前学習モデルよりも頻繁に混同されることである. というのも赤と青は近いという設定で学習をしているため、この推論誤りについては設定どおりに特徴量抽出器が更新されていることを示す一方で、別種の問題を発生させている. 特に赤と青の混同は筋出力を計測する上では注意が必要である. 無傷の MF と壊死したゴーストファイバーは形態的に似ているものの筋出力には大きな差があるため、赤と青の誤分類は筋力測定において大きな悪影響を及ぼす可能性があるからである. 推論誤りにより大きい重みを損失関数に反映させるために、赤と青の類似度 (クラス間距離) を大きく設定することで、赤と青の取り違えを少なくすることができる. つまり類似度の定義をする段階で形態情報だけで決定するのではなく、例えば本課題であれば筋出力に応じてクラス間距離を設定するなど、使用目的に応じて設定を柔軟に対応させることも検討しなくてはならない.

4.6 結言

本章では画像を比較し類似情報を用いるといった対照学習 (Contrastive Learning) の発想をクラス比率学習 (learning from label proportion) に応用することで、クラスの順序尺度に注意した類似比率による特徴量抽出器のファインチューニングを行うことができた。図 4.12 から、学習データとテストデータ間の撮影条件由来のドメインギャップを埋めることに成功し、再生過程筋組織の細胞を分類するための特徴抽出を可能とした。OSLSP (Ordinal Scale Learning from Similarity Proportion) は、ラベルなしでの自己蒸留 (self-distillation with no labels) によってファインチューニングされた DINO モデルや事前学習モデルに比べて分類精度が高いことが確認された。今後は同じようにクラス比率とクラスの類似性が与えられた課題に OSLSP を適用することで、従来の LLP に比べて精度向上が見込めるかの検証する必要がある。

第 5 章 結論

本論文では WSI 解析のコンピュータ支援に向けた弱教師あり学習の検討を行った。WSI コンピューティングは労力の削減や定量的な解析の点で期待される反面、その専門性から教師データの作成が十分に行えず、慢性的なデータ不足に陥っている。特に巨大なピクセルで構成された Whole Slide Image(WSI) はデータサイズが大きいためアノテーションには踏力を伴い、WSI の各領域ごとに教師データを作成することは困難である。不完全データを用いたモデル学習の一つとして、画像化するさいの事前情報を手掛かりとした弱教師あり学習がある。WSI には弱教師となり得る情報を与えられていることが多く、その事前情報を上手くモデル学習に組み込むことで、低コストな推論器の作成を実現する。

本論文では、筋組織の壊死と再生を誘発する毒物 (CTX) をマウスに注入し、マウスを安楽死・画像化させた際に CTX を注入してからの日数を画像と共に記録した。組織の再生具合は局所的であるため、解析目標は各領域が損傷、再生前期、再生後期、安定期のどの段階あるいはクラスなのかを分類することである。これまでの筋組織画像のコンピュータ解析では再生過程における細胞を形態学的に見分けるための明示的な特徴量が不足しており、再生度合いの評価については対象外であった。本研究では自己教師あり学習由来の特徴量抽出モデルを採用し、暗黙的な多次元特徴量によって再生過程時の細胞を見分けるための表現力を獲得した。

第 3 章では薬学研究への応用に焦点を当て、再生過程解析ソフトウェア開発に取り組んだ。CTX 注入からの日数が 3 日など浅い場合、マウスの筋組織の各領域は損傷から再生初期段階が分布しており、日数が経つにつれて回復が進行するため、5 日目、7 日目となるにつれて損傷領域は減っていき、再生後期の領域が増えていく。この事前情報を活用した荒いアノテーションを WSI に施すことで、日付に応じた各クラスの比率を手に入れる。この比率は例えば 3 日目であれば損傷領域が 4 割、再生前期が 6 割といった風に、各領域にクラスラベルが与えられるのではなく、CTX 注入から 3 日目の WSI は損傷 4 割、再生前期 6 割の比率がある、という比率ラベルである。このように個別のデータ点 (インスタンス) には教師ラベルがないが、インスタンス群にクラス比率ラベルが与えられているときに用いる Learning from Label proportion(LLP) を筋組織の各領域における再生度合いの分類の学習パイプラインに組み込んだ。LLP 手法はクラス比率を確率分布と見なして疑似ラベルを各インスタンスに付与する手法に比べて優れた精度を発揮した。また

LLP によって学習したモデルによって構成した解析ソフトウェア MyoRegenTrack は回復傾向の異なる CTX とグリセロールの注入比較実験において、両薬剤の回復の違いを捉えることに成功した。

第 4 章では情報工学的な観点から LLP の持つ 2 つの課題 (1)backbone の未更新, (2) クラスの名義尺度化, に取り組んだ。画像を直接 head(クラス分類器) に入力すると分類するための境界面の候補が無数となるため, backbone と呼ばれる特徴量抽出器に通すことで縦幅 × 横幅 × チャンネル数 の画像は数百次元の空間へと次元圧縮する。(1) しかし LLP の損失関数は head を通した後で計算されるため backbone の更新が行えない。backbone は一般画像で事前学習されたモデルあり, 目的の筋細胞に適した特徴抽出が達成できている保証はない。実際特徴空間の可視化をすると, 学習データと検証データの間にドメインギャップが観測され, 分類クラスを増やした際の精度低下が深刻であった。(2) また head を通した後で計算される損失関数はクラスを名義尺度として扱っており, 再生過程の細胞が持つ順序的な類似性を損なう。生物学的なクラス順序性に注意を払いながら特徴抽出を行うことができれば, 分類器の学習効率ならびに精度は向上する。(1,2) を同時に解決する順序尺度類似比率学習 (Ordinal Scale Learning from Similarity Proportion Loss; OSLSP) の提案を行った。OSLSP ではインスタンス群同士を比較し合うことで得られる類似度分布と, 群の持つクラス比率を掛け合わせることで得られる類似クラス分布, 2 つの分布を比べることで特徴量空間での順序学習を実現した。OSLSP によって特徴量抽出モデルが筋細胞にドメイン適合したことが確認され, 分類精度についても事前学習済みの backbone を凌駕する結果が得られた。OSLSP は順序性とクラス比率を持つ WSI への応用が期待される。

これまで LLP は癌腫瘍領域の判定 [1, 27], 妊娠率を向上させるための胚選択 [77] などに適用されており, 筋組織の再生評価のための学習パイプラインに導入した知見は第 3 章の貢献である。特に再生のトリガとなる毒物を注入してからの日数のみから学習を行えた低コスト化の実現は, 他の WSI コンピューティングにも活かされることを望む。第 4 章では LLP そのものの改新を達成した。クラス比率学習はアノテーションコストが小さく, また胚選択 [77] のようにプライバシーの問題から個別のラベルを得られないケースでは有力な手段となり得る。無論 LLP は WSI にとどまらず, 合成開口レーダー (Synthetic Aperture Radar; SAR) 画像の分類課題 [100] や選挙結果予測 [101] などに用いられている。また順序尺度は時系列データと符合する点が多く細胞追跡 [102–104] や人歩容解析 [105] などの課題への応用も面白く, アノテーションコストの低い比率学習にお

