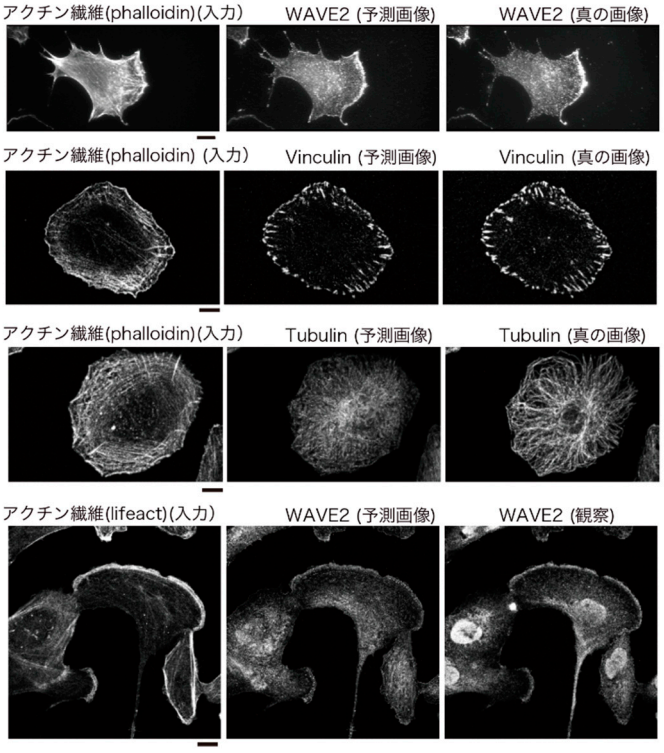
84　深層学習と細胞生物学の融合による細胞運命予測と応用	末次　志郎

【目的】いわゆる人工知能（AI）の一つである深層学習アルゴリズムは、画像間の関係性を判定することや、関連画像を生成することに用いられている。本研究では、深層学習による画像変換を、細胞の分子局在を示す画像に適用し、細胞内の分子局在の相互推定を試みる。さらに、その推定精度から分子間の関係性の推定と分子局在による細胞運命の予測を行う。

【方法】深層学習アルゴリズムによる学習では、細胞の輪郭抽出を行う場合、通常の細胞画像と、教師データとして与えた細胞輪郭の画像（分割の答え）の間の関係の学習から、細胞の輪郭抽出を行う予測器を作成する。この予測器を、新規の画像に適用すると、細胞の輪郭抽出が可能となる。本研究では、関連する二つのタンパク質の局在を示す画像を教師データとして学習させ、二つのタンパク質の局在を示す予測器を作成した。その結果、あるタンパク質の局在から、関連するタンパク質の局在を推定した画像の作成を行うことができることを見出した。

【結果】アクチン繊維と、アクチン繊維の部分構造であるラメリポディアに局在するタンパク質（WAVE2）の画像を学習させると、アクチン繊維の画像から WAVE2 の局在を推定できる予測器を作成することができることがわかった（図）。また、WAVE2 の結合タンパク質である IRSp53 の画像も同様に予測できた。ラメリポディアに存在する細胞接着部位に局在するビンキュリンの予測も可能であった。これに対して、アクチン繊維とは関係の薄いと思われる微小管の曲材は十分に予測できなかった。以上のことから、予測器は分子間の関係性に基づいて、それらの局在を示す画像を生成していることが示唆された。上記の予測は細胞を化学固定したのちに取得した画像を用いてなされたが、生細胞に発現させたアクチン繊維のマーカーである lifeact を用いた画像を用いても、WAVE2 の局在の予測が可能であり、WAVE2 の局在した部位は、ラメリポディアを形成していた。これらの分子局在の予測は、そのさきの細胞運命を示していると考えられ、深層学習により細胞の運命が予測できる可能性が示された。

