

生成 AI が出てきて思うこと

吉浦紀晃

情報メディア基盤センター長

スマホが出てきてからの IT 技術で感じることは、その目的が人間を考えさせないことではないかということです。例えば、最近のゲームアプリはプレイヤーがゲームステージをクリアできないとゲームステージの難易度を自動で下げたりします。この目的はプレイヤーにゲームを続けさせることです。こんなゲームをプレイしてはステージをどのようにクリアするかと考えることもなくなります。ゲームに限りませんが、『スマホ脳』（アンデシュ・ハンセン著）¹で指摘されているように、スマホのアプリは、利用者をスマホに引きつけるための工夫や仕組みが入っていてとにかくスマホを使わせることが重要視されます。使わせることでアプリ提供者は何らかの収入を得ることができるからです。そしてスマホを使わせつつける仕組みには頭を使わせないという側面があります。

さて、ChatGPT 等の生成 AI については様々な議論が起きています。巷で心配されているように、例えば生成 AI を使いレポート課題に取り組む学生が多くなるでしょう。また、職場で生成 AI を使って文書作成する人たちも出てくるでしょう。AI を使った方が文書の出来が良くなるかもしれません。実際に使ってみると改善の余地がたくさんあると思います。しかし、AI 技術の広がりや頭を使わなくてもいい状況を増やしていきます。

実は、この状況はインターネットや検索エンジンの出現のときと似ています。例えば、レポート課題の回答に、検索エンジンの検索結果の上位の Web ページをそのままコピーするといったことが起きていました。頭を使わなくなる状況は既に生まれていたわけです。しかし、一方でインターネットや検索エンジンは情報や知識を得るのには非常に有益でもあります。実際、『ウェブ進化論』（梅田望夫著）²には「知の高速道路」という言葉が出てきており、検索エンジンを上手に利用すると学びの進み方が格段に速くなります。実際、様々な情報をインターネットから得ることにより昔に比べて学べる機会が多くなったと思います。ただし、これも人次第であって、インターネットを学びに活用する人もいれば、レポート課題の回答をコピペで済ませてしまうだけの使い方しかしない人もいます。インターネットや検索エンジンは人々に大きな知力の差を生み出しているのかもしれません。

生成 AI も頭を使わなくても済む状況を作り出しています。一方で上手に利用して学びの速度を上げていく人も出てくるでしょう。実際、将棋界では藤井聡太という非常に強い棋士がいますが、もしも AI 技術に基づく将棋ソフトが存在しなかったら、ここまで強い棋士が出現しなかったかもしれません。でも、これは藤井聡太氏に失礼かもしれません。それでも、将棋ソフトの利用が強さの一因ではないかと思いますし、AI 技術を利用して自らの能力を高めた一例ではないかと思います。

今後、様々な IT 技術が登場してくると思いますが、頭を使わなくても済む状況を増やすことになると思います。そして、その技術で頭を使わなくなるか学びを早く進めるかで、人々を分けていくことになるかもしれません。そして、それはその技術ではなく、その技術の使い方が分けるのだと思います。

検索エンジンが出現し人間の能力の一部が退化したと思うし、考えない人が増えたように思います。そして、ChatGPT などの AI 技術によるツールが人間の退化を促進するのではないかと危惧しています。その一方で、これらのツールを活用し自らの能力をこれまで以上に高める人たちも出てくるだろうとも思います。結局道具は使い道次第であり、どう使うかが非常に重要なのだと思います。

-
1. アンデシュ・ハンセン『スマホ脳』（久山葉子訳）新潮社、2020 年
 2. 梅田望夫『ウェブ進化論』ちくま新書、2006 年

色知覚に関する階層的な脳内情報表現の研究

栗木一郎，有馬隆人¹，上野健一²

1. 埼玉大学大学院理工学研究科

2. 理化学研究所脳神経科学研究センター

1. 背景

私たちは日々の暮らしの中で、多くの色情報を処理し、それらの情報から物体の材質や危険性などの生きる上で必要な様々な情報を得ている。図1に示した2つのオレンジとレモンについて、その色を問うと普通は橙色（あるいはオレンジ）と黄色のように色名を答える。しかし、片方のオレンジが少し赤みを帯びていることや、光が直接当たる部分と影の部分とで明るさが違って見えることなど、詳細な色の違いについてもその感覚を報告することができる。この2つの色に対する感覚はいずれも日常的にごく普通に感じている物だが、前者（色カテゴリー）に比べ、後者（カラーアピランス）の脳内の神経機構はよくわかっていない。

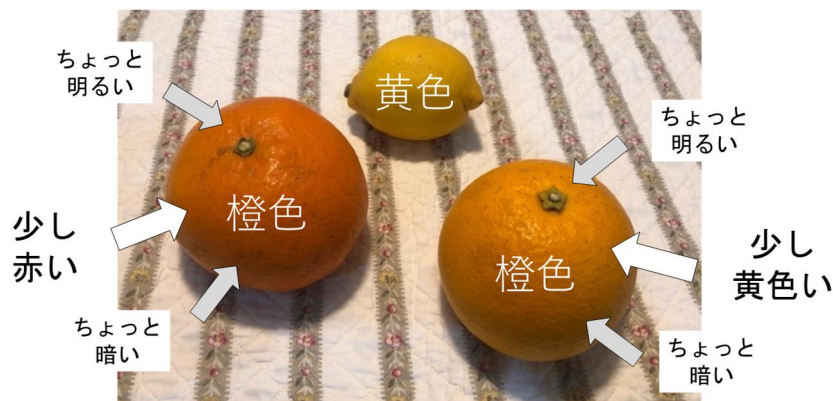


図1. カラーアピランスと色カテゴリーの違い。

カラーアピランスにおける微細に異なる色を全て正確に記憶することは難しく、色の記憶に関する研究では色の扱いがカテゴリー的になっていることが知られている。視覚の情報処理全体の流れは、物体を視覚でとらえ、対象物を認識したのちに記憶する、という段階を経る。すなわち、記憶は視覚情報処理の最後のステージであり、最終的な情報表現は色名のようにカテゴリー的な離散化された表現形式であることがわかる。

Brouwer と Heeger [1] は、視覚刺激に対して色カテゴリー的判断をする課題と色から注意をそらす課題を実施中の実験参加者の脳活動に対して主成分分析を行った結果、後頭葉底部の領野（ventro-occipital area; 以下、VO）と第4次視覚野（V4）がカテゴリー的な応答を示すこと、さらに注意を逸らす課題（注意散逸課題）がVOやV4における色の情報

表現に影響を与えることを報告した。また、Bird ら[2]は両半球の中前頭回（Middle Frontal Gyrus; MFG）がカテゴリカル色知覚に関連する脳領域であると報告した。

他方、細かな色の違いを見分けるカラーアピアランスは色カテゴリーに比べると低次の色情報表現であり、カラーアピアランスは色カテゴリーに比べて連続的な感覚である。このような感覚を調べた先行研究として、DeValois ら[3] は 12 色の実験刺激に対し、色の見えに関する基本色 4 色（赤、緑、青、黄）を 5 回組み合わせて評価する色相スケーリング実験を行った。その結果から、網膜の出力信号である錐体反対色空間の軸と基本色の知覚のピークが一致しないことを明らかにした（図 5(a)）。また Kaneko ら[4] は、初期視覚野の神経活動を反映すると考えられる SSVEP（Steady State Visual Evoked Potentials; 定常視覚誘発脳波）の色相選択性を調べた結果、SSVEP 応答振幅は色相に対して等方的ではなく、ライム-マゼンタ方向に振幅が大きい傾向を示した。計算モデルのフィッティング結果から SSVEP 振幅は錐体反対色信号と知覚的なカラーアピアランスの両方を反映すると報告した。

2. 研究目的

先行研究[1]は、カテゴリカル色知覚の処理に関係する高次視覚野での情報表現が注意散逸課題の影響を受けることを示したが、カラーアピアランスに関する情報表現に対する注意散逸課題の影響についてはまだ研究が報告されていない。そこで本研究では、注意散逸課題がカテゴリカル色知覚とカラーアピアランスに影響を与えるのか、与えるとしたらどの領域でどのような影響を与えるのかを機能的 MRI（functional MRI; 以下 fMRI と略記）という脳機能計測手法を用いて調べた。

3. 実験

3.1 実験手法

脳機能を計測する fMRI 実験では、実験参加者が鏡を通して見ることのできるスクリーンに視覚刺激を投影しつつ、MRI 装置によって脳内での酸素を持つ血液（酸素化ヘモグロビン; oxy Hb）の分布を測定し、刺激を観察している最中に生じる脳活動を計測する。この手法によって得られた脳活動の信号を、BOLD (blood oxygenation level dependent; 血液酸素化レベル依存) 信号と言う。

本研究の実験刺激は、スクリーン中央に色の付いたチェッカーボード図形を呈示し、その中心に 1 秒間隔でランダムに変わるアルファベットを呈示した（図 2）。実験に用いた刺激の色は、錐体反対色空間における等輝度な平面の中で、等エネルギー白色から等距離の円上において、角度（色相）方向に等間隔に選んだ 12 色を選択した。1 回の刺激提示（1 試行）では、その 1 色と背景の灰色により構成されるチェッカーボード図形（4 Hz で反転）を実験刺激として 3 秒間呈示した。同一の視覚刺激に対し、以下に記す 3 つの異なる課題に取り組んでいる時の脳活動を fMRI により計測した。この方法により、条件間の脳活動の違いは課題に関連した脳内の情報処理過程の違いのみを反映することになる。

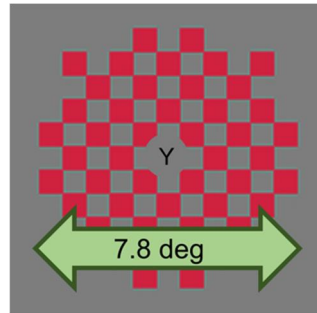


図2．刺激画像の模式図．色の付いたチェッカーボードは 4Hz で反転しながら，3 秒間呈示される．呈示と呈示の間隔は 15 ± 2 秒でランダムに変動させる．中央のアルファベットは 1 秒おきにランダムに交代して呈示される．全ての課題条件において同じ視覚刺激を用いる．

3 つの課題のうち第 1 はカテゴリー課題で，赤/青/緑/黄/紫の 5 色に対応したボタンのうち 1 つを押す方法[1]により色刺激に対応する色名を 1 つ回答した．第 2 はアピアランス課題で，基本色 4 色（赤/緑/青/黄）に対応するボタンを 3 回押す方法[3]により，呈示された実験刺激の色みを多段階で評価する事が求められた．第 3 は注意散逸課題で，1 秒ごとにランダムに切り替わる中央のアルファベットが 2 つ前と同じかを判断するツーバック課題[1]を実施した．この課題は注意負荷が高く，実施中はスクリーンの他の場所すなわち色刺激の場所に注意を向けることができなくなる．初期視覚野は，脳の情報処理の階層の中では比較的低次であり，課題の種類に応じて高次領野からのフィードバック信号を受けることが知られている[1]．注意散逸課題は，色に関する視覚野へのフィードバックを遮断する目的で実施する．