ける順序尺度を考慮した OSLSP の応用先は幅広い。

ニューラルネットワークは人の神経回路から着想を得たわけだが、逆に弱教師あり学習を人のメカニズム解明に役立てられないだろうか。例えば人を運動感覚系に捉え直した場合、知覚システムはある運動・意思決定に応じて直接的な正解の知覚が用意されているわけではなく、認知機能なども絡めた自己教師学習あるいは弱教師あり学習を行っている可能性がある。実際物体の類似度を考慮した自己教師あり学習による物体認識モデルは脳視覚野メカニズムとの妥当性 (biologically plausible) が確認されているとの報告 [29] もある。また弱教師あり学習である LLP で言えば、例えば同調性バイアス [106] という集団にいて同じ行動 (behavior) をとってしまう心理傾向があるが、これは個別の行動ではなく郡単位で比率の多い側に誘導されている現象とも捉え直せる。このとき意思決定者は群に属する個別の他人の行動を逐一観測するのではなく、群全体としての動きから自らの意思決定を行っている。このようにマクロな比率データはミクロなヒト個別データに比べて観測のコストが小さい。データ取得コストを小さくして推論を行う例ではないだろうか。弱教師あり学習の持つ、ある程度確定的な情報を用いて教師が不完全なデータの予測を行う、というメカニズムに人類が未だ解明できぬ謎が隠れていると信じ、本論文の筆を置くこととする。

付録 A Bag 学習

本節では多数のインスタンス群と群のクラス比率が与えられたバッグを用いた学習方法について補足する。クラス比率とは弱教師 (weakly supervise) の一つである。図 1.1にある値とクラスラベル赤または青を持つインスタンスの分類について考える。教師 (supervise) とは図 1.1の”各データに教師あり”が示すように、データ点それぞれに赤と青の分類すべきクラスがある状態を示す。クラス比率とは図 1.1の右図のようにバッグと呼ばれるインスタンス群を形成し、バッグに含まれるインスタンスのクラス”比率”のみを持つ。バッグに含まれる各インスタンスのクラスは与えられておらず、この比率情報のみを元に青と赤を区別する分類モデルを学習することが目標である。

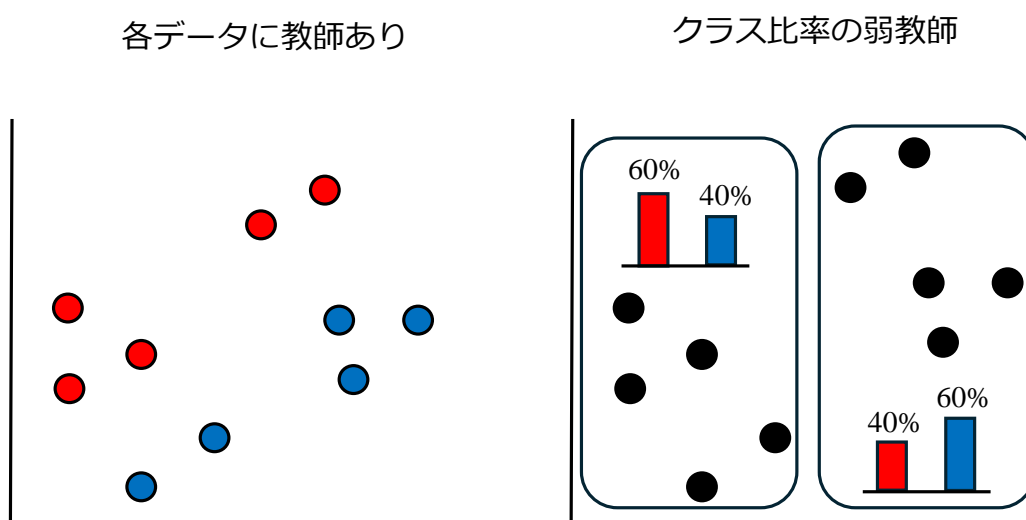


図 1.1 教師ありと弱教師 (クラス比率)

本章ではクラス比率を用いた学習法について紹介する。

A.1 疑似ラベル法

疑似ラベル法 [80] はクラス比率が与えられた際に、比率をクラス確率と見なすことによって、インスタンスに確率に応じた疑似的なクラスを付与することで、疑似的に教師あり学習のパイプラインに載せる手法である。本論文ではすべての学習データのインスタンスには日付ラベル $d = \{day0, day3, day5, day7\}$ と紐づいたクラス比率 $\mathbf{p}_d = [p_1, \dots, p_k, \dots, p_K]$ が与えられている。そこでインスタンスそれぞれの持つ比率をク

ラス確率と見なし、ベクトル内の 1 つの要素が 1、それ以外が 0 のベクトルであるワンホットベクトル (one-hot vector) $\mathbf{y} \in \{0, 1\}^K$ を生成する。このとき one-hot ベクトルの k 番目の要素が 1 になる確率がクラス k の比率に対応する。

$$\mathbf{y} \in \{0, 1\}^K, P(y_k = 1) = p_k \quad (\text{A.1})$$

バッグ単位での疑似ラベル法の損失関数 \mathcal{L}_{pseu} は $\mathbf{x} \in \mathbb{R}^D$ をクラス分類器に入力した結果である $\mathcal{F}(\mathbf{x})$ と one-hot ベクトル $\mathbf{y} \in \{0, 1\}^K$ との交差エントロピーによって計算される。

$$\mathcal{L}_{pseu} = \sum \mathcal{F}(\mathbf{x}) \cdot \log \mathbf{y} \quad (\text{A.2})$$

本手法は疑似的なラベルであることから誤ったラベルを付与する可能性が残り、ノイズとして精度に悪影響が与えられることが指摘されている [90]。のちに説明する LLP をベースとした手法の台頭により疑似ラベル法は最もバニラな手法の一つとしてベースラインの役割を担うことが多い。また特徴量抽出器の更新に疑似ラベル法を用いる IIB-MIL [26] など、一部の利点を活用するケースもある。

A.2 Learning from Label Proportion: LLP

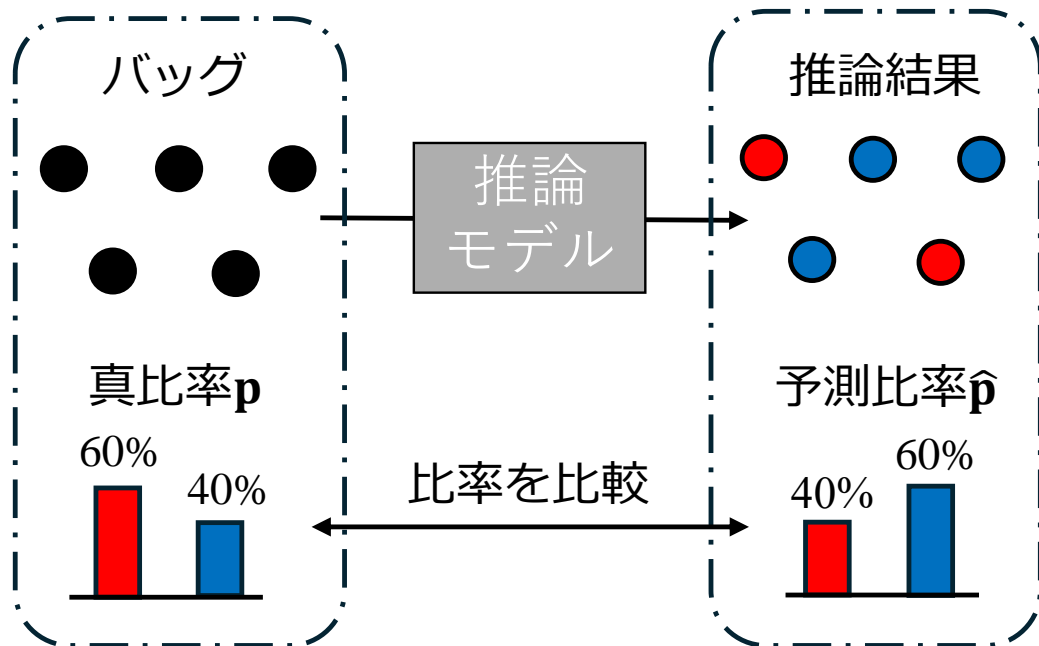


図 1.2 クラス比率学習

個別のインスタンスにクラスラベルを付与するのではなく、インスタンスの予測を集計してバッグ単位の予測クラス比率を計算する手法である。モデルを学習する際、図 1.2 のようにバッグに含まれるインスタンスのクラスを予測した後、予測比率 $\hat{\mathbf{p}}$ を計算し、元々バッグに備わっている真のクラス比率 \mathbf{p} と比較することで損失関数を計算する。