深層学習によるタンパク質局在画像生成



スケールバー　10 μ m

84. 深層学習と細胞生物学の融合による細胞運命予測と応用

末次　志郎

*奈良先端科学技術大学院大学 バイオサイエンス領域 データ駆動型サイエンス創造センター
分子医学細胞生物学

Key words：深層学習，GAN，WAVE2，アクチン繊維，微小管

緒　言

深層学習などの機械学習はさまざまな分野で大きな成功を収めている［1］。生物学分野においても機械学習は、細胞画像の分類に適用されている。明視野細胞画像を使用して、機械学習によって放射線耐性細胞を親細胞と区別した例や［2］、抗がん剤パクリタキセルで処理された乳がん細胞も、機械学習によって未処理の乳がん細胞と区別できることが報告されている［3］。これらの結果は、機械学習が画像から細胞特性に関連する情報を抽出できることを示している。

機械学習は、上記の分類だけでなく、タンパク質の局在化予測にも適用されている。たとえば、インシリコラベリングとして知られる方法では、明視野細胞画像から特定のマーカータンパク質の推定染色画像を生成して、核、神経細胞、生細胞の位置を推定できる［4］。しかし、ある分子の局在を用いて、別の分子の局在を予測することは報告されていない。

生成的敵対的ネットワーク（GAN）は、機械学習による画像分析の方法の一つであり、多数のペア画像を使用したトレーニングによって得られた確率分布モデルが、仮想のペア画像を生成する［5］。ここで、GAN は、A と B のペアの画像を多数学習した後、B の画像から A の同様の画像を生成できる。たとえば、GAN は、「非笑顔と笑顔のペア画像をたくさん学ぶことで、「笑顔」の顔を予測生成できる［6］。GAN は、ジェネレータ（生成器）とディスクリミネータ（区別器）の 2 つのコンポーネントで構成され、ジェネレータとディスクリミネータの競合によって高品質の画像を生成する。Pix2pix は、画像から画像への変換問題における GAN の主要な実装の一つである［7］。Pix2pix は、これまでに、航空画像からの地図生成、白黒画像からのカラー画像予測、風景画像への属性ラベル、MRI のような生物医学画像における臓器のラベルへなど、さまざまな種類のペア画像の生成に成功している

細胞生物学では、Pix2pix は、細胞膜と核のマーカーの画像を用いて、細胞膜と核の標識に適用された［8］。この場合、ラベルの生成は、膜と核の領域をラベルした画像（ラベル画像）と実際の画像のペアを使用して学習することで達成された。しかし、細胞内小器官の解像度での細胞分子の局在を示す画像の生成、すなわち他のタンパク質の局在を示す画像の生成への pix2pix の適用を実証した報告はこれまでにない。本研究では、pix2pix を使用して、タンパク質の局在を生成、つまり予測できることを示すことを試みた［9］。

タンパク質の局在を示す画像として、細胞骨格関連のタンパク質を用いることにした。細胞は、アクチンフィラメントを含む細胞骨格を変化させることにより、細胞分裂、周囲の環境、およびその他のさまざまな状況に基づいて形状を変化させる。細胞内では、アクチンフィラメントはさらに高次の構成に組み立てられる。これは主に、Cdc42、RhoA、Rac1 などの Rho ファミリーの小さな GTPase により制御される。その中で、Rac1 は WAVE2 を介してアクチンフィラメントの分岐を誘導し、ラメリポディアを形成する。IRSp53 は、WAVE2 を介してこのラメリポディアの形成にも関与している。ビンキュリンは接着斑のタンパク質であり、アクチンフィラメントに接続されており、ラメリポディアの形成に不可欠である。この研究では、細胞のアクチンフィラメントの画像を WAVE2、IRSp53、およびビンキュリンの実際の画像に近い画像に変換できることを示した。これは、GAN による画像変換が機能的に関連するタンパク質の画像を生成できることを示唆している。

*現在の所属：奈良先端科学技術大学院大学 先端科学技術研究科 バイオサイエンス領域 分子医学細胞生物学
奈良先端科学技術大学院大学 データ駆動型サイエンス創造センター
奈良先端科学技術大学院大学 デジタルグリーンイノベーションセンター