1 人の被験者は各課題に 10 run ずつ参加した．一連の刺激呈示の開始から終了までの 1 run は 216 秒で，3 秒の色刺激呈示と 15 ± 2 秒の背景と文字以外の視覚刺激を呈示しない空白期間から構成される試行を 12 試行（各色 1 試行ずつ）含んでいた．実験参加者の行う課題は 1 run 中は 1 つに固定され，実験参加者は手元に設置されたボタンを用いて各課題に回答した．MRI 装置は理化学研究所脳神経科学研究センターが所有する静磁場強度 3T のスキャナ（Siemens Prisma）を使用し，機能的画像の撮像では TR が 1000 ms，TE が 30 ms，1 voxel のサイズが 3 mm 立方であった．

実験参加者は，視覚・色覚ともに正常な成人 5 名で，いずれも実験参加にあたり同意書を書面で提出した．実験の手続きについてはヘルシンキ宣言を遵守し，埼玉大学および理化学研究所の倫理委員会の承認を得て行われた．

3.2 解析方法

データ解析では，1 試行中で BOLD 信号値が概ねピークに達する視覚刺激の呈示開始から 7 秒後の信号値を用いて，表象非類似度行列[5]（representational dissimilarity matrix; 以下，RDM）を計算した．RDM は，脳活動データに基づく“脳活動 RDM”と，

課題中のボタン応答に基づく“ボタン応答 RDM”の2種類を作成した。脳活動 RDM は図 3(a)～(c)のように行と列が 12 色の各刺激に対応している。全脳を 1 辺が 5 voxel の立方体に分割した小領域（以下、キューブと呼称； $5 \times 5 \times 5 = 125$ voxels）について、RDM の行と列に対応する 2 つの色刺激に対する脳活動（各 125 次元）の相関係数を 1.0 から引いて非類似度を計算し作成した。つまり、脳活動 RDM は脳内の各小領域であるキューブにおける、2 色の刺激を観察中の脳活動の非類似度を表す。また、ボタン応答 RDM は、カテゴリ課題とアピアランス課題それぞれで得た実験参加者のボタン回答に基づき、12 色のうち行と列に対応する 2 色間の非類似度を表す。具体的には、まず 12 色の実験刺激が呈示された時のボタン押しによる回答結果を、カテゴリ課題では 5 次元ベクトル、アピアランス課題では赤-緑成分と青-黄成分の 2 次元ベクトルで表現した。次に、異なる 2 色の実験刺激の非類似度を、各色に対応したベクトルから得られる $\cos\theta$ ，すなわち 2 つのベクトルの内積をベクトルの大きさで正規化した値を利用して $1 - \cos\theta$ と定義した。図 3 に、1 人の実験参加者の視覚野における代表的なキューブの脳活動 RDM と、同じ実験参加者によるボタン応答 RDM を示す。

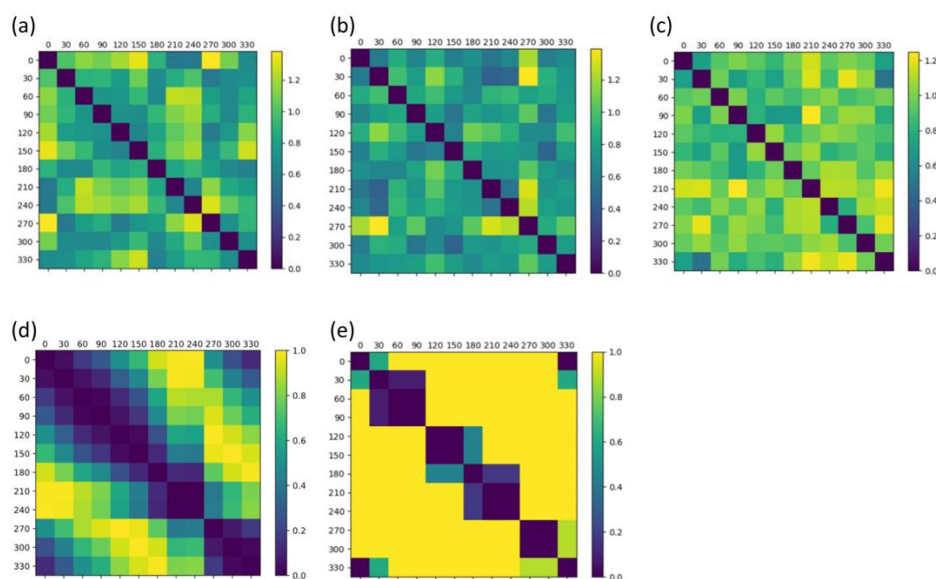


図 3．脳活動 RDM とボタン応答 RDM. 上段 3 つは 1 人の実験参加者の脳活動から求めた RDM ((a) カテゴリ課題, (b) アピアランス課題, (c) 注意散逸課題), 下段の 2 つは脳活動計測実験中のボタン押しによる応答の RDM ((d) アピアランス課題, (e) カテゴリ課題)。

ボタン応答 RDM はカテゴリ課題とアピアランス課題の心理物理的応答を直接反映しているため、それぞれの課題を遂行中の脳活動はカテゴリカル色知覚とカラーアピアランスの情報表現を含んでいると考えられる。従って、ボタン応答 RDM とパターンの類似度が高い脳活動 RDM を持つ脳部位（キューブ）は、それぞれカテゴリカル色知覚とカラーアピアランスに関連の高い情報処理を行っていると予想できる。そこで、脳活動 RDM と

ボタン応答 RDM のパターンの相関係数を求め、相関係数が統計的に有意なキューブを脳画像上にマッピングした。

2つの RDM 間の相関係数が統計的に有意であることを確認する手段として、ランダムパーミュテーション検定 (random permutation test) を行い、統計的に有意と見なせる相関係数の境界値を求めた。具体的には、全脳のデータから時空間的にランダムに、キューブと同数の 125 個の voxel を取り出して作成した架空のデータについて、脳活動 RDM を計算し、実際のボタン応答 RDM との相関係数を求める手順を 10 万回繰り返して null 条件の度数分布を得たのち、上位 5% にあたる相関係数を境界値と定めた (図 4)。以後の解析結果では、全てこの条件を満たしたキューブを用いた結果のみを示す。

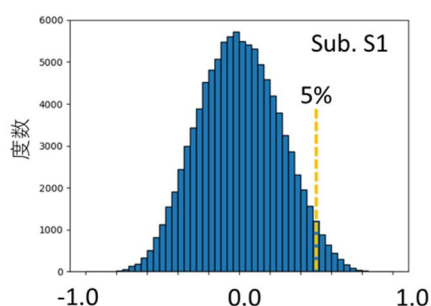


図 4. 相関係数の境界値を求めるための null 分布. 度数分布の上位 5% 点を統計的に有意な相関 ($p < 0.05$) の境界値とする。

4. 結果

4.1 色相スケーリング

アピアランス課題中のボタン応答について、DeValois らの研究[3]と同様の解析を行った結果、図 5 のように先行研究[3]と類似した結果を得た。従って、この課題を実施中の実験参加者の脳内では、カラーアピアランスの評価に関する情報処理が行われていたと考えることができる。

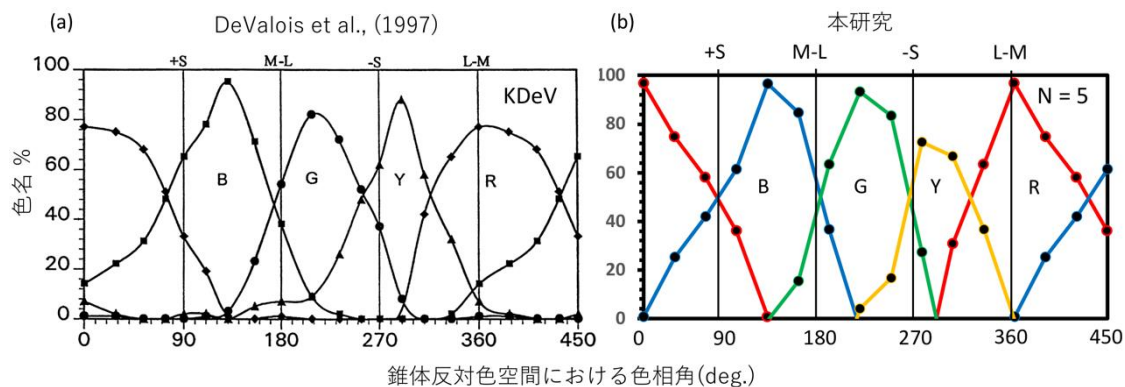


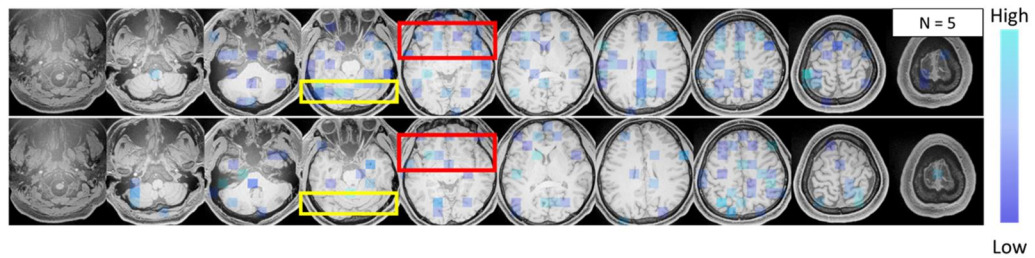
図 5. (a)は先行研究 [3] の結果, (b)が本研究の結果 (参加者 5 名の平均)。横軸は錐体反対色空間における色相角, 縦軸が基本色 (R, G, B, Y) 回答頻度 (%) を示す。90°おきに引かれた縦線は、錐体反対色の

軸方向を示しているが、基本色の応答のピークは一致しない。本研究でも先行研究と同じようにカラーアピアランスの評価ができていることを示す。

4.2 脳活動 RDM とボタン応答 RDM

脳活動 RDM とボタン応答 RDM の相関が高いキューブをマッピングした結果、カテゴリー課題とアピアランス課題でそれぞれ図 6(a), (b) のようになった。

(a) カテゴリー・ボタン応答 vs. (カテゴリー脳活動(上) / 注意散逸課題 脳活動(下))



(b) アピアランス・ボタン応答 vs. (アピアランス脳活動(上) / 注意散逸課題 脳活動(下))