この比較は確立分布を差を計算する手法であれば良いため、L1 ノルム、コサイン類似度、交差エントロピー、KL ダイバージェンス (Kullback-Leibler Divergence)、JS ダイバージェンスなどが候補となる [90]。どの計算法を用いるかの観点の一つとして実際の損失関数の値を見た収束性の観点がある。

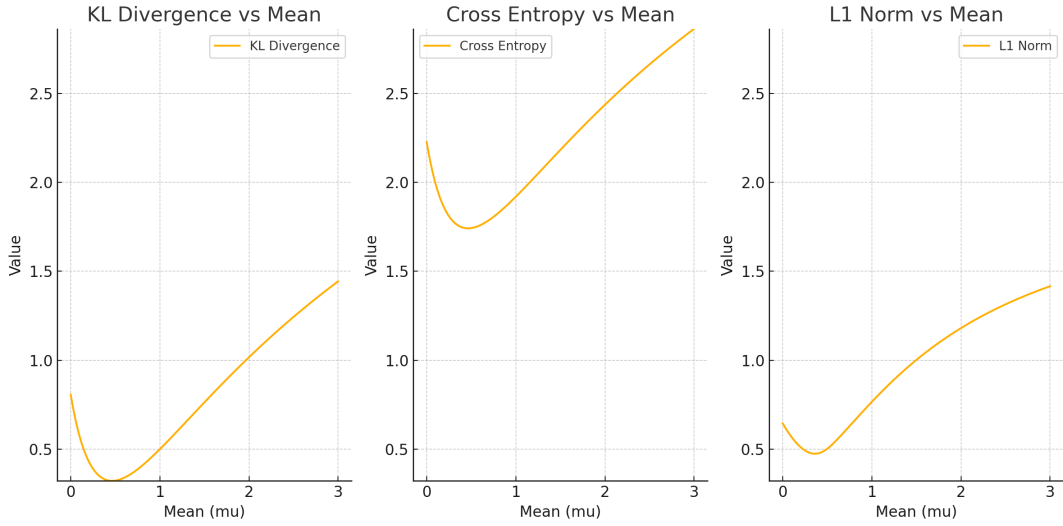


図 1.3 分布差を計算する手法の収束性比較

図 1.3は、平均 0、分散 1 の標準正規分布 $P = N(0,1)$ と、平均 μ を 0 から 3 まで変化した標準正規分布 $Q = N(\mu,1)$ を KL ダイバージェンス、交差エントロピー、L1 ノルムで比較した際の図である。いずれも数学的に非負が保証されている。L1 ノルムは $\mu = 1.5$ 付近で凹凸の変曲点が発生しており、最小値に至る場合の勾配が急になりやすく過学習の危険性が高まる。

$$D_{KL}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (\text{A.3})$$

$$H(P,Q) = - \sum_x P(x) \log Q(x) \quad (\text{A.4})$$

KL ダイバージェンス $D_{KL}(P||Q)$ と交差エントロピー $H(P,Q)$ は勾配については数式からも同じであるが、自己情報量の期待値であるシャノンエントロピー $\sum_x P(x) \log P(x)$

分オフセットが生じる，という見方ができる．本論文の場合，比率の差を限りなく 0 に近づけたい思想から KL ダイバージェンスに決定した．また実験的な確認方法ではあるが，実際に損失関数の推移を確認し，手前のコンピュータ資源や学習の学習率 (learning rate) などのハイパーパラメータも勘案することも大切である．

A.3 IIB-MIL

IIB-MIL [26] は 2023 年に提唱されたインスタンス単位の予測誤差には疑似ラベルロス $\mathcal{L}_{instance}$ を用いて，バッグ単位の予測誤差には従来の LLP ロス \mathcal{L}_{bag} を用いることで，特徴量抽出器 (backbone) とクラス分類器 (head) を同時に学習する手法である．ただし α は重みの割合を調整するためのハイパーパラメータである．

$$\mathcal{L} = \mathcal{L}_{bag} + \alpha * \mathcal{L}_{instance} \quad (\text{A.5})$$

Github 上 (<https://github.com/TencentAILabHealthcare/IIB-MIL>) では限定的なデータに対しての推論コードが公開されており，他データセットでの学習コードがないため，本論文では紹介に留める．

付録 B 多数決学習

本論文では採用しなかったがあるインスタンス群に最も多く含まれるクラス，という弱教師を用いて学習する手法を多数決学習 (majority voting learning) と言われる^{*8} [107]. Majority 学習はバッグ学習よりも更に弱い弱教師であるが，最頻のクラスを指定するだけであるため，クラス比率よりアノテーションコストが小さい.

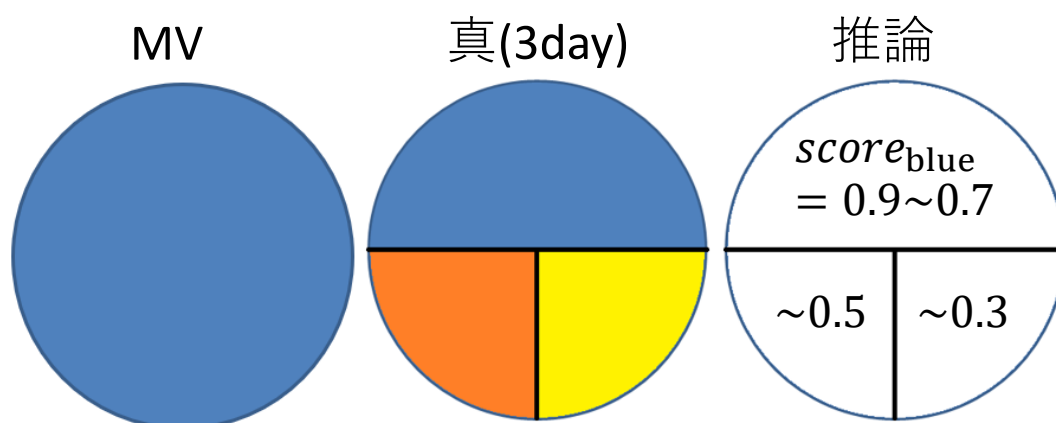


図 2.1 多数決アノテーション

多数決アノテーション (Majority Voting annotation; MV) は図 2.1 のようにあるバッグの真のクラス割合が例えば青 50%，オレンジ 25%，黄色 25% という比率であったとしても，比率は分からず青が最も多いクラスであるということが判っている．この場合学習時はバッグ内に含まれるすべてのインスタンスに青のクラスラベルを施して学習を行う．すると推論時のクラス確信度分布からどの程度青が含まれているのか，ということが推定可能である．このことを利用し，あるバッグに最も含まれているインスタンスが，他のバッグにどの程度含まれているのか，ということを表示するのが多数決学習である．

実際に本研究において，CTX を注入してからの日付ラベルにおいて，Day0 は赤色が最も多く，Day3 は黄色，Day5 は青色，Day7 はオレンジ，という多数決設定で学習を行った結果を示す．多数決の判断に用いたのは専門家によるアノテーション結果である (図 2.13 参照)．なお実際は図 3.8 から判る通り，Day3 に最も多いのは青色，Day5 はオレンジである．なお学習及び推論には MaskRCNN [73] を用いた．図 2.2 の結果では赤は安定状態，青は損傷状態，黄色は再生初期状態，オレンジは再生後期状態であり，真値