方法および結果

1. 条件付き GAN によるアクチンフィラメント画像からの WAVE2 局在の予測

Rac1 の活性化時にラメリポディアを形成するため、Swiss 3T3 細胞を使用した。構成的に活性な Rac1 変異体を Swiss 3T3 細胞に導入し、ラメリポディアを誘導した。化学固定後、細胞をファロイジンと抗 WAVE2 抗体で染色し、それぞれアクチンフィラメントと WAVE2 を可視化した。細胞周辺の扇形のアクチンフィラメント構造は、ラメリポディアであり、WAVE2 が局在していた（図 1)。

アクチンフィラメントと WAVE2 の画像のペアが取得され、pix2pix 条件付き GAN モデルのトレーニングに使用された［10]。変換性能は、アクチンフィラメントと WAVE2 の 772 ペアの画像を使用して、4 分割交差検定によって推定された。各サブセットで、トレーニングセットは 579 枚の画像で構成され、そのうち 15%が検証セットとして使用された。画像を人為的に増やして、学習を強化する画像増強は行わなかった。残りの 193 枚の画像はテストセットに使用された。トレーニングのエポック数に対応する反復回数は 200,000 回とした。反復回数が増えると、生成された WAVE2 画像と実際の WAVE2 画像の類似性が高まった（図 1)。生成された WAVE2 の最終画像は、抗体染色によって得られたものと同様であった。

生成された WAVE2 画像の例を図 2 に示す。WAVE2 は、ラメリポディアへの顕著な局在を示した（図 2)。 pix2pix モデルは、ラメリポディアのサイズに関係なく、WAVE2 の局在を予測した（図 2)。WAVE2 は、ラメリポディアだけでなく、アクチンフィラメントの他の細胞内構造にも局在し、例えば、マイクロスパイクやラメリポディア中に埋め込まれた突起として観察されるフィロポディアの先端は、WAVE2 が局在することが知られている。pix2pix モデルは、ラメリポディアのマイクロスパイク構造での WAVE2 の局在を予測できた。興味深いことに、図 2 の四角で示されているように、ラメリポディアの外側のマイクロスパイクにも WAVE2 が局在すると予測され、実際に WAVE2 が局在した。WAVE2 は細胞間結合でも機能すると報告されている。さらに 2 つの細胞の接触部位に WAVE2 が局在していた（図 2) が、その接着部位における WAVE2 の局在は、予測されることができた。 pix2pix による全体的な予測 WAVE2 局在化は、抗体染色によって検出された実際の WAVE2 局在化と非常に類似していた。総合すると、pix2pix がラメリポディアだけでなく他の細胞構造における WAVE2 の局在を予測できることを示唆した。

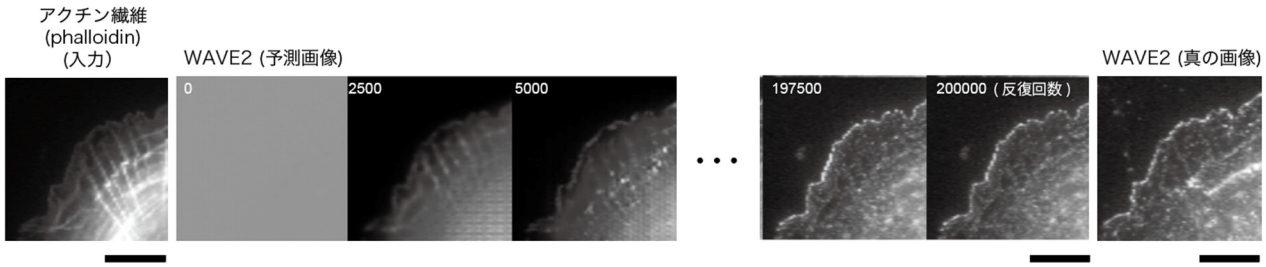


図 1. ファロイジン(phalloidin)によるアクチン繊維画像から WAVE2 画像の生成
ランダムノイズ画像が、反復回数の増加に伴い、WAVE2画像に近づいていく。
（スケールバー：10 μ m）