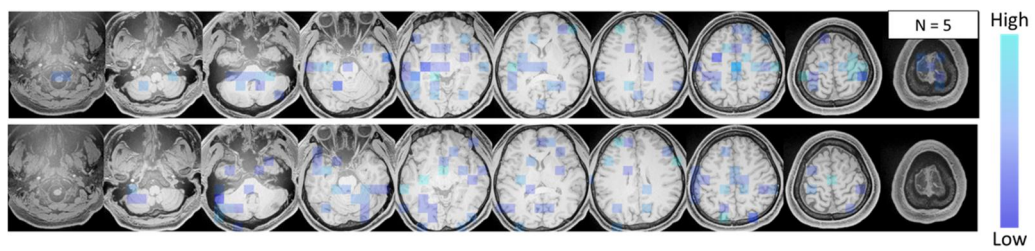


図 6. 知覚課題（カテゴリー課題，アピアランス課題）のボタン応答と脳活動の相関．(a) は色カテゴリー課題のボタン押し RDM との相関のマップ．上段は色カテゴリー課題中の脳活動との相関，下段は注意散逸課題中の脳活動との相関を示す．赤と黄で囲った領域に，上段では相関の高い cube が見られ，下段ではそれが消失している．(b) はカラーアピアランス課題のボタン押し RDM と相関の高い脳部位のマップ．上段はアピアランス課題中の脳活動との相関，下段は注意散逸課題中の脳活動との相関．

色情報処理に関する脳活動が注意散逸課題の影響を受ける場合，先行研究[1]のように情報表現が崩れることが予想される．そこで，カテゴリー課題やアピアランス課題中の脳活動 RDM とボタン応答 RDM との相関係数が高い脳領域において，注意散逸課題中の脳活動 RDM とボタン応答 RDM との相関係数の変化を検証した．具体的には，図 6(a) のように多くの被験者で共通して相関係数が高い下前頭回（Inferior frontal gyrus；以下，IFG）と VO のキューブを解剖学的に定義し，この 2 つの ROI（region of interest；関心領域）に対応するキューブにおいて脳活動 RDM とボタン応答 RDM の間の相関係数の差を t 検定で評価した．その結果，左 IFG ではカテゴリー課題とアピアランス課題の両課題，VO ではカテゴリー課題で相関係数が注意散逸課題中に低下することが分かった（表 1）．

表 1. 相関係数の差分に関する統計的評価

ROI	色課題	t 値 (df=4)	p 値 (両側)
左 IFG	カテゴリー	3.62	0.0224
	アピアランス	2.93	0.0428
右 IFG	カテゴリー	0.760	0.4899
	アピアランス	0.239	0.8232
VO	カテゴリー	4.206	0.0136
	アピアランス	0.403	0.7178

5. 考察

表 1 の結果から、左 IFG と VO における色課題中の脳活動は、注意散逸課題中の脳活動よりもボタン応答と相関が高いことが確認できる。そのため、色カテゴリーやカラーアピアランスに関する脳内情報処理は注意散逸課題の影響を受ける可能性が示された。IFG において注意散逸課題で相関が低くなる原因としては、言語処理の一部を担うブローカ野[7]の存在が考えられる。ブローカ野では言語に関する情報処理が行われているが、今回の実験ではボタン押しによる色課題の応答において、色カテゴリーもしくは基本色名に対応するボタンを押す方法を取ったため、色課題実行中は色名に関する処理が行われていた可能性がある。一方で、注意散逸課題中には、色課題実行時とは異なる処理が行われるため、ボタン応答との相関が低下したと考えられる。

今後の計画として、機能的 ROI の特定と多次元尺度構成法[6] (multi-dimensional scaling; 以後, MDS) を用いた解析を行う予定である。現在の解析では解剖学的に ROI を設定しているが、言語課題によるブローカ野の特定[7]など、機能に基づいて ROI (機能的 ROI) を定義し、ROI ごとの解析を行うことで、分析の精度向上に繋がる可能性がある。高次視覚野の VO 領域において注意散逸課題で相関係数の低下が確認できたことから、初期視覚野でもカテゴリーカル色知覚に関する情報処理が行われている可能性がある。Kaneko らの先行研究[4]で用いられた脳波成分の SSVEP は、初期視覚野 (V1, V2) が信号源であることが知られている。そのことを考慮すると、初期視覚野ではカラーアピアランスに対応した脳活動が認められる可能性がある。そのため、レチノトピーマッピング[8], [9]による視覚野 ROI の特定を行いたい。さらに各 ROI に対して MDS [6] 等を用いた複数 voxel によるパターン解析を行うことにより、先行研究[1]の示した色カテゴリーごとのクラスターや注意散逸課題による変化の有無を検証し、カラーアピアランスに関する情報表現を確認したいと考えている。

謝辞: 本研究は、科学研究費補助金 (課題番号 20H00597 ; 研究代表者・栗木一郎) の補助を受けて行われた。

参考文献

- [1] Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *Journal of Neuroscience*, 33(39), 15454-15465.
- [2] Bird, C. M., Berens, S. C., Horner, A. J., & Franklin, A. (2014). Categorical encoding of color in the brain. Proceedings of the National Academy of Sciences, 111(12), 4590-4595.
- [3] De Valois, R. L., De Valois, K. K., Switkes, E., & Mahon, L. (1997). Hue scaling of isoluminant and cone-specific lights. *Vision Research*, 37(7), 885-897.
- [4] Kaneko, S., Kuriki, I., & Andersen, S. K. (2020). Steady-state visual evoked potentials elicited from early visual cortex reflect both perceptual color space and cone-opponent mechanisms. *Cerebral cortex communications*, 1(1), tgaa059.
- [5] Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 4.
- [6] Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness-of-fit to non-metric hypotheses. *Psychometrika*, 29(1), 115-29.
- [7] Pang, E. W., Wang, F., Malone, M., Kadis, D. S., & Donner, E. J. (2011). Localization of Broca's area using verb generation tasks in the MEG: Validation against fMRI. *Neuroscience Letters*, 490(3), 215-219.
- [8] Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, 56(2), 366-383.
- [9] Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex*, 7(2), 181-192.

樹皮画像解析による樹種判別システムの開発と評価

長谷川 有貴

1. はじめに

近年、画像処理を用いた種類判別システムの研究が盛んに行われており、植物の分野でも植物を写した画像から、植物種を判別するシステムの開発や実用化が進められている。実用化された例の一つである「LeafNet」[1]では、全 21 層の Convolutional Neural Network (以下、CNN) を構築し、大量の植物の葉の画像を学習させることで、高精度な植物種判別を可能としているが、同属間の植物における判別精度の低下が課題となっている。また、葉の画像による植物種判別では、撮影時に背景を単一色にしなければならないなど撮影環境に制約がある[2]。そこで本研究では、撮影環境や撮影時季に制約がなく、樹種によってその特徴が異なる[3]樹皮画像を用いた樹種判別システムの構築を目的として、すでに大量のデータによって学習済みの CNN の一部のみを再学習させることで目的に合わせた新たな学習モデルを形成するファインチューニングにより、樹種判別に適した CNN モデルを構築した。

実験では、埼玉大学大久保キャンパス構内に植樹されている 34 種の樹木を対象として大規模なテストデータセットを作成し、混同行列を用いた判別結果の可視化、並びに Grad-CAM を用いてモデルの判別結果をヒートマップで表し、判別根拠の可視化を行うことで、本研究で構築したシステムの有用性と汎用性を評価した。

2. 開発環境および判別結果の評価手法

2.1 開発環境

開発環境として、Python のディストリビューションである Anaconda と、Jupyter notebook を使い、python 3.8.12 を使用した。使用した PC スペックは、CPU Intel Core i7-10700 (2.90GHz)、RAM 32.0GB、OS Windows 10 で、樹皮画像の撮影には iPhone 11 Pro を使用した。

2.2 Grad-CAM (Gradient-weighted Class Activation Mapping)

Grad-CAM とは、モデルの判別をヒートマップによって可視化してモデルの判別根拠を探る手法である。Grad-CAM では、最後の畳み込み層のモデルの判別に対する勾配を求めることによって、可視化を行っており、式(1)、(2)で表される。

$$a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (1)$$

$$L_{Grad-CAM}^c = ReLU \left(\sum_k a_k^c A^k \right) \quad (2)$$

式(4)では、クラス c の出力 y^c を特徴マップ k の (i,j) ピクセルにおける強度 A_{ij}^k で微分して勾配をとり、全ピクセルについて平均することにより、特徴マップ k のクラス c に対する重み係数 a_k^c を求める。式(5)では、式(4)で求めた重み係数 a_k^c と特徴マップ k の出力 A^k の乗算をすべて加算し、活性化関数 $ReLU = \max\{0, x\}$ による出力をヒートマップ出力 $L_{Grad-CAM}^c$ としている。

2.3 混同行列

本研究では、モデルの判別結果を可視化するために混同行列を用いた。表 1 に、2 クラスを分類した時の混同行列を示す。混同行列から、モデルの正答率 (Accuracy)、精度 (Precision)、再現率 (Recall) を算出でき、それぞれ式(3)-(5)で示される。

表 1 2 クラス分類の混同行列

		予測 (Predicted)	
		Positive	Negative
正解 (True)	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

3. 判別システムの概要

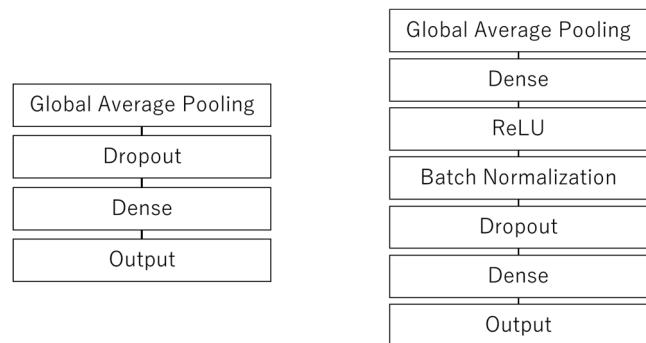
3.1 ファインチューニングモデルと諸設定

本研究では、ファインチューニングを行う CNN モデルとして、ImageNet と呼ばれる大規模画像データセットによって学習済みの「EfficientNet」[4]を採用した。EfficientNet は、モデルの深さ、広さ、解像度の 3 つの要素のバランスをとりながらスケールアップすることで、総パラメータ数を抑えながら、精度を向上することができる。EfficientNet には、スケールの大きさ順に、B0～B7 の 8 種類のモデルがあるが、効率のよい精度向上が見込めるとされている B0 から B4 の性能を比較した結果、最大正答率が 86.56%と、最も高かった EfficientNet B4 を採用した。

EfficientNet B4 のモデル構造の概略図を図 1 に示す。本研究では、最終層 (Final Layers) を、図 2 のように再構築した。また、EfficientNetB4 の学習の初期設定を表 2 に示す。



図 1 EfficientNet B4 の構造の概略図



(a) Final Layers

(b) 再構築した Final Layers

図 2 Final Layers の比較

表 2 EfficientNet B4 の学習の初期設定

パラメータ	設定値
入力画像サイズ	380×380 px
バッチサイズ	64
最適化アルゴリズム	SGD (確率的勾配降下法)
学習率	0.001
モーメンタム	0.9