^{*8} 研究が盛んでないことから命名に定まりがない

は図 2.13である。図 2.3は日付ごとに推論されたクラスの比率を円グラフにしたものである。

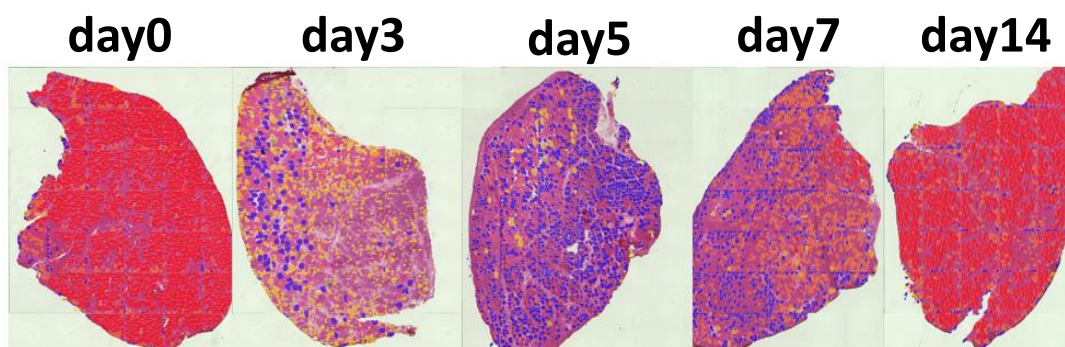


図 2.2 多数決アノテーション結果

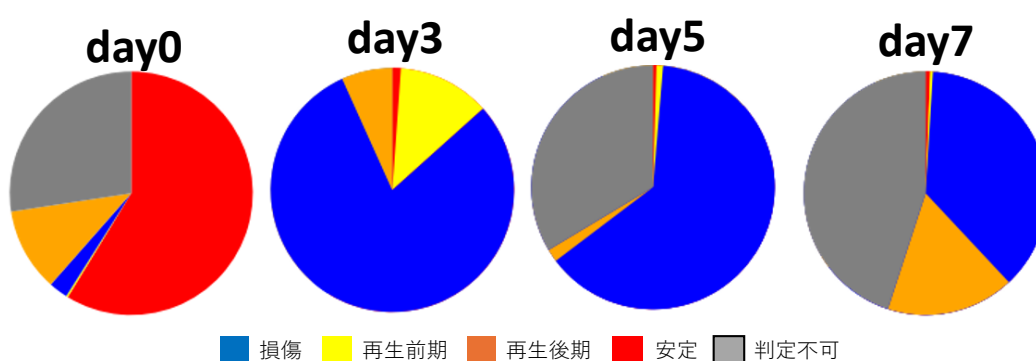


図 2.3 各日付の推論クラス比率

図 2.4は各日付ごとに各クラスの確信度スコアの確率密度分布を示している。例えば Day0 は高い赤の確信度が多く検出されたことを示しており、確信度スコアが小さい分布が多くなるほど、推論モデルの確信度にばらつきが見られ、推論に迷いが生じてことになる。図 2.3では各日付の WSI に含まれた細胞のクラス予測結果の割合であり、3 日目以降で最も多く推論されたクラスは青色であることが示されている。また灰色の割合は本来細胞があるにもかかわらず MaskRCNN で検出できなかった割合である。

0 日目と 14 日目の安定状態は精度良く推論できている点は、真のクラス比率で赤が 100% であることとも一致する反面、3 日目から 7 日目の推論結果には課題が残る。同様に比較的多数決アノテーションと真のクラス比率に近い 7 日目などはオレンジの検出が見られる。3 日目から 7 日目にかけては青のクラスが支配的となり、多数決アノテーションと真クラス比率の差が大きい場合は推論が成功しないことから、WSI に最頻クラスが

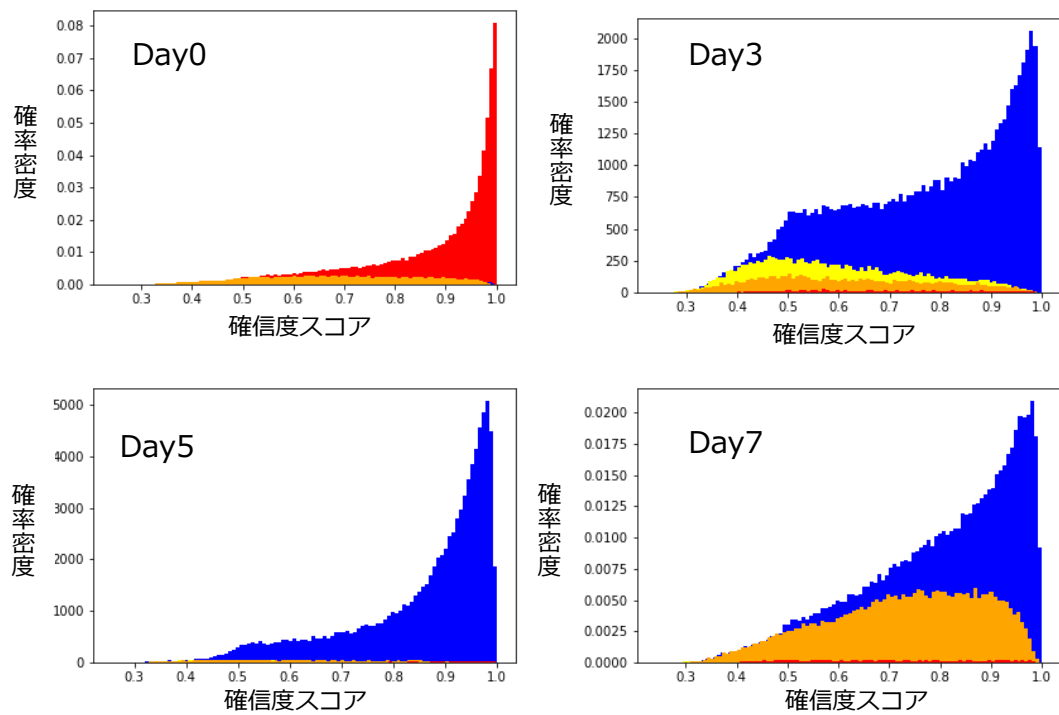


図 2.4 確信度スコアの確率密度分布

少なくとも 70% 近くなければ学習が成功しない傾向が確認された。多数決学習はよりクラス比率の偏りが大きい学習に適する。

付録 C 多クラス分類の評価指標

本章では多クラス単一ラベル分類における評価指標について補足する。micro 平均 (Micro-Averaged Accuracy) は評価指標を全クラス同時に計算するため、全体の正解数を全サンプル数で割ったものである。反対に macro 平均は各クラスごとに指標を計算してから平均を取る。このため指標計算時、各クラスの反映度合いすなわち重みについて、micro はサンプル数の多いクラスほど反映される一方、macro は各クラスの重みが等しくなる。クラスごとのサンプル数に偏りがある場合は、少ないサンプル数のクラスの過度な影響を避ける為に、micro-accuracy を採用することになる。多クラス分類の micro-accuracy 計算手順は、全クラスにおける正解数 (True Positives) の合計をカウントし、全サンプル数でその合計を割る。

$$\text{Micro-Averaged Accuracy} = \frac{\sum_k \text{TP}_k}{\text{Total Number of Samples}}$$

TP_k はクラス k において正しく分類されたサンプル数 (True Positives) であり、分母は全サンプル数の合計である。

例えば 3 クラス (A、B、C) 分類で、以下の混同行列があるとします。この例では C は非常にレアなクラスである。

真値 \ 予測	Pred A	Pred B	Pred C
Actual A	30	5	3
Actual B	4	25	6
Actual C	2	4	1

- 各クラスの True Positives:

$$\text{TP}_A = 30, \quad \text{TP}_B = 25, \quad \text{TP}_C = 1$$

- 全サンプル数の合計:

$$30 + 5 + 3 + 4 + 25 + 6 + 2 + 4 + 1 = 80$$

- ミクロ平均の Accuracy:

$$\text{Micro-Averaged Accuracy} = \frac{30 + 25 + 1}{80} = 0.7$$

これに対して macro accuracy を仮に 2 値分類の定義を流用して下記のように定めた場合,

$$\text{macro accuracy} = \frac{1}{K} \sum_{k=1}^K \frac{\text{TP}_k + \text{TN}_k}{\text{TP}_k + \text{TN}_k + \text{FP}_k + \text{FN}_k}$$