2. lifeact によるアクチンフィラメント染色と IRSp53 局在への応用

この方法の一般化を調べるために、神経膠腫 U251 細胞と WAVE2 の別の制御因子である IRSp53 を使用して同様の解析を行った。IRSp53 ノックアウト U251 細胞を作製し、GFP-IRSp53 の安定した発現により IRSp53 の発現を回復させた。次に、細胞は、アクチンフィラメントを視覚化するために、mCherry でタグ付けされた lifeact（lifeact-mCherry）を安定発現させた。これらの U251 細胞を血清中で培養し、固定した。 WAVE2 の局在は抗体染色によって同定された。 lifeact によるアクチンフィラメント、GFP による IRSp53、および抗体染色による WAVE2 の画像を、pix2pix を使用した画像変換のための機械学習にかけた。アクチンフィラメント（lifeact）の画像を WAVE2 と IRSp53 に変換し

た。変換パフォーマンスは、4 分割交差検定によって推定された。合計 100 枚の画像が取得され、75 枚がトレーニングに用いられた。75 枚の画像は、90 度のステップでの回転と垂直および水平方向の反転によって、7 倍に拡張され（augmentation:増強され）、525 枚の画像で構成されるトレーニングデータセットが作成され、そのうち 15%が検証セットとして使用された。残りの 25 枚の画像はテストセットに使用された。結果は、lifeact 画像が WAVE2 と IRSp53 の画像を生成できることを示した（図 3）。ただし、核周辺の染色など、一部の特徴は予測されない場合もあった。

