

研究終了報告書

「ディープラーニングを利用した植物表現型の定性的・定量的計測技術の開発」

研究期間：2017年10月～2021年3月

研究者：戸田 陽介

1. 研究のねらい

近年におけるイメージング技術や計測機器の性能の著しい向上、さらには導入コストの低下によって、植物に関連したあらゆる表現型を画像として記録することが可能となった。顕微鏡を用いて得られる細胞・器官単位の画像から、ドローンによる植物個体(群)を対象とした空撮画像まで、対象物の規模は目的に応じて多様である。特に後者に関しては、ICT 農業への利用が検討されており、目的とする表現型を画像から自動的に抽出する技術の需要は今後ますます高まることと考えられる。しかしながら従来技術では、撮影環境を完全制御できる条件下で、かつ解析の目的が対象物の面積や輝度といった低次表現型である場合にのみ、自動画像解析が適用可能であった。あらゆる局面で画像解析が植物表現型解析におけるボトルネックとなりつつあるが、抜本的な解決法は現在に至るまで呈示されていない。申請者は、深層学習を画像解析に導入することによって、植物表現型解析の技術的限界の突破を試みる。

深層学習は既存の画像解析技術では対応できない複雑な画像を扱うことのできる手法として近年脚光を浴びている。本手法を用いると、着目すべき物体の特徴の設計をコンピュータにまかせることができることから、低次表現型ではなく、「画像が一体何を意味しているのか」といった、これまでは人でしか判断できなかったかのような高次表現型を高精度で解析できると期待されている。医療・工学分野では積極的にその利用が進められているが、植物分野における当該技術の活用事例は未だ限定的であり、その力は未知数である。申請者はこれまでに、本手法を利用した植物の気孔の定量的解析技術の開発に従事してきた経験から、当該分野における親和性は極めて高いと確信した。

そこで本提案研究では、植物特徴量の定性的・定量的解析に特化した画像解析技術の開発に取り組む。植物病虫害診断や種子測定技術など、従来技術では困難であった植物フェノタイピング法の実現、およびその利用による当該分野の研究の一層の加速を狙う。さらには、付随して発生する一連の課題を明確化し、逐次解決していくことで基盤的知見の整備を目指す。

2. 研究成果

(1)概要

本研究では、取り組むべき解決課題を以下のように設定した。

- a. 教師データ収集コストの高さ
- b. 教師データのアノテーションコストの高さ
- c. 作成した深層学習モデルのブラックボックス性による解釈性や真正性の不透明さ

深層学習を活用し、かつ汎化性能が高いモデルを作成するためには、従来手法と比べ、多量の教師データが必要となることが多い。車や人などが対象とする画像については入手が比較的容易であるものの、作物に関連した画像については季節的な制約や、栽培現場でのデータ収集の労力コストが問題となっていた。したがって、次項(2)の研究テーマ A「集合知を利用した病害虫診断技術の開発」において、画像を用いた作物の病害虫診断技術をケーススタディとし、インターネットより教師データを収集することで人的収集コストを代替する可能性について検証した。本項目は、作成されたモデルの利用により、作物の病害虫診断能力の暗黙知の形式知化や、篤農知識の非継承問題という側面を解決するための研究課題としても位置づけられる。また、上述の手法を通じて有効なモデルが作成できたとしても、ニューラルネットワークが人間の意図する画像特徴量を学習したかは不明であり、所謂ニューラルネットワークのブラックボックス性がモデルの真正性や解釈性の困難さが新たな課題として想定された。そこで、次項(2)の研究テーマ B「病害虫診断モデルの特徴量可視化」において、病害虫画像の分類能力を獲得したネットワークをモデルケースとして、ネットワークの特徴量解釈性について網羅的な検証を行った。

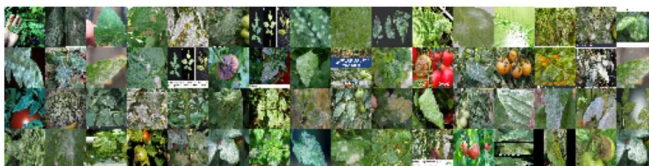
さらに、一般的な教師あり学習においては、収集したデータだけでは意味をなさず、画像と対となるアノテーション情報が必要となる。画像のアノテーション作成、いわゆるラベル付け作業の人的コストの高さや、ラベルの正確性を担保するクオリティコントロールの難しさが当該研究分野においても課題として挙げられていた。したがって次項(2)の研究テーマ C「教師データラベリングコスト低減技術の開発」において、種子計測をケーススタディとし、機械学習モデルの性能を維持したままラベル付け作業を低減化する手法に取り組んだ。

