

深層学習を用いたスギ花粉の発芽率測定の効果化

1. はじめに

林木育種センターで進めている林木ジーンバンク事業では、有用な育種素材として林木育種に貢献するもの、ニーズが高く早期に新たな需要が見込まれるもの、生息域内の保存のみでは種または集団の絶滅のリスクが高いものを対象として「生物資源」として成体や種子、花粉での保存を行っています。収集した花粉は、発芽検定により発芽率を測定し、その結果をデータベースに記録し、同時に含水率を調整した後、-80℃で長期保存し、将来の利用に備えます。花粉の発芽検定では、発芽培地に花粉を蒔き、発芽に適した条件に置いた後、顕微鏡下で直接または取得した画像上で、発芽した花粉の数、発芽しなかった花粉の数を数え、その割合を発芽率として求めています。花粉の発芽の有無は人の目で判断しており、時間がかかる作業です。ジーンバンク事業を継続的に運営していくためには、このような作業を効率化していく必要があります。

最近では深層学習（ディープラーニング）が発展し、様々な分野で利用されるようになっていきます。深層学習は、人工知能（AI）の一つである機械学習に含まれる手法の一つです。従来の機械学習においては、例えば画像の分類を行う場合、画像の着目すべき場所（これを特徴量と呼びます）を事前に決めて、これを数値化し

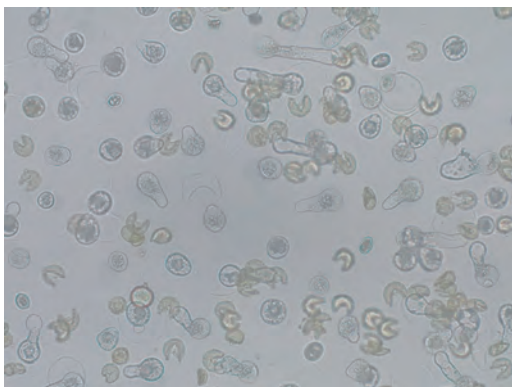


図1 発芽検定に用いた画像の一例

ておく必要があります。例えば花粉の発芽であれば、花粉の画像から花粉の形や模様などを数値化する必要があります。この特徴量の数値化が多くの場合困難で、画像に従来の機械学習の手法を適用するのは簡単ではありませんでした。しかし、深層学習においては特徴量の抽出まで自動で行うことができるため、画像の分類といった従来は簡単にはできなかった作業が可能となります。特に、人の目でみれば判断できる画像の分類には有効であり、花粉の発芽試験にも利用可能であると考え、その試行を行いました。本記事は武津ら（2023、関東森林研究）で報告した内容を紹介したのになります。

2. 材料と方法

深層学習による画像認識技術にはいくつかの種類があります。最も簡単なのは画像分類で、例えば花粉の画像が与えられた場合に発芽した花粉なのか、発芽していない花粉なのかを分類するために用います。しかし、実際には花粉の画像は一つずつ撮影するわけではなく、図1のように複数の花粉が含まれる画像を取得します。この画像内で複数花粉について、発芽しているか発芽していないかを認識する必要があります。このような技術を物体検出と呼びます。物体検出を行うことのできるプログラムは多く公開されています。今回は、Yolov5 (<https://github.com/ultralytics/yolov5>) というプログラムを用いました。

深層学習により物体検出を行うためには、正解となるデータを使用してモデルを作成します。そのためには、目的とする物体をあらかじめラベリングした注釈データを準備する必要があります。これは、実際に画像のどの場所に目的とする物体（今回の場合は発芽した花粉と発芽していない花粉）があるかを手作業で印をつけていく作業となります。今回は、東北育種基本区でそれぞれ異なる遺伝子型をもつ42個体