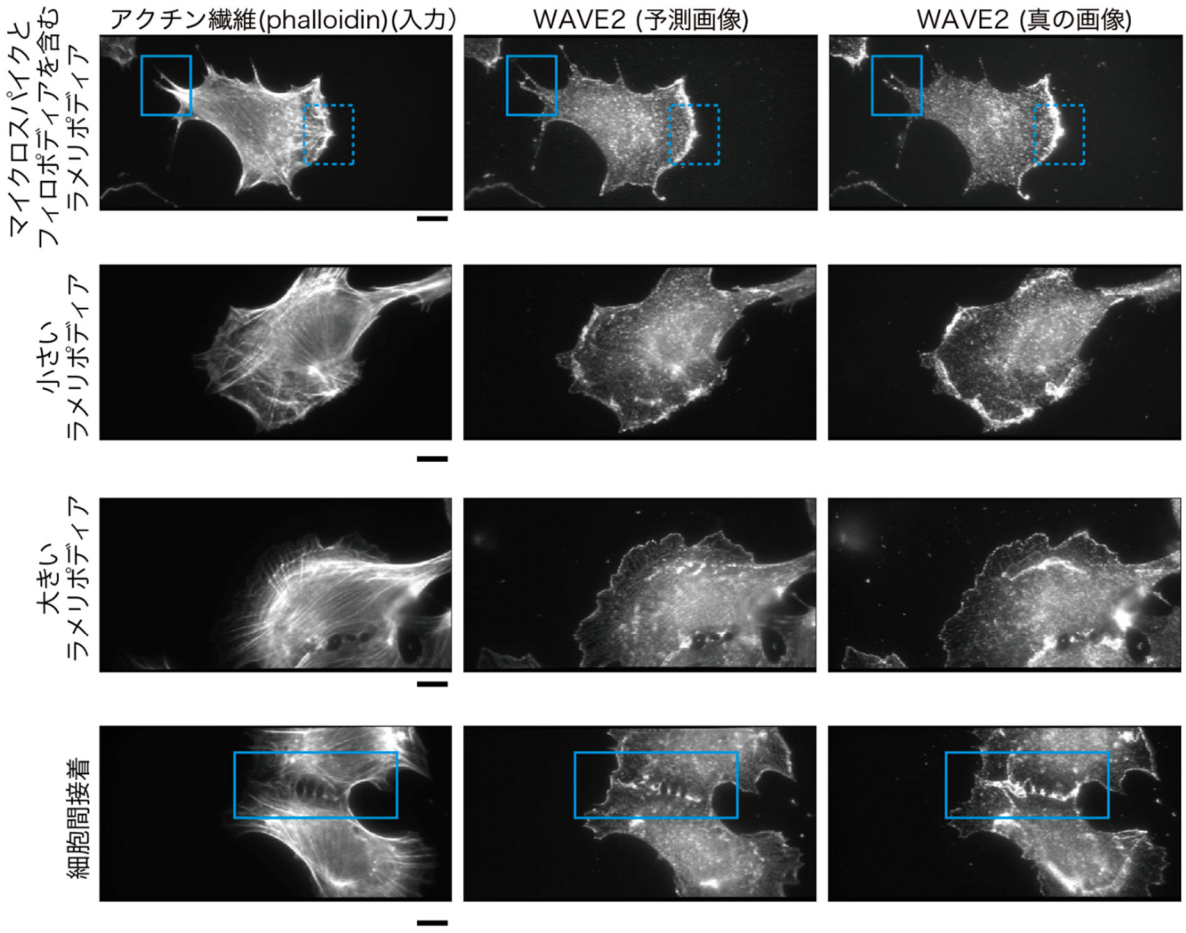


図 2. WAVE2 予測画像の詳細

（1 段目）マイクロスパイク（実線の四角）とフィロポディア（点線の四角）を含むラメリポディア、（2 段目）小さいラメリポディア、（3 段目）大きいラメリポディア、（4 段目）細胞間接着（四角）。固定細胞のファロイジン染色によるアクチン繊維画像を用いた WAVE2 の予測画像と、実際の抗体染色による WAVE2 の局在を示す（スケールバー：10 μm）。

生細胞内の lifeact-mCherry を観察して、ラメリポディアを特定した。次に、細胞を固定し、透過処理し、WAVE2 局在化のために染色した。アクティブなラメリポディア領域は WAVE2 で染色され、これらのラメリポディアのアクチンフィラメント画像は、上記のようにトレーニングされたモデルを使用して WAVE2 画像を生成することができた（図 3）。これらの結果から、pix2pix は、さまざまな条件下で、ラメリポディアでの WAVE2 の局在を具体的に予測できると結論付けた。