多クラス分類では各クラスの TN を導入する場合, 各クラスを「ポジティブクラス」と見なし残りを「ネガティブクラス」とする 1 対多 (one-vs-all) 手法を用いることになる. 混合行列で TN_A を計算すると TN の定義「予測は A ではなく, 実際も A ではない数」から

$$\text{TN}_A : 25 + 6 + 4 + 1 = 36$$

となってしまう, (予測, 真値) の組のうち (B,C)=4, (C,B)=6 は誤りであるにも関わらず, TN に属してしまうため accuracy の計算をすると,

- A クラスの macro accuracy(仮):

$$\text{accuracy}_A = \frac{\text{TP}_A + \text{TN}_A}{\text{TP}_A + \text{TN}_A + \text{FP}_A + \text{FN}_A} = \frac{30 + 36}{30 + 36 + (4 + 2) + (5 + 3)} = \frac{66}{80}$$

となり不適であることが判る. よって多クラス単一ラベル分類において Accuracy を計算する場合は micro 平均 accuracy を採用することになるまた macro 平均では実際のサンプル数は少ないクラス C が, 精度指標に他のクラスと同じ重みで計算されてしまう問題点がある. これは例えば各クラスの recall を計算した場合,

- A クラスの recall:

$$\text{recall}_A = \frac{30}{30 + 5 + 3} = \frac{30}{38} \quad 0.789$$

- B クラスの recall:

$$\text{recall}_B = \frac{25}{4 + 25 + 6} = \frac{25}{35} \quad 0.714$$

- C クラスの recall:

$$\text{recall}_C = \frac{1}{2 + 4 + 1} = \frac{1}{7} \quad 0.143$$

- macro recall:

$$\text{recall}_{\text{macro}} = \frac{1}{3} \times \left(\frac{30}{38} + \frac{25}{35} + \frac{1}{7} \right) \quad 0.549$$

Micro-Averaged Accuracy が 0.7 であることから少数クラス C の影響を受けていることが判る. 本研究の検証データでは真値が青クラス (損傷) のデータが 3000 程度に対して, オレンジクラス (回復後期) が 1 万以上あるなどクラス不均衡であったため, micro-accuracy の計算も入れた.

謝辞

本論文は著者が 2022 年 4 月に人間情報工学講座からゲノム情報工学講座研究室に移籍してから現在に至るまでに行った研究成果をまとめたものです。他の研究室から来た私を受け入れて下さった当時の松田研究室のスタッフや学生の皆様に感謝申し上げます。

論文投稿，国際会議，博士論文に至るまで全過程のご指導頂いた大阪大学大学院情報科学研究科バイオ情報工学専攻 松田秀雄 教授に感謝申し上げます。研究室を移籍するにあたっての事前の相談から移籍後に至るまで大変ご迷惑をおかけしたと思います。この度博士審査に望めたことそれ自体が僥倖であり，先生のお力あつてのことと思っております。本当にありがとうございました。

骨格筋のテーマを与えてくれ，論文執筆や発表資料の基礎の指導だけでなく，研究者としての生き方について教えを頂戴した大阪大学大学院情報科学研究科バイオ情報工学専攻 瀬尾茂人 准教授に御礼申し上げます。研究者としてのまさに骨格を教えて頂いたと感じています。

また画像の提供のみならず論文執筆に大いなる尽力を頂いた大阪大学薬学研究科 再生適応学分野 深田 宗一郎 教授に感謝を申し上げます。薬学研究について不慣れな私に実験設備の説明から懇切丁寧にご教授頂いたことが，論文にも活かされているのだと思います。さらにご協力いただいた同研究室の岩森歌奈子様にも感謝申し上げます。

大阪大学大学院情報科学研究科バイオ情報工学専攻 繁田浩功 助教からは論文執筆にとどまらず研究における実装について知見を頂きました。とりわけ SSH と Docker について移籍後門外漢であった私に丁寧に教えて頂いたことは，のちの研究成果に結びつくものでありました。

また前研究室で指導教員であった大阪大学大学院情報科学研究科バイオ情報工学専攻 前田太郎 教授から受けた「科学とは再現性の学問である」という言葉は移籍しても常に持ち続けていた心得であります。同研究室の古川正紘 准教授が D1 ポスター時にコメントした「尺度」という言葉が，のちの医療画像解析のトップ会議で世界初となる LLP に順序尺度を設けた OSLSP の発表に繋がったと思います。人に関する知見が逆にニューラルネットワークに活かされるということを体現できるのは，2つの研究室に在籍し，2つの視点を学ぶことができたからなのだと思います。その運命に感謝したいと思います。同研究室の助教である原彰良先生には学生時代から実験系，論文，予算書の相談に乗って頂きました。今後のご活躍を祈念致します。

大阪芸術大学 安藤英由樹 教授には学部一年の頃よりお世話になっており、安藤教授が大阪大学を去られた後も議論を交わす関係となったことは、大阪大学で得たかけがえのない出会いの一つであります。博士論文提出という報告を以て10年にわたる指導への感謝の意とさせていただきます。

研究室を移籍する際に特にご相談をさせていただいた大阪大学大学院情報科学研究科バイオ情報工学専攻 清水浩 教授には移籍後も度々B棟エレベーターの中で会話をさせていただきました。先生に良い報告をできることが研究室移籍の意義を示す点で目標になっていたと思います。本博士論文の審査にあたって頂き誠にありがとうございました。

大阪大学大学院情報科学研究科バイオ情報工学専攻 松田史生 教授にはM1のポスターの頃から話をする機会があり、移籍し研究内容が変われど忌憚なくご意見を頂戴し、研究の質を向上させてくれたものと思います。

大阪大学大学院情報科学研究科バイオ情報工学専攻ゲノム情報工学講座 事務補佐員 小林加代子様 堀安 真紀代様には日々の事務作業で私を支えてくださいました。誠にありがとうございます。研究室移籍に伴い他の学生よりも多くの事務作業を強いてしまったことと存じます。また人間情報工学講座のときにあらゆる事務作業を完全にこなしてくれた水岡真奈美様には、本文が届くこと叶わぬとしてもここで御礼を申し上げないわけにはいきませんゆえ、書かせて頂きます。本当にお世話になりました。

ヒューマンウェアの同期とその関係者各位の皆様にも、感謝申し上げます。同期には心理面から支えられてきました。博士課程という言葉ば未知の航海を共に進めたことに感謝したいです。ヒューマンウェアプログラムからは同期との出会いだけでなく経済的な支援もありました。その支援がなければ博士課程を終えることは難しかったと思います。

最後に CHAN Weng Ian さん。研究室を変えてから数カ月、君との出会いから探究の歯車を動かすことができました。幾たびもの季節を廻り今博士論文の謝辞を書いていると、これまでの記憶が蘇ります。研究室を移籍し、今まで貯めた研究を博論の業績リストから放棄したとき、正直に言ってすぐに研究をしようという気持ちにはなれませんでした。どれほど優れたとされる論文を読んでも、以前の研究ならばこう問題提起する、というある種の傲慢が脳裏にちらつくのです。しかし君といることで研究の本質について考え直す機会を与えられました。何が面白くて、何を知りたくて、僕は実験をしていたのか、思い出すことができたのです。そして top tier の会議に出す文化の系譜を受けたこと。高い難易度の国際会議に挑む精神は一人では醸成できません。君と肩を並べて世界と戦えたことを誇りに思います。ありがとう。