3.2 データセット

本研究では、判別対象として 34 種類の樹木の樹皮画像を各 30 枚用意し、モデルの学習に用いる学習データと、判別精度を測るための検証データとして 2 : 1 の割合で分割した。また、汎化性能を評価するための大規模テストデータを 1811 枚用意した。なお、学習データについては、ImageDataGenerator という Keras ライブラリに用意されたデータ拡張用クラスを用いて、画像の拡大縮小、ランダムに左右反転、回転、輝度変化を行い、水増しした。

4. モデルチューニングと性能評価

4.1 Grad-CAM を用いた再学習開始位置の検討

本実験では、Grad-CAM を用いて、モデルの各層における注目度を可視化した。可視化は、Conv ブロック 2 から Conv ブロック 7 までの各 Conv ブロックの最後の Convolution 層、Top_Conv 層に行った。可視化の結果例を図 3 に示す。図 3 から、ブロック 2、3 では、画像全体のあらゆる箇所のエッジや明るい部分に注目しており、ブロック 4 では、樹皮への注目も強いが、背景部分への注目も強いことが分かる。ブロック 5 では、背景部分への注目が減り、ブロック 6 以降は樹皮に注目していることが分かる。以上のことから、Conv ブロック 4 までの層では、エッジや明暗などの汎用的な特徴を、Conv ブロック 5 以降の層では、樹種判別特有の複雑な特徴を抽出していると考え、モデルの再学習開始位置を Conv ブロック 5 以降とした。

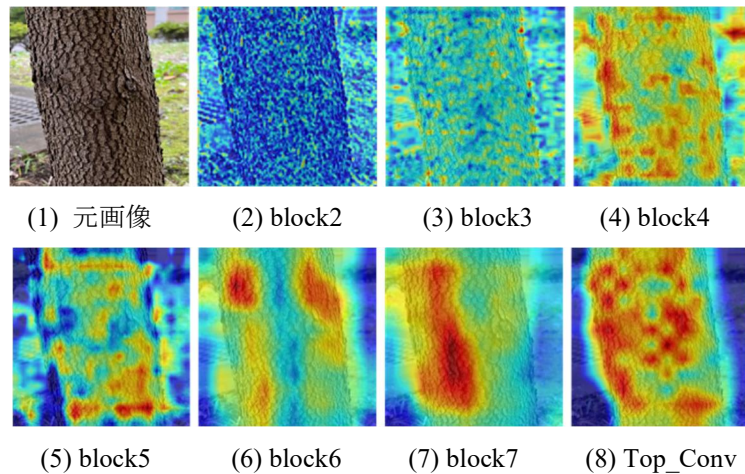


図3 各層での注目度の可視化

4.2 ハイパーパラメータのチューニング

正答率の向上のため、構築したモデルのハイパーパラメータの一つである入力画像サイズとバッチサイズの最適なサイズについて検討を行った。

まず、入力画像サイズを 380×380 px から 430×430 px まで 10 px ずつ変化させ、それぞれ 100 エポック学習させ結果を比較した。その結果、 410×410 px のモデルの正答率が 95.63% となり最も高くなった。

次に、バッチサイズを 8、16、32、64、128 としたモデルでそれぞれ 100 エポック学習させた結果を比較した。その結果、バッチサイズが 16 のモデルの正答率が 97.32% と最も高く、初期設定のバッチサイズ 64 のモデルでの正答率 (86.56%) を大きく向上するとともに、学習時間には大きな差はないことがわかった。バッチサイズが大きいと、データの特徴が平均化され、各データの個性が失われる可能性があるが、バッチサイズが小さいと、ノイズの影響を受け、局所解に収束してしまうリスクがあり、各データに敏感になると考えられる。したがって、同種類間のデータにおいて、ばらつきが少ない樹種判別では、バッチサイズを小さくしても、ノイズの影響を受けず、さらに、各データの個性を失わずに学習することができ、正答率が向上したと考えられる。

4.3 大規模なテストデータによる性能評価結果と考察

本研究で構築したモデルの汎用的な性能を評価するため、大規模なテストデータを用いた判別を行い、その結果を混同行列で表し (図4)、考察した。

図4の混同行列では、正解 (縦軸) に対して予測されたクラス (樹木、横軸) の数を濃淡で示しており、対角成分が最も濃く表されるのが理想的な結果となる。図4から、誤判別がほとんどないクラスも複数見られるが、誤判別されているクラスも少なくない。

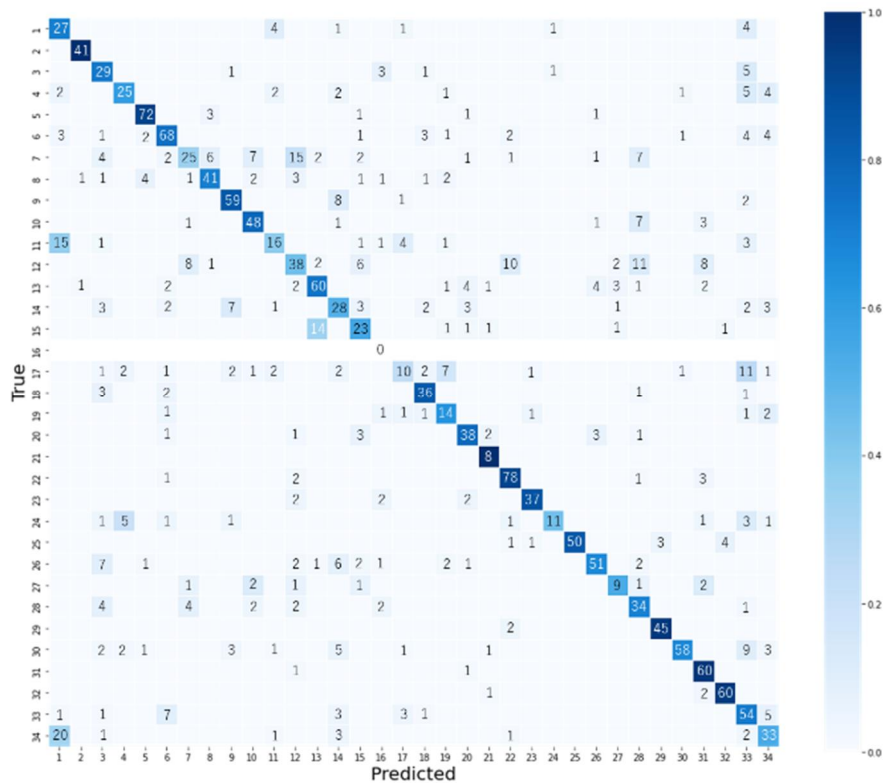


図4 テストデータの混同行列

例えば、1 番ツバキは 11 番サザンカと 34 番モッコクと、17 番コブシは 33 番ヤマモモとの誤判別が多いことが分かった。また、精度と再現率を算出した結果から、1 番ツバキは最も精度が低く、17 番コブシは、再現率が最も低いことが分かった。なお、全体の正答率は 71.5%となり、汎用的な利用にはさらなる性能向上の必要性があることがわかった。

5. まとめ

本研究では、樹皮画像を用いた樹種判別システムの開発を目的として、一般的なスマートフォンで撮影した 34 種の樹皮画像を用いて、既存の CNN モデルをファインチューニングすることでモデルの再構築を行った。また、大規模テストデータセットを作成し、モデルの汎用性を評価することで、樹皮画像を用いた樹種判別の有用性を評価した。

ファインチューニングモデルとして EfficientNet B4 を採用し、判別させた結果、正答率は 86.56%だった。そこで、再学習開始位置、画像サイズ、バッチサイズをチューニングした結果、正答率は 97.32%に向上した。最後に、1811 枚の大規模テストデータセットを判別させた結果、正答率は 71.5%となり、さらなる性能向上の必要性が示された。

以上の結果から、本研究で開発した樹皮画像を用いた樹種判別システムの有用性が示された。樹皮画像を用いた判別の場合、入力画像の精細さが重要となるため、今後、データセットの質を高めることで、さらなる精度向上ができると考えられる。

謝辞

本レポートは、2023 年 3 月に理工学研究科博士前期課程を修了した小林巧暉さんの修士論文を基に作成しました。

参考文献

- [1] Pierre Barré, Ben C. Stöver, Kai F. Müller, &Volker Steinhage: “LeafNet: A computer vision system for automatic plant species identification”, Ecological Informatics,40.50~56, 2017.
- [2] Wang-Su Jeon, and Sang-Yong Rhee: “Plant Leaf Recognition Using a Convolution Neural Network,” Int. J. Fuzzy Log. Intell. Syst., vol.17, no.1, pp.26-34, MAR 31 2017.
- [3] 梅本浩史,樹皮ハンディ図鑑,株式会社永岡書店, 2010.
- [4] Mingxing Tan, Quoc V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, International Conference on Machine Learning, vol.97, pp6105-6114, Jun. 2019.

ノイズのある原子間力顕微鏡データからの探針形状推定法の開発

松永 康佑

1. はじめに

原子間力顕微鏡 (Atomic Force Microscopy; AFM) は、生理的条件下でステージ上の試料分子の表面形状を計測することのできる強力な計測手法である[1]。最近、高速 AFM (High Speed AFM; HS-AFM) が開発され、画像の取得速度が劇的に向上し (1 秒あたり数十フレームに達する)、静的な表面形状だけではなく生体分子が動いている様子を観察することが可能となっている[2, 3]。HS-AFM によって、他の方法では観察できない生体分子の詳細な構造と機能の関係が解析でき、HS-AFM は現代の生物物理学において欠かせない技術となっている。HS-AFM の代表的な応用例としては、ミオシン V がアクチンフィラメントの上を歩行する様子[4]、F1-ATPase の回転触媒[5]、天然変性タンパク質の構造ダイナミクス[6]、CRISPR-Cas9 の機能ダイナミクス[7]の計測などがある。現在、HS-AFM 機器の空間分解能は横方向で約 2 nm、AFM ステージに対する垂直方向で約 0.15 nm である[8]。

機器の解像度とは別に、計測される分子表面形状の解像度は、計測に用いる探針と試料との相互作用によって影響を受ける。探針の先端形状、測定される形状情報(以後、画像プロファイルと呼ぶことにする)、分子表面の関係を図 1 に示す。探針の先端形状が十分に細い場合、得られる画像プロファイルは試料表面とほぼ同じになる。一方、先端形状が試料のスケールと比較して太い場合、画像プロファイルは試料表面からぼやける。先端の形状がわかっているならば、画像プロファイルを **deconvolution** して分子表面形状を近似的に再構築する **dilation** と呼ばれる演算が提案されている。したがって、分子表面形状を再構築するには、探針の先端形状を正確に知ることが重要である。また、先端形状の推定は、画像プロファイルから表面形状だけでなく、3D 分子構造を推測する上でも重要である。最近の AFM 画像プロファイルの分析では、(X 線結晶学や分子動力学シミュレーションなどの異なる実験または計算技術で得られる) 3D 分子構造と、仮定された先端形状から擬似 AFM 画像が作成され、実験的データと比較される。この分析では、実験データに最も似た擬似 AFM 画像を生成した 3D 構造が分子構造の最良の推定値として選ばれる。このような分析では、擬似 AFM 画像の精度が探針先端形状に依存する。したがって、先端形状はできるだけ正確に決定されるべきである。