(2) 詳細

研究テーマ A「集合知を利用した病害虫診断技術の開発」

機械学習を活用した画像による病害虫診断技術のためには、病害虫画像を収集する必要が生じる。これまで国内外で数多くの研究プロジェクトが進められており、論文やデータセットの公開という形で報告されている。しかしながら、それらの多くは研究グループが有するフィールドや栽培現場で多くの人的資源と年月をかけて収集されたものである。さらには、発生頻度が稀である病害の種については教師データの量が不十分であるという問題がある。したがって申請者は、それらの課題を解決するため、インターネットより病害虫の画像を収集し、整備することによって現場での教師データ収集を代替できないか検証した。8作物種650病種についての画像をデータセット化することを目的とし、インターネットより病害画像を合計44万枚収集した。そのうち、訓練に不敵なノイズを除去した7クラスについて分類モデルを作成したところ、テストデータセットにおいて72.5%の正確さを達成した。このことは、インターネット経由のデータセットを用いても病害虫診断技術が作成可能であることを示唆した。このようなデータセットを用いることで、どのような病種であれ、病害虫診断モデルの作成コストが一層低下することが強く期待される。現在、データセットのさらなるクレンジングを進めており、論文、公開データベースとして公開する予定である(下図)。

Tomato Powdery Mildew



Rice Blast



図1 病害虫画像データセットのなかから2ラベルに属する画像をランダム抽出した様子

研究テーマ B「病害虫診断モデルの特徴量可視化」

近年、深層学習を活用したニューラルネットワークモデルの学習により、病害虫診断モデルを訓練する事例が報告されるようになってきた。しかしながら、ネットワーク自体が学習した特徴量の解釈までを行った事例は少なく、ブラックボックス性を残したままであることが多い。したがって本研究では、病害虫診断能力を獲得したニューラルネットワークを用いて、これまでに報告されている可視化技術の網羅的適用による解釈性の探索を行った。その結果、病斑のサイズや特徴に応じてその様態がネットワークに学習されていることが判明した(下図)。さらには、既存の一般物体と比べた病斑の特徴の「相対的非複雑性」により、既存技術の適用限界など明らかとなり、ネットワークの浅い層を活用するよう可視化技術を改変することで解釈性が増すことなどを見出した。それだけでなく、可視化技術を活用したデータセットバイアスの発見や、推論に不要なネットワーク層を取り除くという新たなパラメータ削減の方法などをも見出すことができ、病害虫診断モデルを作成するにあたって極めて有益な基盤知識を整備することができた。本研究成果は、国際科学誌 *Plant Phenomics* に 2019 年 3 月に掲載された。

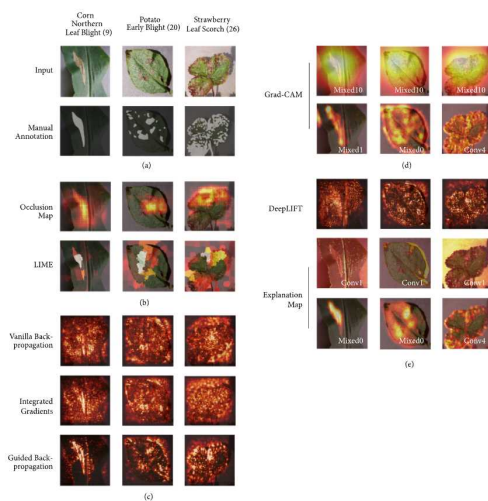


図2 多様な深層学習特徴量可視化技術を作物病害画像に適用した様子

研究テーマ C「教師データラベリングコスト低減技術の開発」

現在では、多様な画像解析を実現するための深層学習技術が整備されつつあり、(機械学習の観点において)綺麗な教師データを十分量整備することができれば、技術開発の目処が立つ事例が増えてきた。しかしながら、教師データのアノテーション、所謂ラベリングの労力コストおよび要する時間は未だ解決されていない。例えば植物の種子の形状を計測するためには、輪郭をトレースした教師データを用意する必要があるが、品種や種そのものが異なる場合、その都度新たな教師データを用意することが求められる。そこで、少数の種子のサンプルのみから大量の教師データを自動生成し、機械学習(深層学習)モデルの効率的な訓練を行う手法を構築した。本手法は、ドメインランダム化と呼ばれるアプローチを応用しており、機械学習モデル開発の高速化に役立てるものである。提案手法では、形状の情報があらかじめ付与された少数の種子のサンプル画像を仮想空間上にランダムに配置し、合成画像を多数作成する(下図)。実際に合成データのみを用いて学習を行ったところ、手作業で測定するときと同程度の精度で、画像からの種子の検出と形状の抽出が可能であることを示した。さらに、同様の手法は、イネ、コムギ、エンバク、レタスなど数多くの品種の種子の測定に容易に応用が可能であることも、本研究で明らかとなった。この結果は、作物種を問わず、多様な種子の自動測定が実現可能であることを強く示唆している。本研究成果は、国際科学誌 *Communications Biology* に 2020 年 4 月に掲載された。