本研究の一部は大阪大学博士課程教育リーディングプログラムの補助，住友化学株式会社高度情報人材育成奨学金，独立行政法人日本学術振興会特別研究員制度の助成を受けて進められました．

参考文献

- [1] S. Matsuo, D. Suehiro, S. Uchida, H. Ito, K. Terada, A. Yoshizawa, and R. Bise. Learning from partial label proportions for whole slide image segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2024.
- [2] KTK. りんご. <https://www.ac-illustr.com/>, 2024.
- [3] StockNova. 手書き風いちご. <https://stock-nova.com/ja/illustr/36521>.
- [4] D. Miyazaki, A. Nakamura, K. Fukushima, K. Yoshida, S. Takeda, and S.-i. Ikeda. Matrix metalloproteinase-2 ablation in dystrophin-deficient mdx muscles reduces angiogenesis resulting in impaired growth of regenerated muscle fibers. *Human Molecular Genetics*, Vol. 20, No. 9, pp. 1787–1799, 2011.
- [5] ChatGPT and DALL·E. AI-generated illustration:[マウスの回復]. Generated using OpenAI’s ChatGPT and DALL·E, 2024.
- [6] A. Mayeuf-Louchart, D. Hardy, Q. Thorel, P. Roux, L. Guéniot, D. Briand, A. Mazeraud, A. Bouglé, S. Shorte, B. Staels, F. Chrétien, H. Duez, and A. Danckaert. MuscleJ: a high-content analysis method to study skeletal muscle with a new Fiji tool. *Skeletal Muscle*, Vol. 8, p. 25, 2018.
- [7] L. W. Babcock, A. D. Hanna, N. H. Agha, and S. L. Hamilton. MyoSight—semi-automated image analysis of skeletal muscle cross sections. *Skeletal Muscle*, Vol. 10, pp. 1–11, 2020.
- [8] M. Reinbigler, J. Cosette, Z. Guesmia, S. Jimenez, C. Fetita, E. Brunet, and D. Stockholm. Artificial intelligence workflow quantifying muscle features on Hematoxylin–Eosin stained sections reveals dystrophic phenotype amelioration upon treatment. *Scientific Reports*, Vol. 12, No. 1, p. 19913, 2022.
- [9] A. Waisman, A. M. Norris, M. Elías Costa, and D. Kopinke. Automatic and unbiased segmentation and quantification of myofibers in skeletal muscle. *Scientific Reports*, Vol. 11, No. 1, p. 11793, 2021.
- [10] L. Encarnacion-Rivera, S. Foltz, H. C. Hartzell, and H. Choo. Myosoft: an automated muscle histology analysis tool using machine learning algorithm utilizing FIJI/ImageJ software. *PLOS ONE*, Vol. 15, No. 3, p. e0229041, 2020.

- [11] G. Sanz, L. M. Martínez-Aranda, P. A. Tesch, R. Fernandez-Gonzalo, and T. R. Lundberg. Muscle2View, a CellProfiler pipeline for detection of the capillary-to-muscle fiber interface and high-content quantification of fiber type-specific histology. *Applied Physiology*, Vol. 127, No. 6, 2019.
- [12] M. Rahmati and A. Rashno. Automated image segmentation method to analyse skeletal muscle cross section in exercise-induced regenerating myofibers. *Scientific Reports*, Vol. 11, No. 1, p. 21327, 2021.
- [13] T. Desgeorges, S. Liot, S. Lyon, J. Bouvière, A. Kemmel, A. Trignol, D. Rousseau, B. Chapuis, J. Gondin, R. Mounier, B. Chazaud, and G. Juban. Open-CSAM, a new tool for semi-automated analysis of myofiber cross-sectional area in regenerating adult skeletal muscle. *Skeletal Muscle*, Vol. 9, pp. 1–12, 2019.
- [14] Y. Yamaoka, W. I. Chan, S. Seno, K. Iwamori, S. Fukada, and H. Matsuda. Quantifying the recovery process of skeletal muscle on hematoxylin and eosin stained images via learning from label proportion. *Scientific Reports*, Vol. 14, No. 27044, 2024.
- [15] N. Hashimoto, D. Fukushima, R. Koga, Y. Takagi, K. Ko, K. Kohno, M. Nakaguro, S. Nakamura, H. Hontani, and I. Takeuchi. Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3852–3861, 2020.
- [16] R. Hooke. *Micrographia: or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses*. The Royal Society, 1665.
- [17] 山本剛. 道場山本の「ちょっと！これっ教えて♡♡」. (株)GramEye, 2024.
- [18] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, Vol. 65, No. 6, p. 386, 1958.
- [19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.
- [20] A. Dosovitskiy. An image is worth 16x16 words: Transformers for image recog-

- nition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [21] A. Gu and T. Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
 - [22] L. Breiman. Random forests. *Machine Learning*, No. 45, pp. 5–32, 2001.
 - [23] H. J. Jin, X. K-Means clustering. *Encyclopedia of Machine Learning*, 2011.
 - [24] 備瀬竜馬. 教師なし・半教師あり・弱教師あり学習の最先端と バイオ医療画像応用. MEDICAL IMAGING TECHNOLOGY, 2021.
 - [25] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, Vol. 89, No. 1, pp. 31–71, 1997.
 - [26] Q. Ren, Y. Zhao, B. He, B. Wu, S. Mai, F. Xu, Y. Huang, Y. He, J. Huang, and J. Yao. IIB-MIL: Integrated instance-level and bag-level multiple instances learning with label disambiguation for pathological image analysis. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 560–569, 2023.
 - [27] G. Dulac-Arnold, N. Zeghidour, M. Cuturi, L. Beyer, and J.-P. Vert. Deep multi-class learning from label proportions. *arXiv preprint arXiv:1905.12909*, 2019.
 - [28] Y. Yamaoka and H. Ando. Simulation of pseudo inner reading voices and evaluation of effect on human processing. In *Communications in Computer and Information Science*, Vol. 1224, 2020.
 - [29] C. Zhuang, S. Yan, A. Nayebi, M. Schrimpf, M. C. Frank, J. J. DiCarlo, and D. L. K. Yamins. Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, Vol. 118, No. 3, p. e2014196118, 2021.
 - [30] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon. A survey on contrastive self-supervised learning. *Technologies*, Vol. 9, No. 1, p. 2, 2020.
 - [31] D. Bosnakovski, Z. Xu, W. Li, S. Thet, O. Cleaver, R. C. Perlingeiro, and M. Kyba. Prospective isolation of skeletal muscle stem cells with a Pax7 reporter. *Stem Cells*, Vol. 26, No. 12, pp. 3194–3204, 2008.
 - [32] A. Uezumi, S.-i. Fukada, N. Yamamoto, S. Takeda, and K. Tsuchida. Mes-

- enchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. *Nature cell biology*, Vol. 12, No. 2, pp. 143–152, 2010.
- [33] M. Yamamoto, N. P. Legendre, A. A. Biswas, A. Lawton, S. Yamamoto, S. Tajbakhsh, G. Kardou, and D. J. Goldhamer. Loss of MyoD and Myf5 in skeletal muscle stem cells results in altered myogenic programming and failed regeneration. *Stem Cell Reports*, Vol. 10, No. 3, pp. 956–969, 2018.
- [34] S. B. Heymsfield, M. C. Gonzalez, J. Lu, G. Jia, and J. Zheng. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. *Proceedings of the Nutrition Society*, Vol. 74, No. 4, pp. 355–366, 2015.
- [35] 宇多野病院関西脳神経筋センター. 筋ジストロフィーのしくみ.
- [36] R. Piccirillo, F. Demontis, N. Perrimon, and A. L. Goldberg. Mechanisms of muscle growth and atrophy in mammals and drosophila. *Developmental Dynamics*, Vol. 243, No. 2, pp. 201–215, 2014.
- [37] H. V. Tanaka, N. C. Y. Ng, Z. Yang Yu, M. M. Casco-Robles, F. Maruo, P. A. Tsonis, and C. Chiba. A developmentally regulated switch from stem cells to dedifferentiation for limb muscle regeneration in newts. *Nature Communications*, Vol. 7, No. 1, pp. 1–8, 2016.
- [38] H. Yin, F. Price, and M. A. Rudnicki. Satellite cells and the muscle stem cell niche. *Physiological Reviews*, Vol. 93, No. 1, pp. 23–67, 2013.
- [39] Y. Tanaka, S. Kita, H. Nishizawa, S. Fukuda, Y. Fujishima, Y. Obata, H. Nagao, S. Masuda, Y. Nakamura, Y. Shimizu, M. Ryohei, N. Tomoaki, F. Tohru, R. Barbara, F. So-Ichiro, M. Norikazu, and S. Ichiro. Adiponectin promotes muscle regeneration through binding to T-cadherin. *Scientific Reports*, Vol. 9, No. 1, p. 16, 2019.
- [40] L. W. Babcock, A. D. Hanna, N. H. Agha, and S. L. Hamilton. Myosight—semi-automated image analysis of skeletal muscle cross sections. *Skeletal Muscle*, Vol. 10, , 2020.
- [41] 齋藤昭彦. 骨格筋の構造. 理学療法科学, Vol. 18, No. 1, pp. 49–53, 2003.
- [42] A. Mauro. Satellite cell of skeletal muscle fibers. *The Journal of Biophysical*

- and *Biochemical Cytology*, Vol. 9, No. 2, p. 493, 1961.
- [43] O. Guardiola, G. Andolfi, M. Tirone, F. Iavarone, S. Brunelli, and G. Minchiotti. Induction of acute skeletal muscle regeneration by cardiotoxin injection. *JoVE (Journal of Visualized Experiments)*, No. 119, p. e54515, 2017.
 - [44] M. T. Webster, U. Manor, J. Lippincott-Schwartz, and C.-M. Fan. Intravital imaging reveals ghost fibers as architectural units guiding myogenic progenitors during regeneration. *Cell Stem Cell*, Vol. 18, No. 2, pp. 243–252, 2016.
 - [45] Y. Tsuchiya, Y. Kitajima, H. Masumoto, and Y. Ono. Damaged myofiber-derived metabolic enzymes act as activators of muscle satellite cells. *Stem Cell Reports*, Vol. 15, No. 4, pp. 926–940, 2020.
 - [46] L. R. Smith and E. R. Barton. SMASH—semi-automatic muscle analysis using segmentation of histology: a matlab application. *Skeletal Muscle*, Vol. 4, pp. 1–16, 2014.
 - [47] Y. S. Lau, L. Xu, Y. Gao, and R. Han. Automated muscle histopathology analysis using CellProfiler. *Skeletal Muscle*, Vol. 8, pp. 1–9, 2018.
 - [48] Y. Wen, K. A. Murach, I. J. Vechetti Jr, C. S. Fry, C. Vickery, C. A. Peterson, J. J. McCarthy, and K. S. Campbell. MyoVision: software for automated high-content analysis of skeletal muscle immunohistochemistry. *Journal of Applied Physiology*, Vol. 124, No. 1, pp. 40–51, 2018.
 - [49] V. Laghi, V. Ricci, F. De Santa, and A. Torcinaro. A user-friendly approach for routine histopathological and morphometric analysis of skeletal muscle using CellProfiler software. *Diagnostics*, Vol. 12, No. 3, 2022.
 - [50] F. Liu, A. Mackey, R. Srikuea, K. Esser, and L. Yang. Automated image segmentation of haematoxylin and eosin stained skeletal muscle cross-sections. *Journal of Microscopy*, Vol. 252, No. 3, pp. 275–285, 2013.
 - [51] L. Kametsky, T. R. Jones, A. Fraser, M.-A. Bray, D. J. Logan, K. L. Madden, V. Ljosa, C. Rueden, K. W. Eliceiri, and A. E. Carpenter. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics*, Vol. 27, No. 8, pp. 1179–1180, 2011.
 - [52] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, Vol. 18, No. 1, pp. 100–

106, 2021.

- [53] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.
- [54] R. L. Davis, H. Weintraub, and A. B. Lassar. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, Vol. 51, No. 6, pp. 987–1000, 1987.
- [55] D. C. Stephens, M. Mungai, A. Crabtree, H. K. Beasley, E. Garza-Lopez, L. Vang, K. Neikirk, Z. Vue, N. Vue, A. G. Marshall, et al. Protocol for isolating mice skeletal muscle myoblasts and myotubes via differential antibody validation. *STAR Protocols*, Vol. 4, No. 4, p. 102591, 2023.
- [56] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. Albumentations: fast and flexible image augmentations. *Information*, Vol. 11, No. 2, p. 125, 2020.
- [57] A. Briguët, I. Courdier-Fruh, M. Foster, T. Meier, and J. P. Magyar. Histological parameters for the quantitative assessment of muscular dystrophy in the mdx-mouse. *Neuromuscular Disorders*, Vol. 14, No. 10, pp. 675–682, 2004.
- [58] V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, pp. 988–999, 1999.
- [59] B. M. Carlson and E. Gutmann. Regeneration in free grafts of normal and denervated muscles in the rat: morphology and histochemistry. *The Anatomical Record*, Vol. 183, No. 1, pp. 47–61, 1975.
- [60] Y. Yoshimoto, M. Ikemoto-Uezumi, K. Hitachi, S.-i. Fukada, and A. Uezumi. Methods for accurate assessment of myofiber maturity during skeletal muscle regeneration. *Frontiers in cell and developmental biology*, Vol. 8, p. 267, 2020.
- [61] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, Vol. 33, pp. 21271–21284, 2020.
- [62] V. Kim, N. Adaloglou, M. Osterland, F. M. Morelli, M. Halawa, T. König,

- D. Gnutt, and P. A. M. Zapata. Self-supervision advances morphological profiling by unlocking powerful image representations. *BioRxiv*, pp. 2023–04, 2023.
- [63] C. McQuin, A. Goodman, V. Chernyshev, L. Kametsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, et al. CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biology*, Vol. 16, No. 7, p. e2005970, 2018.
- [64] M. A. Mahdy, H. Y. Lei, J.-I. Wakamatsu, Y. Z. Hosaka, and T. Nishimura. Comparative study of muscle regeneration following cardiotoxin and glycerol injury. *Annals of Anatomy-Anatomischer Anzeiger*, Vol. 202, pp. 18–27, 2015.
- [65] M. A. Mahdy, K. Warita, and Y. Z. Hosaka. Early ultrastructural events of skeletal muscle damage following cardiotoxin-induced injury and glycerol-induced injury. *Micron*, Vol. 91, pp. 29–40, 2016.
- [66] L. Lukjanenko, S. Brachat, E. Pierrel, E. Lach-Trifilieff, and J. N. Feige. Genomic profiling reveals that transient adipogenic activation is a hallmark of mouse models of skeletal muscle regeneration. *PLOS ONE*, Vol. 8, No. 8, p. e71084, 2013.
- [67] A. M. Norris, K. E. Fierman, J. Campbell, R. Pitale, M. Shahraj, and D. Kopinke. Studying intramuscular fat deposition and muscle regeneration: insights from a comparative analysis of mouse strains, injury models, and sex differences. *Skeletal Muscle*, Vol. 14, No. 1, p. 12, 2024.
- [68] N. Arsic, S. Zacchigna, L. Zentilin, G. Ramirez-Correa, L. Pattarini, A. Salvi, G. Sinagra, and M. Giacca. Vascular endothelial growth factor stimulates skeletal muscle regeneration in vivo. *Molecular Therapy*, Vol. 10, No. 5, pp. 844–854, 2004.
- [69] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pp. 234–241. Springer, 2015.
- [70] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI*