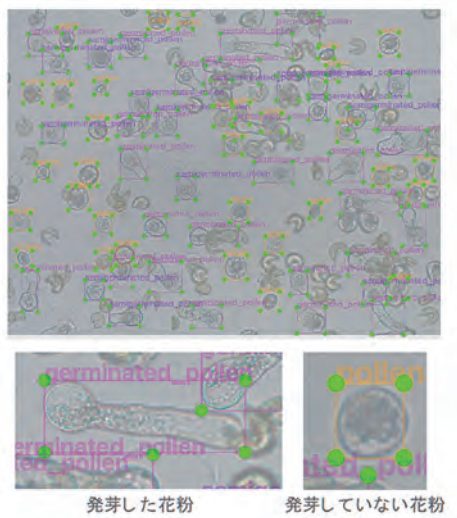


図2 注釈データ作成の様子

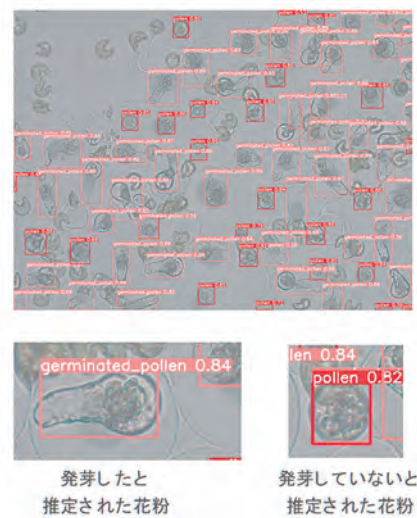


図3 作成されたモデルによる検出の例

から収集されたスギ花粉42サンプルを、通常の方法で発芽培地上で発芽させ、顕微鏡下で撮影して画像を取得しました。124枚の画像から画像から発芽した花粉を合計1,346箇所、発芽していない花粉を合計4,361箇所ラベリングしました。ラベリング作業には、ソフトウェアLabelImg (Tzutalin 2015) を用いました(図2)。

注釈データを学習データセット、検証データセット、テストデータセットに分割し、学習データセットおよび検証データセットを用いてYolov5によりモデルを作成しました。作成されたモデルをテストデータセットに適用し、目で実測した発芽率と予測された発芽率とを比較しその精度を検証しました。

3. 結果

作成されたモデルをテストデータセットの画像に適用し、物体検出を行った例を図3に示しました。1枚の画像内の多数の場所に物体が検出されています。これらの検出された物体について、画像上の位置と推定されたラベル(今回は発芽した花粉と発芽しなかった花粉)がデータとして得られます。得られたデータから発芽率(%)を計算した結果を図4に示しました。誤差はあるものの、ある程度の精度で発芽率が予測されていることがわかります。予測の精度を示す指標の一つである二乗平均平方根誤差を

発芽率(%)について計算したところ、約5ポイントとなり、平均発芽率に対する誤差率約11%で発芽率が推定されることがわかりました。また、細かくみていくと、花粉が重なっている部分ではうまく推定されないなど、モデルの改善の余地があることもわかりました。

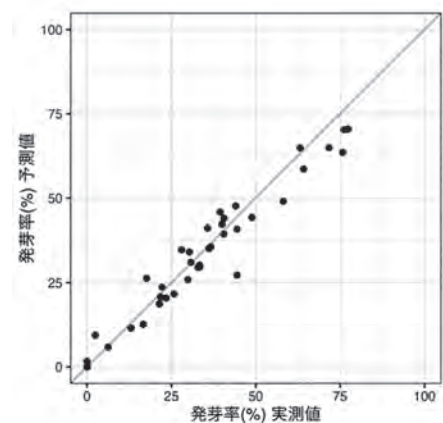


図4 発芽率の実測値と予測値の関係

4. おわりに

本記事では、ジーンバンク事業の効率化に向けて、発芽検定作業の一部について深層学習を用いた自動化の試行について紹介しました。実際に事業に利用していくためには、モデルの改良・他樹種への展開に加えて、顕微鏡下でリアルタイム推定により画像を取得する手間を省くといった工夫が今後必要となります。

(遺伝資源部 探索収集課 武津英太郎)