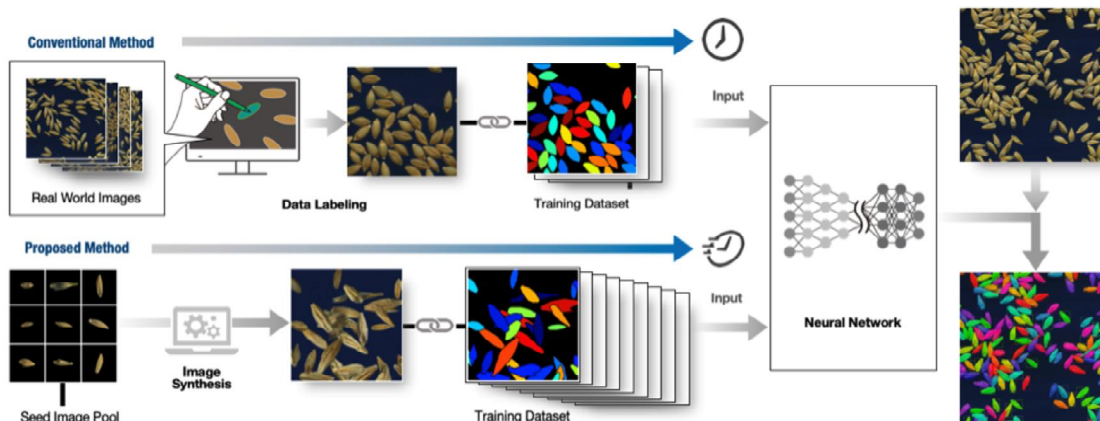


図3 一般的な学習データ作成(上段)と、提案システムの流れ(下段)

3. 今後の展開

本研究課題を通じて分野で共有すべき価値の高い病害虫画像データセットが整備されつつある。今後は、国内サーバーへのホスティングに向けてデータセットを成形し、産学問問わずどのようなアイデンティティでもデータセットにアクセス可能な状態へと昇華させることを目指す。さらには、データセットの逐次更新を可能とするよう、コミュニティベースでの維持管理システムの構築を目指す。それらを利用することで、世界中の誰でも容易に病害虫診断モデルを構築可能とする「作物病害虫版 ImageNet」を整備する。

4. 自己評価

- 研究目的の達成状況

病害虫診断モデルを作成するための基盤データセット作りに関しては、研究開始時点の目標より対象種を拡大し、巨大データベースの整備の見通しを立てることができた。特に、インターネット由来のデータソースを利用した深層学習モデルの性能も実証することができた。さらにまた、病害虫診断をケーススタディとしたブラックボックス問題に対する解釈の試みも論文としてまとめることができ、モデルパラメーターの新たな削減方法の発見など予期せぬ成果にも結びついた。しかしながら、一部成果の論文化やデータセットの公開に関し、新型コロナウイルスによる研究成果取りまとめにおける遅延が生じたため、研究課題期間終了後も、それらについて優先的に取り組んでゆく。一方、植物フェノタイピング技術をすすめるにあたって課題として見出した教師データラベル付コストについても、合成画像の利用による労力低減の手法提案などについても論文化することができた。あわせて、研究目標は概ね達成できたと思われる。

- 研究の進め方(研究実施体制及び研究費執行状況)

本研究課題を通じて、技術補佐員1名を専属で雇用し、データ整理やアノテーション作業の補助業務を担当した。また、申請者が情報科学分野の研究を単独で遂行できるよう、計算サーバーなど各種機材を揃え研究に専念することができた。さきがけの専任研究員としてエフォートの大部分を本研究課題の遂行のために使用し、適切な研究活動を実施したと自己評価する。

- 研究成果の科学技術及び社会・経済への波及効果

インターネットソース由来の病害虫画像データベースに関しては本研究成果に匹敵する規模のものは現時点で存在しておらず、公開することで、日本のみならず世界中の研究者が利用できるインパクトの高いものとなる。また、一連の研究成果に関しては論文化を行い、例えば、病害虫診断におけるブラックボックス性について網羅的解析を行った Toda et al., 2019 Plant Phenomics については、現在までに37件の引用がされ、さらには当該誌の2019年に最も読まれた論文として選出され、当該分野におけるプレゼンスを増すことができた。