現在、先端形状を推定するためには大まかに 3 つの方法がある。第一の方法は、走査電子顕微鏡 (SEM) または透過電子顕微鏡 (TEM) を使用した探針先端のイメージングである。しかし、SEM と TEM の両方とも試料の二次元投影しか提供しないので、さまざまな角度から先端をイメージングするための特別な装置を設置する必要があるが、それを日常のルーチンとして確立することは現実的ではない。さらに、先端は AFM 測定中に時間とともに

損傷する可能性があるため、AFM 測定前または後の先端形状の決定は必ずしも適切ではない。第二の方法は、特別な条件で AFM 測定を行い、先端形状を推定する。具体的には、事前に幾何学がわかっているサンプルを混ぜて計測を行い、探針形状を推定する。しかし特定のサンプルを用意できないケースや、サンプルを混ぜることで目的の試料分子に影響を与えることを考えると、常に使用できる手法ではない。

第三の方法は、事前知識なしに AFM データの画像プロファイル解析のみにより先端形状を推定するものである。Villarrubia によって開発された **Blind Tip Reconstruction (BTR)** は、AFM 画像から任意の先端形状を推定する古典的な方法である[9]。このアルゴリズムの背後にある考え方は、画像プロファイルの凸凹の細かさと深さが先端の鋭さよりも滑らかでなければならないというものである。そして、アルゴリズムは初めに非常に太い先端形状を仮定し、画像プロファイルの凸凹の細かさと深さに応じて先端形状を「彫る」ことで形状を推定する（詳細は以下で説明する）。しかし、BTR はノイズのない AFM 画像に対して完璧に機能するが、アルゴリズムはノイズに対して敏感であり、ノイズに対する閾値パラメータを決定するのが困難である、という弱点がある。これは、HS-AFM の場合特に従来の AFM よりもノイズに対して敏感であるため、重要な問題となる。Villarrubia の論文から四半世紀の間に、より良い推定のためにいくつかの改善や新しい方法が提案されているが、未だにノイズに弱い問題があり、BTR は HS-AFM データ解析には実用に至っていない。

最近、我々はこのノイズの問題を克服するために、オリジナルの BTR を拡張した微分可能 BTR(End-to-end differentiable BTR)を提案した[10]。この微分可能 BTR では、ノイズを適切に扱うために、近年の機械学習とニューラルネットワークの技術を用いていることが特徴である。具体的には、ノイズを考慮した損失関数によって探針を「彫る」か決定するとともに、探針形状の最適化には自動微分を用いることで複雑な損失関数を最適化することに成功している。本稿では、該当論文から微分可能 BTR の手法と応用例を紹介したい。

2. 手法

Villarrubia によるオリジナル BTR をまず簡単に紹介し、その後で微分可能な BTR を紹介する。まず、AFM ステージを xy 平面とし、ステージに対して垂直な高さを z 座標とし、ステージ位置を $z=0$ とする。次に、座標 (x, y) における試料表面の高さを $s(x, y)$ とし、分子表面の形状(高さ)を $i(x, y)$ とする。次に、探針形状のほうの座標系を導入する。 uv 平面を xy 平面に平行とし、その原点を探針の先端頂点に置く。すなわち、 uv 座標系では探針先端頂点が $z=0$ に対応する。この uv 座標系は探針と共に移動する移動座標である。探針表面の高さは $t(u, v) \geq 0$ とし、これが探針の形状を表す。便宜上、原点を通して探針の形状を反転させ、反転した探針の表面を $p(u, v) = -t(-u, -v) \leq 0$ と定義する。すると、mathematical morphology の dilation という以下の演算によって画像プロファイル $i(x, y)$ が得られる。

$$i(x, y) = \max_{u, v} [s(x - u, y - v) + p(u, v)].$$

これは概念的には、これは分子表面 $s(x, y)$ をカーネル関数 $p(u, v)$ で “convolve” したものと見なすことができる。しかし、dilation は畳み込みとは異なり非線形変換であり、探針と試料の間の物理的相互作用を表していることに注意されたい。更に、mathematical morphology の用語を借りて、 $s(x, y)$, $i(x, y)$, $p(x, y)$ 上の点またはそれより下に存在する点の集合を umbras と呼びそれぞれ $S = \{(x, y, z) \mid z \leq s(x, y)\}$, $I = \{(x, y, z) \mid z \leq i(x, y)\}$, $P = \{(u, v, z) \mid z \leq p(u, v)\}$ と定義する。すると dilation は、 $I = S \oplus P$ と書くことができる。測定された画像プロファイル $i(x, y)$ と反転した探針の形状 $p(u, v)$ の umbras I と P が与えられた時、試料表面 S は、mathematical morphology の別の演算子である erosion によって $S^{(r)} = I \ominus P$ と近似的に再構築できる。これは以下のように定義される、

$$s^{(r)}(x, y) = \min_{u, v} [i(x + u, y + v) + t(u, v)] = \min_{u, v} [i(x + u, y + v) - p(u, v)].$$

Erosion は、概念的には画像プロファイル $i(x, y)$ を $p(u, v)$ で “deconvolve” したものと見なせる。しかし、dilation と同様に erosion も非線形変換であることに注意されたい。Erosion は、用いている探針の形状 P よりも小さい凸を彫って (erode して) 取り除く性質がある。この erosion の性質から、 $S^{(r)} \oplus P = (I \ominus P) \oplus P = I$ および $S^{(r)} \supseteq S$ を導くことができる。これは、 $S^{(r)}$ が探針の形状 P で画像プロファイル I を再現する試料表面の (最小の) 上限であることを意味する。一般に、umbra B による umbra A の erosion の dilation は opening と呼ばれ、シンプルに $(A \ominus B) \oplus B = A \circ B$ と書けるので、上記の関係は単に $I \circ P = I$ としてまとめることができる。ただし、 $I \circ P = I$ が必ずしもどんな P に対しても成り立つわけではないことに注意されたい。この関係は I を生成した正しい探針形状 P またはそれよりも細い形状 (例: δ 関数的に細い形状) の場合にのみ成り立つ。一般に、opening は探針の形状 P よりも小さい凸を取り除く。なぜなら、erosion がそのような小さい凹みを一旦取り除いてしまうと、次の膨張ではそれらを再構築できないからである。したがって、 $I \circ P = I$ は、例えば、正解の探針形状よりも太い形状の場合には成り立たない。

Villarrubia の BTR アルゴリズム (以下、オリジナル BTR と呼ぶ) は、AFM 画像から探針の形状の上限を推定するものである。BTR は、 $I \circ P = I$ の関係から始まり、この関係から P の上限条件を構築する。Opening の重要な性質として、 $I \circ P = \bigcup \{P + \mathbf{a} \mid P + \mathbf{a} \subset I\}$ がある。ここで $\mathbf{a}$ は 3 次元の並進である。これを $I \circ P = I$ と組み合わせると、 $I = \bigcup \{P + \mathbf{a} \mid P + \mathbf{a} \subset I\}$ となる。これは、 I の全ての点が P の並進に含まれる一方で、 P の一部が I の表面より上に延びないように、その並進は制限されていることを意味する。特に、 I の表面上の全ての点は、 P のこれらの平行移動の一つ以上の表面によって触れられなければならない。並進した探針 P が I の表面 $\mathbf{c} \in I$ で触れると仮定し、移動する uv 座標の原点 (探針の先端) から $\mathbf{c}$ へのベクトルを $\mathbf{d}$ と定義する。この場合、並進した P は $P + \mathbf{c} - \mathbf{d}$ と書け、 P の形状に対する上限は、並進した P が I の表面上またはその下にあるということになる。すなわち、

$$P \subseteq I - \mathbf{c} + \mathbf{d}.$$

この条件は、 $\mathbf{d}$ が存在する I の表面または内部の任意の $\mathbf{c}$ に対して成り立つ。

オリジナル BTR は、この P の上限条件を満たすアルゴリズムを与える。アルゴリズムは、真の P を部分集合として含む **umbra** を初期条件として、太い形状（通常は平らな上部を持つ正方形の柱）から始める。探針が I の表面の点 $i(c_x, c_y)$ に触れると、移動する uv 座標系における I の高さは、 $\mathbf{d}$ における $i(c_x, c_y) = p(d_u, d_v)$ であり、他の (u, v) の位置における I の高さは $i(c_x + u - d_u, c_y + y - d_v)$ である。この 2 つの高さの差に、接触点 (d_u, d_v) での P の高さを加えると、 $dill = i(c_x + u - d_u, c_y + y - d_v) - i(c_x, c_y) + p(d_u, d_v)$ が得られる。すべての可能な $\mathbf{d}$ に対して $dill$ を計算すると、 $dill$ の最小値が得られ、それを $dill_{min}$ と書くことにする。 $dill_{min} + thresh < p(u, v)$ の場合、探針の形状は、より小さい高さで更新され、 $dill_{min} + thresh < p(u, v)$ となる。ここで、 $thresh > 0$ は、画像のノイズによる凸凹と探針形状の凸凹の程度から決定すべきパラメータである。以上の計算は、画像プロファイル $i(c_x, c_y)$ 上のすべての (c_x, c_y) 位置で行われる。すべての計算ステップは、探針の形状が収束するまで繰り返される。最終的に、真の形状 P の上限である $P^{(r)} \supseteq P$ が再構築される。このアルゴリズムでは、 $thresh$ は探針の推定精度にとって重要なパラメータである。実際、画像にノイズが含まれている場合、再構築された探針の形状は、ノイズの凸凹に強く影響される。ベストな $thresh$ 値を選ぶためには、分子表面の凸凹とノイズの凸凹をうまく分離するスケールを見つける必要があり、これは特に初心者にとって困難な課題である。

以下では、我々の微分可能 BTR を紹介する。オリジナル BTR において探針の再構築がノイズに影響される事実は、機械学習においては過学習の問題と見なすことができる。機械学習において、過学習を防ぐ典型的なアプローチは、正則化を含む適切な損失関数を使用することである。アイデアのポイントは、オリジナル BTR のように、画像プロファイルの凸凹から探針を更新する(削る)かどうかを個別の座標で決定するのではなく、ノイズを考慮した損失関数に従って、画像全体から探針を更新する(削る)かどうかを統計的に決定することにある。オリジナル BTR に倣って、微分可能 BTR は $I \circ P = I$ の関係に基づくが、ノイズの存在下では等号が成り立たないため、条件を緩和する。具体的には、ノイズが空間的に独立なガウスノイズであると仮定し、最小化すべき損失関数は平均二乗誤差を用いる。

$$MSE(p) = \frac{1}{N} \|i^{(r)} - i\|^2 = \frac{1}{N} \sum_{x,y} (i^{(r)}(x, y) - i(x, y))^2.$$