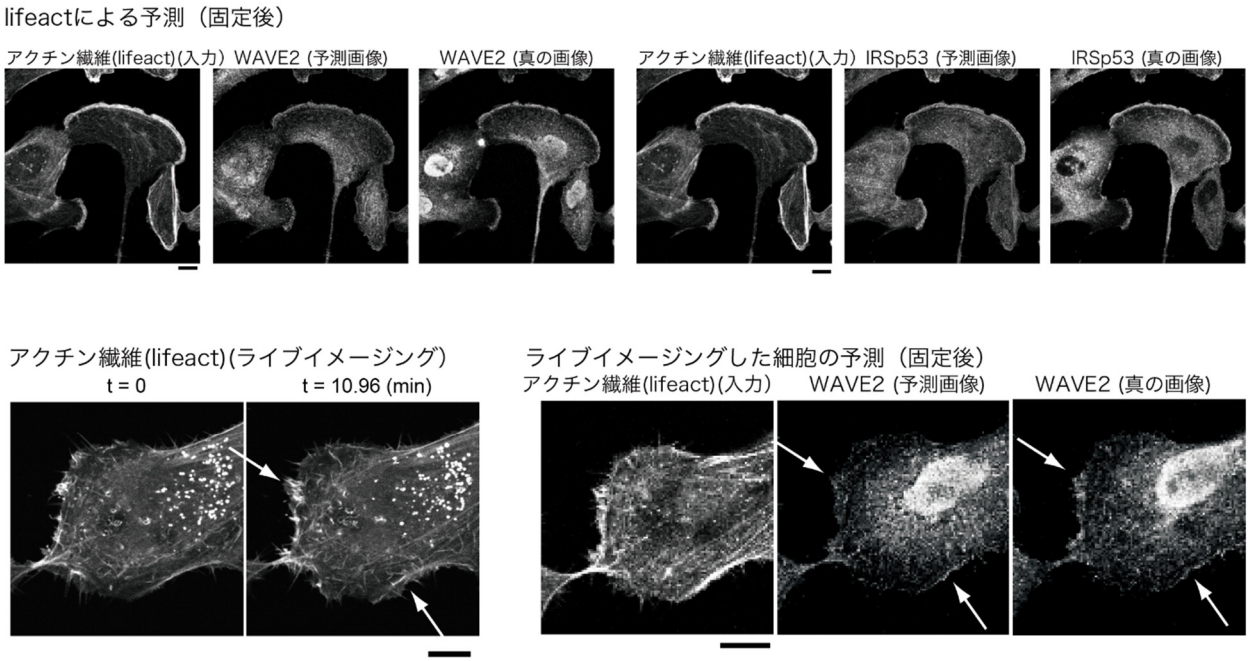


図 3. Lifeact 画像による予測

上段）固定細胞のlifeactによるアクチン繊維画像を用いたWAVE2の予測画像と、実際のWAVE2の局在を示す。

下段）生きた細胞のlifeactによるアクチン繊維画像を示す。

同一の細胞を固定してWAVE2局在を予測し、また、実際の抗体染色によるWAVE2の局在を示す（スケールバー：10μm）。

3. ビンキュリン局在とチューブリン局在への応用

GAN がラメリポディアに関連する他の分子に適用できるかどうかを調べるために、アクチンフィラメントとビンキュリン染色の間でモデルをトレーニングした。100 対のアクチンフィラメントと U251 細胞のビンキュリン画像によって訓練されたモデルの変換性能は、4 分割交差検定によって推定された。 75 枚の画像が 7 倍に拡張され、525 枚の画像で構成されるトレーニングデータセットが作成された。そのうち 15%が検証セットとして使用された。残りの 25 枚の画像はテストセットに使用された。 pix2pix は、アクチンフィラメント画像からビンキュリン画像を生成することに成功した。同様の方法を用いて、アクチンフィラメント画像からチューブリン画像を生成することを試みたが、ビンキュリンや WAVE2、IRSp53 と比べて、成功したとは言い難い。

考　察

本研究では、機械学習手法である pix2pix を介した画像変換を使用して細胞内構造を特定した。Pix2pix は GAN の実装であり、WAVE2、IRSp53、およびビンキュリンは、ラメリポディアの確立された調節因子である。したがって、画像変換によるこれらの局在の予測は、アクチン細胞骨格画像からラメリポディアの部分を同定できることが示唆された。この分野の専門家は、ラメリポディアのアクチンフィラメントと非ラメリポディアを簡単に区別するが、訓練を受けていない人の目には、ラメリポディアがはっきりしない場合があり、顕微鏡画像の判断基準は、異なる研究室の研究者間で異なるが、少なくとも、GAN メソッドは、画質が類似している場合、同等の結果を再現する。

WAVE2 の局在の予測は、ラメリポディアのサイズとは無関係であることがわかった。本研究で用いたアルゴリズムは、4 pixel ごとに計算を行い、その面積は、約 1.9μm^{2}に相当する。したがって、このサイズを超える連続的な特徴を予測することは困難であると予想される。ただし、ラメリポディアでのアクチンフィラメントの特徴はこのウィンドウ内にあると考えられているため、さまざまなサイズのラメリポディアに、予測される WAVE2 の局在化によって注釈を

付けることができる。ビンキュリンと IRSp53 の局在も、4 ピクセルウィンドウのアクチンフィラメントのこのような特徴で予測可能であると考えられた。一方、微小管の局在予測において、予測された微小管局在はフィラメント状ではなかった。これらの非連続な微小管の予測は、カーネルサイズに起因する可能性がある。あるいは、WAVE2、IRSp53、ビンキュリンなどのアクチンフィラメントのレギュレーターとは対照的に、微小管はアクチンフィラメントに直接関連していないため、予測が不正確になることも考えらえる。このため、GAN による局在予測は、特定の細胞構造に関連するタンパク質の関連を決定／予測できる可能性がある。