- 領域独自の評価項目

当研究領域の目標設定として、「現在、農学・植物科学と情報科学の融合分野は人材が極めて限定的」であるため、「農学・植物科学研究者は情報科学を、情報科学研究者は農学・植物科学を 研究期間内に自ら積極的に学び、異分野の知見を取り入れつつ自身の研究を深化」することが求められ、また、「セミナーでの講演や成果公表等、積極的なアウトリーチ活動への協力」も推奨されていた。

申請者は植物科学分野を専攻とし、独学で情報科学的手法に挑戦し始めた研究者であったが、本研究期間を通じて、植物フェノタイピングという新たな研究分野において精力的に活動し、研究成果をまとめることができたと考えている。

それだけでなく、セミナーやハンズオンの技術的な教師活動含め、多数のアウトリーチを行

ってきた。さらには、植物科学・農学と情報科学の学際領域に関する全ての情報共有のため、SLACK のワークスペースを 2018 年に立ち上げた。現在では、メンバー数 200 名近くとなり、そのワークスペースを起点とした新たなコラボレーションが生まれている。一方、植物フェノタイピング分野の研究者のネットワーク構築のため、本領域の関係者を含む研究者らと Japan Plant Phenotyping Network を幹事として立ち上げに寄与し、アイデアソンを始めとした交流活動を推進してきた。現在では、学会化への昇華の声も聞こえてきている。このように、申請者は当該分野のミックスアップ活動にも一定の貢献ができたと自己評価する。引き続き当該研究分野を様々な方面から盛り上げていくことで、日本におけるサイエンスの一層の推進に貢献したい。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

- Toda Y, Okura F, Ito J, Okada S, Kinoshita T, Tsuji H and Saisho D. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. Communications Biology. 2019: 3,173.
- Toda Y and Okura F. How Convolutional Neural Networks Diagnose Plant Disease. Plant Phenomics. 2019: Article ID 9237136, 14.
- 大倉史生、水谷未耶、野下浩司、戸田陽介 植物科学の「人工知能」との関わり方を考える bsj-review 2019.

研究期間累積件数: 4件

1. Toda Y, Okura F, Ito J, Okada S, Kinoshita T, Tsuji H and Saisho D. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. Communications Biology. 2019: 3,173.

画像解析におけるドメインランダム化の手法を応用し、1 枚の画像から一度に数百粒の種子の形を自動かつ高精度に測定する事を可能にする機械学習訓練システムを提唱しました。種子の形は、作物の収量や品質と密接に関連のある重要な農業形質であり、画像から種子の形状を自動測定して評価することは、効果的な作物品種の創出に欠かせない技術です。本報告では、大麦種子画像を仮想空間上にランダムに合成することで、教師データに使われる画像を大量に自動生成し、機械学習(深層学習)モデルの効率的な訓練を行う手法を構築しました。実際に合成データのみを用いて学習を行ったところ、手作業で測定するときと同程度の精度で、画像からの種子の検出と形状の抽出が可能であることを示しました。

2. Toda Y and Okura F. How Convolutional Neural Networks Diagnose Plant Disease. Plant Phenomics. 2019: Article ID 9237136, 14.

植物科学・農学分野での深層学習を利活用するにあたり障壁となる事柄として、ニューラルネットワークがどのような画像特徴量を学習したか不明であるという、いわゆるブラックボックス

ス問題がある。本報告では、病害虫診断能力を獲得した分類問題を解くニューラルネットワークを対象とし、多様な可視化技術を適用することによってどのような特徴量が形成されているかを網羅的に調査した。病斑に最適な既存可視化技術の改変を要することを見出しただけでなく、特徴量可視化の利用によるデータセットバイアスの発見や、新たなパラメーター削減手法の提案を行うことができた。

(2)特許出願

特になし

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

・学会発表

- “A unified framework for image-based plant phenotyping under controlled growth environment: From image acquisition to phenotype.” 6th International Plant Phenotyping Symposium. 2019年10月23日
- 「人工知能で切り開く植物科学の近未来」日本植物学会主催シンポジウム 2019年9月17日

・プレスリリース

- 「合成画像のAI学習で種子の形態評価を効率化 ～植物科学・農学分野における機械学習モデルの開発の高速化に期待～ 名古屋大学プレスリリース 2020年4月16日