ここで、 N はピクセルの数であり、 $\| \|^2$ は L2 ノルムである。 $i^{(r)}(x, y)$ は、ある探針の形状 P による opening $I^{(r)} = I \circ P = (I \ominus P) \oplus P$ であり、展開して以下のように書き直すことができる。

$$i^{(r)}(x, y) = \max_{u,v} \left[\min_{u',v'} [i(x - u + u', y - v + v') - p(u', v')] + p(u, v) \right].$$

微分可能な BTR は、その勾配を通じて探針の形状に対する平均二乗誤差を最小化することで、最適な探針の形状 $p(u, v)$ を見つける。Dilation と erosion に現れる最大および最小

の関数は一見微分不可能に見えるが、畳み込みニューラルネットワークの Max Pooling レイヤーの実装に従って微分する。たとえば、ピクセルのインデックスで離散化された 1 次元の探針の形状 p_u を考える。 $h = \max[p_1, p_2, p_3]$ の勾配は、 p_3 が最大値であれば、 $\partial h / \partial p_1 = 0$ 、 $\partial h / \partial p_2 = 0$ 、 $\partial h / \partial p_3 = 1$ であり、これは argmax 関数と同等である。他の画像処理の研究では、 $\max$ や $\min$ にエントロピー正則化を加えて滑らかにした上で微分しているものもあるが、本研究では単純な argmax および argmin 関数を使用する。

残念ながら、上記の損失関数を最小化する問題は、不良設定問題である。例えば、正しい探針の形状以外にも、 δ -関数のような非常に細い探針が損失関数の値をほぼゼロにすることができる。そのような細い形状は、ノイズのある AFM 画像に対する過学習の結果と見なすことができる。したがって、損失関数を小さく保ちつつ、過学習を避けて探針形状の上限を見つけるために、正則化項を導入する。本研究では、探針の形状 P の L2 ノルムを正則化項として追加する。最終的に、微分可能な BTR で使用される損失関数は、次のようになる。

$$\text{MSE}(P) + \lambda \|p\|^2 = \frac{1}{N} \sum_{x,y} \left(i^{(r)}(x,y) - i(x,y) \right)^2 + \lambda \sum_{u,v} p^2(u,v).$$

正則化項を適用することで、細い探針は大きな $p^2(u,v)$ 値を持つため、鈍い探針の形状が得られると期待される。正則化パラメータ λ の適切な値を見つけるためには、 $\lambda=0$ または非常に小さい λ から始めて、 $\text{MSE}(P)$ を小さく保ちつつ、出来る限り太い探針を見つけるために λ を徐々に増やしていくことが一般的である。適切な λ 値をより体系的に見つけるためには、本研究では **one standard error rule** を使用することを提案する(後ほど説明する)。

更に、探針の再構築の初期条件として、オリジナル BTR と同じように、正方形の柱を使用する。この初期条件の選択は、太い探針または探針の形状の上限を見つけるのに寄与する。最適化アルゴリズムは、ニューラルネットワーク/機械学習のフレームワークである Julia プログラミング言語の Flux.jl パッケージで実装した。具体的には、**dilation** と **erosion** を微分可能な関数から成る Flux のカスタム層を独自に定義し、ニューラルネットワークの訓練と同じように自動微分により誤差逆伝播法を適用した損失関数を最適化した。

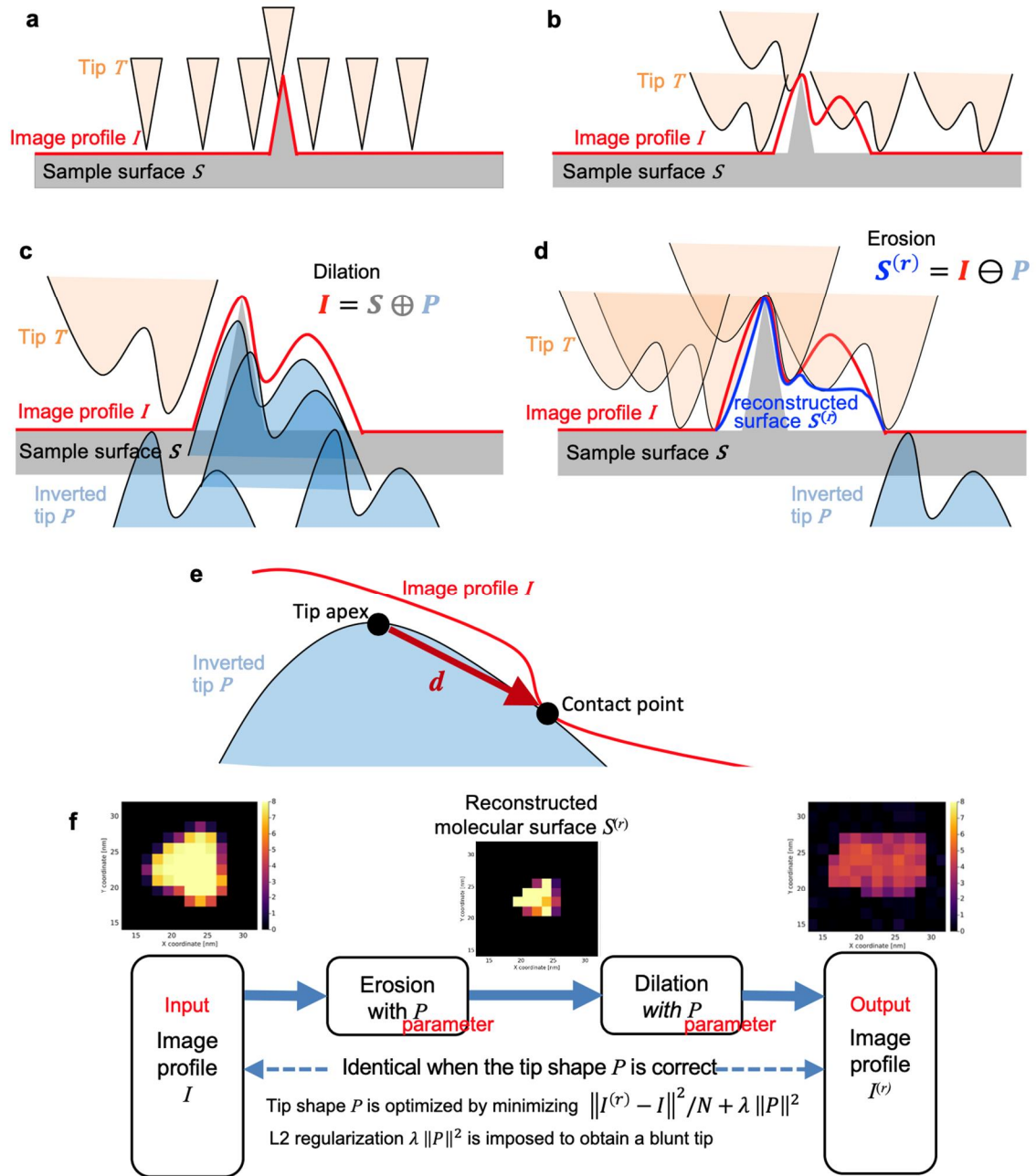


図 1: Mathematical morphology の演算子と微分可能 Blind Tip Reconstruction (BTR) の概要。a, 細い探針で表面を走査して得られた画像プロファイル。b, a と同じ表面を太い先端で走査して得られた画像プロファイル。c, Dilation を示す模式図。d, Erosion を示す模式図。ここでは、直感的な説明のため、erosion は反転した探針ではなく元の探針で表現している。e, オリジナル BTR の背後にある考え方を説明するための、接触点と先端頂点の位置関係。f, 微分可能 BTR の模式図。文献[10]より CC BY 4.0 で引用。

3. 結果

ノイジーな疑似 AFM データの解析

以下では既知の分子構造から人工的に疑似 AFM 画像プロファイルを生成し、画像プロファイルから探針形状や分子表面形状を再構成できるか調べる双子実験を行い、オリジナル BTR と微分可能 BTR の精度を比較した。双子実験には、既知の分子構造としてミオシン V モータードメイン (PDB ID: 1OE940) を用いて、20 フレームから成る疑似 AFM 画像を作成した。探針形状としては一般的な形 (半球と円錐の切断体を組み合わせたもの) を用いた。各フレームで構造の向きはランダムに変化させた。ノイズに対する頑健性を調べるために、全てのフレームの疑似 AFM 画像に標準偏差 $\sigma = 0.3 \text{ nm}$ の空間的に独立したガウスノイズを加えた。ここで、 $\sigma = 0.3 \text{ nm}$ は、現在の HS-AFM 測定での典型的なノイズのサイズである。こうして、100 個の異なるノイズの実現値を用いて、ノイズのある疑似 AFM データ (各々が 20 フレームを含む) を 100 セット作成した。BTR の計算では、パラメータ ($thresh$ と λ) と $I \circ P = I$ との関係性を調べるため、広い範囲のパラメータ値を適用し、MSE をプロットした。ここで MSE の平均と標準偏差は交差検証によって求めた。一方で、再構成された探針形状と表面の精度は、異なるノイズの実現値 (100 セット) を用いて評価した。

図 2b と 2c を見ると、両方の BTR で、 $MSE(P)$ はノイズの実現値に依らず同じように急激に増加する。しかし、オリジナル BTR の探針形状をプロットしてみると、再構成された探針形状はパラメータだけでなくノイズの実現値にも依存しているように見える (図 2d)。実際、同じ $thresh = 4.0 \text{ nm}$ でも、オリジナル BTR はノイズの実現値に応じて探針形状を真の形状よりも細い場合と太い場合がある (図 2d 右)。また、 $thresh \ll 1$ の場合では、探針形状は細くなる (図 2d 左)。これは、オリジナル BTR アルゴリズムが疑似 AFM 画像のノイズに対して探針の形状を過剰適合したことを意味する。真の探針形状からの RMSD (Root Mean Square Displacement) は、 $thresh \approx 4.0 \text{ nm}$ あたりが最良のパラメータ値であることを示している (図 2f)。 $thresh$ の値の選び方は以前の研究で議論されている。Villarrubia は、単一フレームの疑似 AFM 画像の分析から、 $thresh$ に $2 \sim 4\sigma$ を使用することを提案している。Villarrubia が単一フレームの画像のみを分析したが、フレーム数によって $thresh$ を超えるノイズレベルの確率は変わるため、フレーム数に依存する可能性がある。この意味で、 $thresh = 4.0 \text{ nm}$ は現在の 20 フレーム AFM 画像に適切なものであると考えられる。また、Dongmo らは、 $thresh$ パラメータを低い値から高い値に変えることで探針の体積の変化を監視することを提案している [11]。彼らは、推定された探針の体積の変化が大きくなるパラメータが最適なものを見なされるであろうと議論した。現在の文脈では、そのパラメータ値は $MSE(P)$ が急激に増加し始める値 ($thresh \approx 4.0 \text{ nm}$ 、図 2b) に対応するだろう。