- conference on artificial intelligence*, Vol. 31, 2017.
- [71] A. Kendall, M. Grimes, and R. Cipolla. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pp. 2938–2946, 2015.
 - [72] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML’10, p. 807–814, Madison, WI, USA, 2010. Omnipress.
 - [73] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
 - [74] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun. YOLOX: Exceeding YOLO series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.
 - [75] D.-H. Lee, et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, p. 896. Atlanta, 2013.
 - [76] N. Quadrianto, A. J. Smola, T. S. Caetano, and Q. V. Le. Estimating labels from label proportions. In *Proceedings of the 25th International Conference on Machine Learning*, ICML ’08, p. 776–783, New York, NY, USA, 2008. Association for Computing Machinery.
 - [77] J. Hernández-González, I. Inza, L. Crisol-Ortíz, M. Guembe, M. Iñarra, and J. Lozano. Fitting the data from embryo implantation prediction: Learning from label proportions. *Statistical Methods in Medical Research*, Vol. 27, , 05 2016.
 - [78] Q. Ye, Q. Zhang, Y. Tian, T. Zhou, H. Ge, J. Wu, N. Lu, X. Bai, T. Liang, and J. Li. Method of tumor pathological micronecrosis quantification via deep learning from label fuzzy proportions. *IEEE Journal of Biomedical and Health Informatics*, Vol. 25, No. 9, pp. 3288–3299, 2021.
 - [79] J. Romano, J. D. Kromrey, J. Coraggio, and J. Skowronek. Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen’ sd for evaluating group differences on the NSSE and other surveys. In *annual meeting of the Florida Association of Institutional Research*, 2006.

- [80] Z. Shao, Y. Wang, Y. Chen, H. Bian, S. Liu, H. Wang, and Y. Zhang. LNPL-MIL: Learning from noisy pseudo labels for promoting multiple instance learning in whole slide image. In *International Conference on Computer Vision (ICCV)*, pp. 21438–21438, 2023.
- [81] G. Dulac-Arnold, N. Zeghidour, M. Cuturi, L. Beyer, and J.-P. Vert. Deep multi-class learning from label proportions. *arXiv preprint arXiv:1905.12909*, 2019.
- [82] A. Uezumi, S.-i. Fukada, N. Yamamoto, S. Takeda, and K. Tsuchida. Mesenchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. *Nature cell Biology*, Vol. 12, No. 2, pp. 143–152, 2010.
- [83] J. Massenet, E. Gardner, B. Chazaud, and F. J. Dilworth. Epigenetic regulation of satellite cell fate during skeletal muscle regeneration. *Skeletal Muscle*, Vol. 11, pp. 1–16, 2021.
- [84] B. Wang, Y. Sun, and Q. Tong. LLP-AAE: Learning from label proportions with adversarial autoencoder. *Neurocomputing*, Vol. 537, pp. 282–295, 2023.
- [85] D. Kotzias, M. Denil, N. d. Freitas, and P. Smyth. From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, p. 597–606, New York, NY, USA, 2015. Association for Computing Machinery.
- [86] F. Yu, D. Liu, S. Kumar, J. Tony, and S.-F. Chang. SVM for learning with label proportions. In *International Conference on Machine Learning*, pp. 504–512. PMLR, 2013.
- [87] D. Kotzias, M. Denil, N. d. Freitas, and P. Smyth. From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, p. 597–606, New York, NY, USA, 2015. Association for Computing Machinery.
- [88] E. M. Ardehaly and A. Culotta. Co-training for demographic classification using deep learning from label proportions. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 1017–1024. IEEE, 2017.
- [89] K.-H. Tsai and H.-T. Lin. Learning from label proportions with consistency

- regularization. In *Asian Conference on Machine Learning*, pp. 513–528. PMLR, 2020.
- [90] A. Brahmabhatt, M. Pokala, R. Saket, and A. Raghuveer. LLP-Bench: A large scale tabular benchmark for learning from label proportions. *arXiv preprint arXiv:2310.10096*, 2023.
- [91] D. S. Takanori Asanomi, Shinnosuke Matsuo and R. Bise. MixBag: Bag-level data augmentation for learning from label proportions international conference on computer vision. In *International Conference on Computer Vision (ICCV)*, pp. pp. 16570–16579, 2023.
- [92] S. Kubo, S. Matsuo, D. Suehiro, K. Terada, H. Ito, A. Yoshizawa, and R. Bise. Theoretical proportion label perturbation for learning from label proportions in large bags. *arXiv preprint arXiv:2408.14130*, 2024.
- [93] L. Chan, M. S. Hosseini, C. Rowsell, K. N. Plataniotis, and S. Damaskinos. HistoSegNet: Semantic segmentation of histological tissue type in whole slide images. In *International Conference on Computer Vision (ICCV)*, pp. 10661–10670, 2019.
- [94] 備瀬竜馬. 最新の研究動向 2023 病理画像処理. 信学技報, Vol. 123, No. 411, pp. 50–52, 2023.
- [95] Y. Yu, C. Weng Ian, S. Shigeto, F. So-ichiro, and M. Hideo. Learning from similarity proportion loss for classifying skeletal muscle recovery stages. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2024.
- [96] Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, X. Ji, and z. yongbing. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In *Advances in Neural Information Processing Systems*, 2021.
- [97] N. Quadrianto, A. J. Smola, T. S. Caetano, and Q. V. Le. Estimating labels from label proportions. In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, p. 776–783, New York, NY, USA, 2008. Association for Computing Machinery.
- [98] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, Vol. 8, No. 1, p. 13890, 2017.

- [99] B.-B. Gao, C. Xing, C.-W. Xie, J. Wu, and X. Geng. Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing*, Vol. 26, No. 6, pp. 2825–2838, 2017.
- [100] Y. Ding, Y. Li, and W. Yu. Learning from label proportions for SAR image classification. *EURASIP Journal on Advances in Signal Processing*, Vol. 2017, , 2017.
- [101] T. Sun, D. Sheldon, and B. O’Connor. A probabilistic approach for learning with label proportions applied to the US presidential election. In *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 445–454, 2017.
- [102] 長村徹, 瀬尾茂人, 藤本健二, 繁田浩功, 松田秀雄ほか. 動画像シミュレータを介した強化学習による細胞追跡手法. 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 16, No. 2, pp. 11–22, 2023.
- [103] K. Fujimoto, S. Seno, H. Shigeta, T. Mashita, M. Ishii, and H. Matsuda. Tracking and analysis of fuCCI-labeled cells based on particle filters and time-to-event analysis. *IJBBB*, Vol. 10, pp. 2–94, 2020.
- [104] K. Fujimoto, T. Mizugaki, U. Rajkumar, H. Shigeta, S. Seno, Y. Uchida, M. Ishii, V. Bafna, and H. Matsuda. A CNN-based cell tracking method for multi-slice intravital imaging data. In *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 1–7, 2021.
- [105] X. Li, Y. Makihara, C. Xu, and Y. Yagi. End-to-end model-based gait recognition using synchronized multi-view pose constraint. 2021.
- [106] 山下駿野, 入舟広大, 畑中健志, 和佐泰明, 平田研二, 内田健康. 同調バイアスを有する集団避難行動のモデル化と受動性に基づくナッジ設計. 計測自動制御学会論文集, Vol. 58, No. 3, pp. 120–128, 2022.
- [107] D. Tao, J. Cheng, Z. Yu, K. Yue, and L. Wang. Domain-weighted majority voting for crowdsourcing. *IEEE transactions on neural networks and learning systems*, Vol. 30, No. 1, pp. 163–174, 2018.