真の画像と予測画像の違いが相互に独立した局在と機能に起因する場合、GAN は、学習されたペア分子の機能とは独立した機能を発見するためのツールとして使用できる可能性がある。例えば、U251 細胞における WAVE2 の核染色の予測の失敗は、実際に WAVE2 のアクチンフィラメント非依存性の機能を示唆している可能性がある（図 3）。WAVE1 の核局在化が報告されているように、WAVE2 の核染色は、非特異的染色または WAVE2 の核局在化のいずれかを表す可能性がある（Miyamoto and Gurdon、2013）。興味深いことに、このような WAVE2 の核染色は、Swiss 3T3 細胞では観察されなかった。これは、U251 細胞の核 WAVE2 が非特異的染色の結果である可能性も示唆する（図 3）。したがって、予測には、細胞の種類に応じたトレーニングが必要であると結論され、細胞の特定の特徴や細胞ごとに特有の機能を反映している可能性がある。まとめると、生成された画像と実際の画像の間の WAVE2 局在の一致や不一致は、不十分な学習、実験エラーのランダム性、または観察された分子のペアの機能的な独立性に起因する局在の違い、のいずれか、または組み合わせが原因である可能性がある。

タンパク質の局在の予測は大きな可能性を秘めている可能性がある。抗体による標識は通常、いくつかのタンパク質に限定される。対照的に、このような予測による局在予測では、1 回の染色で無制限の数のタンパク質の局在を予測できる。これは、多くの分子間の関係を検出するのに役立つ可能性がある。分子局在の予測は、タンパク質とその機能の局在関連を分析するときに、研究の方向性を探すのに少なくとも役立つ可能性が高い [9]。また、この方法を使用して、生細胞内のラメリポディアにラベルを付けラメリポディア形成の程度を定量化したり、また同様の方法を用いて、ある時間先の画像を予測したりすることも可能であると考えられ、細胞の運命の推定にも役立つと考えることができる。

共同研究者

本研究の共同研究者は、奈良先端科学技術大学院大学情報科学領域の佐藤嘉伸、大竹義人、Soufi Mazen、日朝裕太及び奈良先端科学技術大学院大学バイオサイエンス領域の重根桂、Suphamon Janewanthanakul、西村珠子である。

文　献

1） Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. Nat Methods. 2019;16(12):1233-46. Epub 2019/05/28. doi: 10.1038/s41592-019-0403-1. PubMed PMID: 31133758.

2） Toratani M, Konno M, Asai A, Koseki J, Kawamoto K, Tamari K, et al. A Convolutional Neural Network Uses Microscopic Images to Differentiate between Mouse and Human Cell Lines and Their Radioresistant Clones. Cancer Res. 2018;78(23):6703-7. Epub 2018/09/27. doi: 10.1158/0008-5472.Can-18-0653. PubMed PMID: 30254144.

3） Kobayashi H, Lei C, Wu Y, Mao A, Jiang Y, Guo B, et al. Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning. Sci Rep. 2017;7(1):12454. Epub 2017/10/01. doi: 10.1038/s41598-017-12378-4. PubMed PMID: 28963483; PubMed Central PMCID: PMCPMC5622112.

4） Christiansen EM, Yang SJ, Ando DM, Javaherian A, Skibinski G, Lipnick S, et al. In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. Cell. 2018;173(3):792-803.e19. Epub 2018/04/17. doi: 10.1016/j.cell.2018.03.040. PubMed PMID: 29656897; PubMed Central PMCID: PMCPMC6309178.

5） Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al., editors. Generative adversarial

nets. Advances in neural information processing systems; 2014.
6) Sagawa Y, Hagiwara M. Face Image Generation System using Attributes Information with DCGANs. Transactions of Japan Society of Kansei Engineering. 2018;17(3):337-45.
7) Isola P, Zhu J-Y, Zhou T, Efros AA, editors. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
8) Tsuda H, Hotta K, editors. Cell Image Segmentation by Integrating Pix2pixs for Each Class. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2019.
9) Shigene K, Hiasa Y, Otake Y, Soufi M, Janewanthanakul S, Nishimura T, et al. Translation of cellular protein localization using convolutional networks. Front Cell Dev Biol. 2021(in press).
10) Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N, Sato Y. Automated Muscle Segmentation from Clinical CT using Bayesian U-Net for Personalized Musculoskeletal Modeling. IEEE Transactions on Medical Imaging. 2019:1-. doi: 10.1109/TMI.2019.2940555.