一方、微分可能 BTR によって得られた探針形状は、オリジナル BTR よりもパラメータ依存性が少なく、ほとんどのノイズの実現値に対してより安定して同様の探針形状 (地面の真実に近い) を推定できている (図 2e)。興味深いことに、 λ が大きくなると正則化のペナ

ルティによって（図 2e 右）探針の形状は太くなるが、 λ が小さくなくても探針形状が細くはならなかった。微分可能 BTR におけるパラメータに対するこのロバスト性は、初心者のユーザーが広い範囲からパラメータ値を選んでも正確な形状推定を行うのに役に立つだろう。オリジナル BTR での再構成された探針形状の平均 RMSD は $thresh = 4.0 \text{ nm}$ において 2.0 nm であり、微分可能 BTR では $\lambda = 10^{-7}$ において 0.4 nm であった（図 2f および 2g）。

図 2h は、再構成された探針のものを使用して擬似 AFM 画像（図 2a）の 1 フレームを deconvolve することで得られた分子表面を示す。オリジナル BTR によって再構築された分子表面は、再構成された探針の形状が真の形状よりも大きな形状のエッジを erosion するため、ややぼやけて見える。一方、微分可能な BTR は、ミオシン V モータードメインの形状の輪郭をうまく deconvolve している。真の表面形状からの平均 RMSD は、オリジナル BTR で $thresh = 4.0 \text{ nm}$ において 2.1 nm 、微分可能 BTR での $\lambda = 10^{-4}$ において 1.1 nm であった（図 2i）。

これまで、ノイジーな AFM 画像でさえも λ の下限を見つけていないが、ノイズレベルが現在の損失関数に対して十分に小さいため、過剰適合を避けることが示唆される。 λ の下限を探るため、意図的にノイズレベルを増加させ、別のツイン実験を実施した。このツイン実験では、標準偏差 $\sigma = 1.2 \text{ nm}$ の大きなノイズレベルを適用した。結果から、 λ の小さい値では推定精度が悪くなることがわかったが、これは λ の低い値を持つ探針がノイズに過剰適合する傾向があることを示している。

この状況で λ の適切な値を選ぶため、いわゆる 1 標準誤差のルール 37 を使用することを提案する。このルールでは、最初に交差検証を実行し、 $MSE(P)$ の平均と標準偏差を計算する。次に、最小の $MSE(P)$ を持つ λ 値を見つけ、最小値の 1 標準偏差内または周辺に平均が落ちる最大の λ を選ぶ。このルールの背後にある考え方は、 $MSE(P)$ の観点からはほぼ等しいように見える探針の形状のセットから最も鈍い探針の形状を選ぶことである。このツイン実験の場合、one standard error の境界よりもわずかに上の $\lambda = 10^{-3}$ が、最も正確な探針の再構築をもたらすことがわかった。

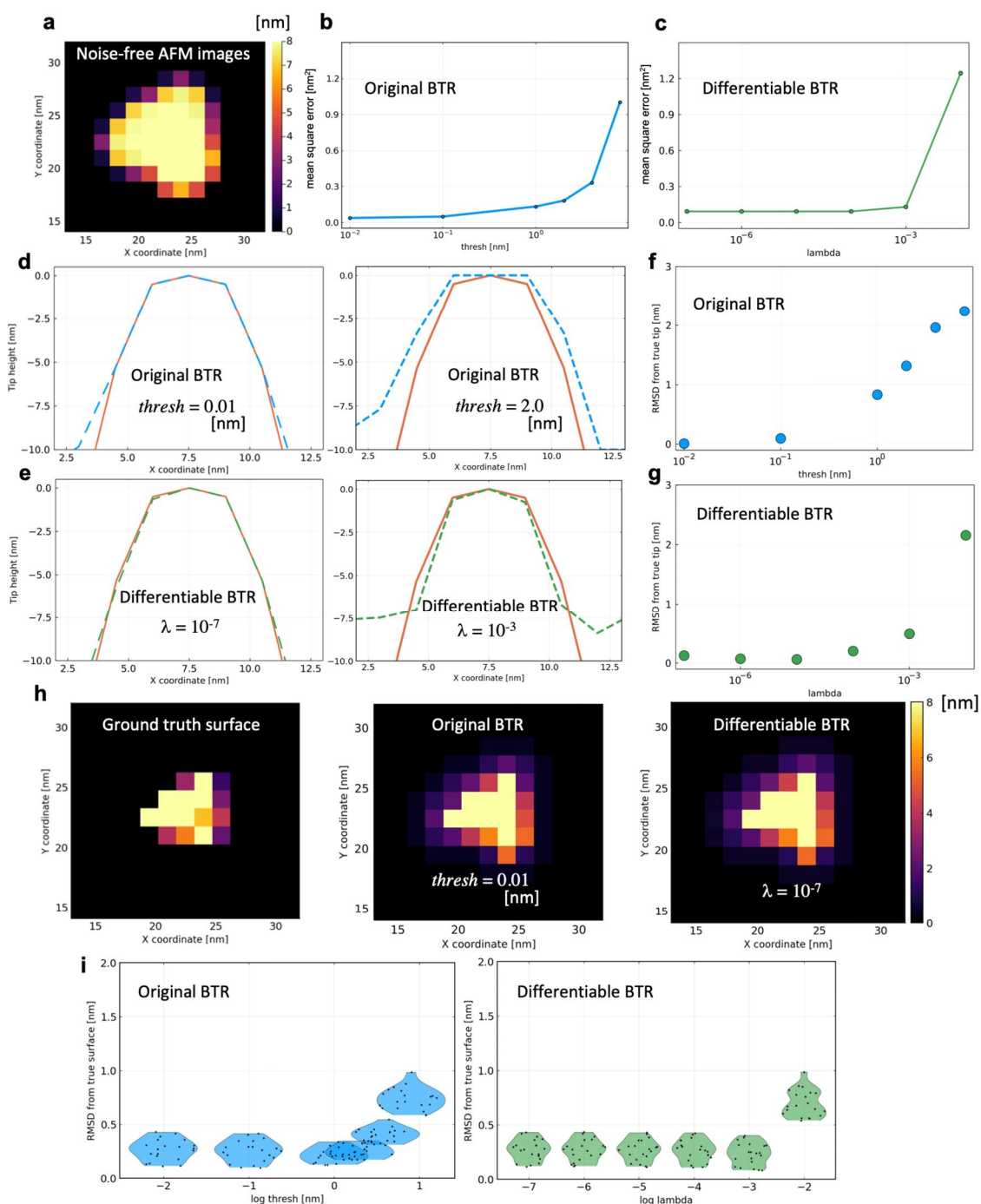


図 2: ノイジーな人工 AFM データの解析結果。a, 用いられた 20 枚の画像の 1 枚目。b-c, さまざまなパラメータ値で最適化された MSE。MSE の平均と標準偏差は、交差検証を用いて 20 フレームの画像の単一セットから計算された。影のエリアは一標準誤差の範囲を示す。d-e, オリジナル BTR と微分可能 BTR（それぞれ破線の青と破線の緑で示される）による x 軸に沿った再構築された先端形状、基準値（赤線）との比較。影のエリアは標準偏差を示す。逆畳み込みに用いられた先端形状は黒線で示される。f-g, 基準値からの再構築先端の二乗平均平方偏差 (RMSD) がバイオリンプロットで可視化されている（同じ color スキーム）。h, 再構築された先端による逆畳み込みで得られた分子表面。i, 全フレーム（20 枚の画像）の基準

値からの逆畳み込みされた分子表面の RMSD がバイオリンプロットで可視化されている。文献[10]より CC BY 4.0 で引用。

実際の実験データの解析：Myosin-V Walking on an Actin Filament

次に、実際の AFM データを分析することによって 2 つの BTR アルゴリズムを比較した。ここでは、Kodera らによって測定されたミオシン V の HS-AFM データ[4]を分析した。ミオシン V は、ATP 加水分解自由エネルギーによってアクチンフィラメントに沿って直線的に移動するホモ二量体分子モーターである。ミオシン V は、二量体の中央に取り付けられた貨物を輸送する分子機能を持つ。Kodera らの HS-AFM ムービーは、ミオシン V が ATP 依存的な構造変化と結合してアクチンフィラメント上を歩行する様子をうまく捉えたものである。歩行ステップ中のミオシン V の分子表面の詳細な形状を得るために、AFM 画像からの 探針形状推定と deconvolution を行う。ここでは、ミオシン V の一步のステップを捉えた HS-AFM ムービーの 30 フレームを分析した。

図 3a に示すように、HS-AFM 画像には、分子から x 軸方向へ向かって数多くの水平な縞が見える。この縞は探針の走査方向に対応しており、パラシュート効果と呼ばれる。パラシュートとは、チップの走査速度が大きい場合、試料分子表面の高さがチップの走査方向に沿って突如減少したとき、探針が分子表面から完全に離れることで起こる。探針が別の地点でステージまたは試料表面に着地するまでに時間がかかるため、走査方向に縞が発生する。試料分子の熱揺らぎによって探針がステージから持ち上げられることで起こるため、縞の発生は確率的である。一方、パラシュート効果が発生した後、傷跡の高さの減少パターンは決定論的であり、dilation の一部としてモデル化できる。したがって、ここでは縞の高さの減少パターンはシグナルの一部として取り扱い、実効的な探針形状の一部と見なすことにする。

図 3b と 3c は、オリジナル BTR および微分可能 BTR を用いて、さまざまな *thresh* および λ パラメータで 30 フレームの HS-AFM 画像から計算された $MSE(P)$ の平均値と標準偏差を示す。図 3d と 3e は、両 BTR による実際の HS-AFM 画像から再構築されたチップ形状の x 軸と y 軸の断面を示す。オリジナル BTR で再構築されたチップ形状は、 x 軸方向に沿って厚くなる傾向がある (図 3d 左) が、 y 軸方向に沿っては非常に薄くなる (図 3d 右)。これは主にパラシュート効果に起因する。パラシュート効果が確率的に発生した後、探針の高さの減少パターンは、 x 軸に沿って dilation の一部としてモデル化できる。しかし、 y 軸に沿っては、パラシュート効果が現れるかどうかは、熱で揺動する分子表面がたまたま高い (大きな z 座標を持つ) 場合だけである。したがって、 y 軸に沿っては、パラシュート効果と非パラシュートラインの違いによる凸凹は確率的なノイズとして現れる。実際、オリジナル BTR で推定された探針形状を見ると、 y 軸に沿った大きなへこみをチップ形状に起因するものと誤解し、チップ形状を過度に erode してしまっていることがわかる。一方、微分可能な BTR によって再構築されたチップ形状は、 x 軸の右側でのパラシュートによる滑らか

な減衰をよく捉える (図 3e 左) が、 y 軸に沿ってもオリジナル BTR のように細くならず、ある程度の太さを保ち、パラメータ依存性が少ない ($\lambda=0.0010$ を除く、図 3e 右)。

図 3h および 3i は、オリジナル BTR と微分可能 BTR で再構築された先端形状で逆畳み込みされた HS-AFM ムービーの第 27 フレームの分子表面を示す。第 27 フレームは、以前の研究でこのフレームのために Flexible fitting 分子動力学シミュレーションを実行した画像である[12]。大きなノイズレベル ($\sigma=1.2$ nm) の双子実験の結果に従い、ここでは微分可能 BTR の λ 値を選択するために One standard error rule を使用した。具体的には、One standard error rule の境界または境界よりわずかに上回る $\lambda=0.0051$ を選択した (図 3c を参照)。オリジナル BTR では、このパラメーター値の先端半径が、 $\lambda=0.0051$ の微分可能 BTR のそれと同等 (またはわずかに大きい) であるため、 $thresh = 1.95$ nm の先端を使用した (図 3d および 3e)。オリジナル BTR の探針形状で deconvolve した分子表面は (図 3h)、ミオシン V の脚は元々の画像プロファイルよりも細くなるが、パラシューティング効果のアーティファクトによる水平な縞がまだ存在する。一方、微分可能 BTR の先端形状で deconvolve された分子表面は、パラシューティングの縞が少なくなる (図 3i)。これは、 y 軸方向に沿った探針形状が太く、 y 軸に沿った画像プロファイルの erosion による更新が広範囲に行われるためである。

Deconvolution で得られた分子表面はどちらが正確だろうか？これを検証するために、これらの表面と構造モデルとを比較した。ここでは、同じデータの第 27 フレームに対して、以前の Flexible fitting 分子動力学シミュレーションの研究から取られた 10 個の構造を使用した[12]。その研究では、ミオシン V とアクチンフィラメントの構造をホモロジーモデリングによってモデル化し、次に 27 フレームの画像に一致するように束縛を課して粗粒子分子動力学シミュレーション(Flexible fitting)を実行した。Flexible fitting で使用された探針形状は現在のものとは同じではないので、単一構造ではなく 10 個の構造を候補構造として使用した (図 3f)。それぞれの構造について、網羅的に並進・回転を行って疑似 AFM 画像を生成し、それらの画像と HS-AFM データの第 27 フレームとの相関係数を計算した。そして、10 の構造の最大相関係数と対応する並進・回転を記録した。図 3g は、両方の BTR の 10 の構造の最大相関係数を比較する。微分可能 BTR による係数が統計的に有意に改善していることがわかる。図 3j および 3k は、2 つの BTR によって deconvolve された分子表面および最大相関係数を持つ構造モデルによって得られた分子表面の断面を比較している。図では、アクチンフィラメントはおおよそ $y=13$ nm の位置にある。オリジナル BTR の再構築された探針形状は、 y 軸に沿って形状が細いため、画像プロファイルを erode することがほとんどできていない (赤と青のラインは図 3j でほぼ同じ)。一方、微分可能 BTR の先端は、画像プロファイルを十分に erode している。特に、微分可能 BTR で deconvolve された分子表面は、構造モデルと良く一致していることがわかる。

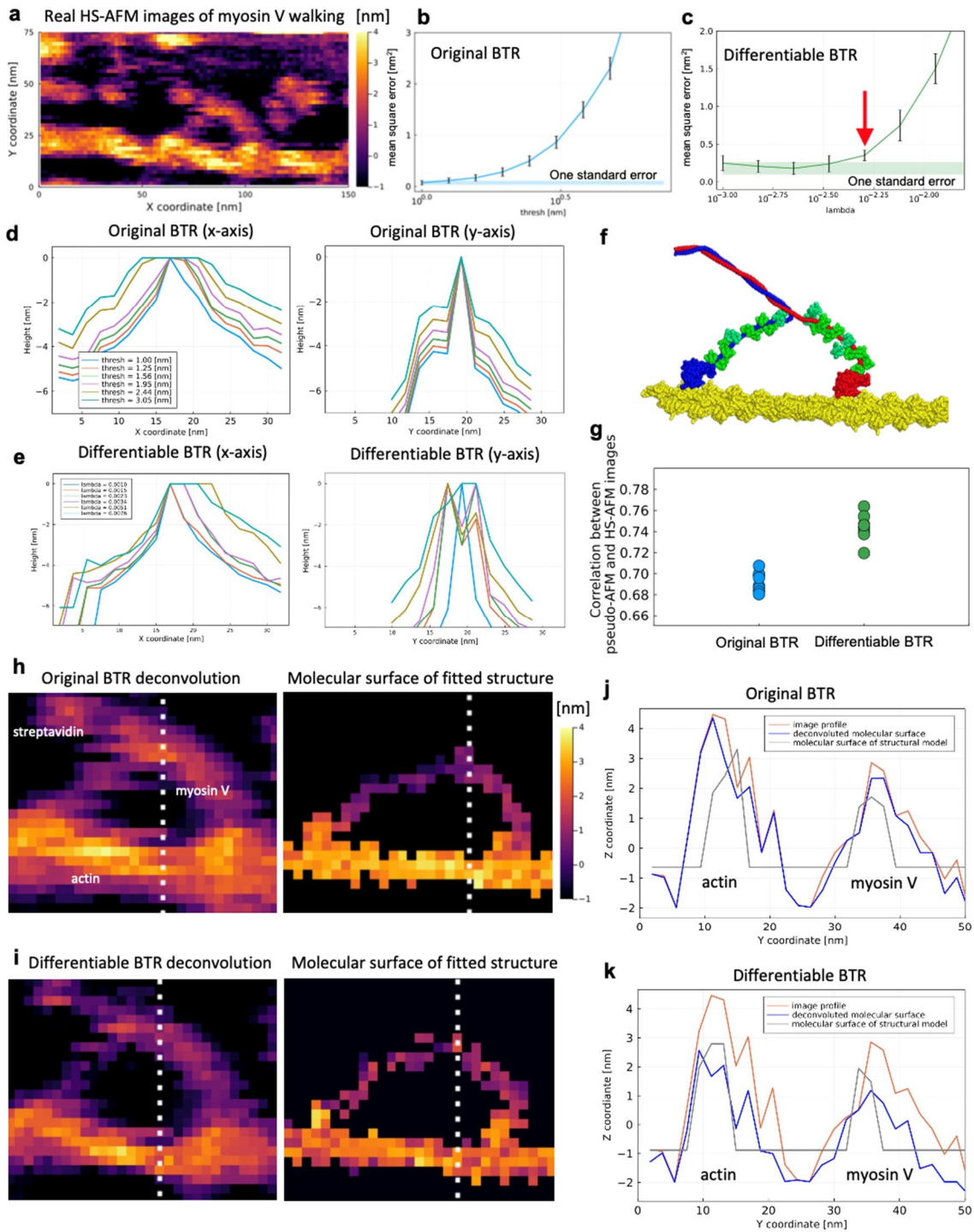


図 3: ミオシン V の歩行の HS-AFM データの Blind Tip Reconstruction (BTR)。a, 解析に用いた 30 枚の画像の 27 枚目。b-c, 各パラメータ値での MSE。MSE の平均と標準偏差は、交差検証により計算された。影のエリアは一標準偏差の範囲を示す。赤い矢印は逆畳み込みのために選ばれた λ を示す。d-e, x 軸に沿ったオリジナル BTR と微分可能 BTR による再構築された先端形状。f, 以前の柔軟なフィッティング分子動力学シミュレーションから選ばれた 10 構造の代表構造。g, 27 枚目の HS-AFM 画像と各 10 構造から生

成された擬似 AFM 画像との最大相関係数。h, BTR で推定された探針形状による deconvolution で再構築された分子表面と、最大相関係数を有する構造モデルの分子表面との比較。白い破線は j および k で使用された断面を示す。j-k, y 軸に沿った分子表面の断面。赤線は画像プロファイル、青線は再構築された分子表面、灰色の線は構造モデルから作成された分子表面である。文献[10]より CC BY 4.0 で引用。

4. まとめ

この研究では、AFM 画像のノイズを考慮した正則化項を含む損失関数に基づいて、過剰適合を避ける端から端の微分可能 BTR 方法を提案した。定義した損失関数を、深層学習研究の最近の進展によって開発されたフレームワークを利用して最適化した。双子実験の結果は、微分可能 BTR がオリジナル BTR よりもノイズに対して頑健であることを示している。実際の HS-AFM データへ応用した結果、微分可能 BTR が先端形状を再構築し、deconvolve された分子表面が構造モデルと良く一致していることを示した。微分可能 BTR は非常に一般的であるため、ノイズの多い AFM 画像だけでなく HS-AFM データの解析に対する標準的な方法になると期待される。

5. 謝辞

共同研究者である埼玉大学の大金智宏氏と、京都大学の湊上壮太郎特任准教授、高田彰二教授に感謝します。また、Myosin V の HS-AFM データを提供してくださった金沢大学の古寺哲幸教授に感謝します。本研究は JST CREST「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用」領域（課題番号: JPMJCR1762）による支援を受けています。

参考文献

- [1] Binnig, G., Quate, C. F. & Gerber, Ch. Atomic Force Microscope. Phys. Rev. Lett. 56, 930–933 (1986).
- [2] Ando, T., Uchihashi, T. & Fukuma, T. High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Progress in Surface Science 83, 337–437 (2008).
- [3] Ando, T. High-Speed Atomic Force Microscopy in Biology. (Springer, 2022).
- [4] Kodera, N., Yamamoto, D., Ishikawa, R. & Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. Nature 468, 72–76 (2010).
- [5] Uchihashi, T., Iino, R., Ando, T. & Noji, H. High-Speed Atomic Force Microscopy Reveals Rotary Catalysis of Rotorless F₁-ATPase. Science 333, 755–758 (2011).
- [6] Kodera, N. et al. Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy. Nat. Nanotechnol. 16, 181–189 (2021).
- [7] Shibata, M. et al. Real-space and real-time dynamics of CRISPR-Cas9 visualized by

- high-speed atomic force microscopy. *Nat Commun* 8, 1430 (2017).
- [8] Ando, T. High-speed atomic force microscopy and its future prospects. *Biophys Rev* 10, 285–292 (2018).
- [9] Villarrubia, J. S. Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. *J. Res. Natl. Inst. Stand. Technol.* 102, 425 (1997).
- [10] Y. Matsunaga, S. Fuchigami, T. Ogane, and S. Takada, End-to-End Differentiable Blind Tip Reconstruction for Noisy Atomic Force Microscopy Images, *Sci Rep* 13, 129 (2023).
- [11] Dongmo, L. S. et al. Experimental test of blind tip reconstruction for scanning probe microscopy. *Ultramicroscopy* 85, 141–153 (2000).
- [12] Fuchigami, S. & Takada, S. Inferring Conformational State of Myosin Motor in an Atomic Force Microscopy Image via Flexible Fitting Molecular Simulations. *Front. Mol. Biosci.* 9, 882989 (2022).