

植物生育記録の自動化に向けた Keypoint 検出の検討

平原健太郎¹ 加納多佳留² 海野博也² 北川寛人²
工藤史堯³ 永井彰³ 峰野博史⁴

概要：栽培と育種の両面において、作物の生育状態を把握するための定期的な生育調査が重要である。定期的な生育調査では、植物の成長速度、植物のストレス状態を定量化することで現在の植物の状態を想像することができる。また、植物の状態を定量化した過去データを参照できるようになれば、現在行うべき農作業の判断に役立ち、より品質の高い作物を栽培に繋がる。しかし、節間の伸長の計測や花数のカウントといった植物の生育調査は、手作業で行うにはとても手間のかかる作業である。近年の農業従事者の高齢化、新規就農者の減少による労働力不足もあり、より重要度の高い収穫・出荷といった農作業に追われ、手間のかかる定期的な生育調査は十分実施できていないと言えない。本研究では、様々な理想的な生育調査の中でも「長さ」を測る項目に着目し、Keypoint 検出を用いて自動的に距離測定を行う技術の確立を目指す。まずはモデル植物として一般的であるトマトの長段ハウス栽培を対象とし、ハウス内の栽培ベッド間を自走する吊り下げ型ロボットで撮影された動画に対し、Keypoint 検出によって茎や枝の節を見つける手法を検証した。Keypoint 検出によって画像中から特徴点を見つけられるため、植物ならではの特徴点の結びつきを定義することで、植物の茎や枝の節間距離の自動推定を実現する。植物に対する茎や枝の節といった Keypoint を定義し基礎評価を実施した結果、植物の画像中から茎や枝の節部分の検出を高精度に実現できる可能性を確認した。今後、植物ならではの Keypoint 検出精度向上に向けた検討を深め、距離推定の検証を進めていく。

キーワード：Keypoint 検出, Detectron2, 節間距離, 長さ推定, 植物生育記録

1. はじめに

近年、栽培と育種の両面において、作物の生育状態を的確に把握するための技術開発の重要性が高まっている。作物の生育状態を定期的に調査することで、品質や状態を推定するのに関係する指標を定量化し、時系列データとして蓄積できるようになる。定量化し時系列データとして蓄積された過去の生育状態のデータを比較して栽培評価することで、現在の作物の品質や状態に応じて適切な栽培や環境制御に繋げられ、より品質の高い作物を栽培するための試行錯誤が可能となる。このような作物の生育状態を把握するために行われる生育調査には、株ごとの節間の伸長計測、果実や花数のカウントなどがあるが、人手で実施するには手間のかかる作業で、多くの時間と労力を必要とするものが多い。さらに、昨今の国内農業分野における高齢化の進行と新規就農者の伸び悩みによって、農業従事者の高齢化と減少も進んでいる。そのため、定期的な生育調査は、品質の高い作物の栽培に重要な作業であるにも関わらず、より重要度の高い収穫・出荷解いた農作業に追われ、手間のかかる定期的な生育調査は十分実施できていないと言えない。そこで、植物の様々な栽培ステージにおいて、作物の生育状態を定量的に評価する技術が期待される。

本研究では、モデル植物として一般的であるトマトの長段ハウス栽培を対象とし、生育調査業務に含まれる「節間距離」「葉長」といった長さを測る項目について、Keypoint

検出技術を用いて、画像から自動で長さ推定を行う手法を検討する。Keypoint 検出は、これまで主に人の姿勢検出やハンドトラッキング等に活用されてきた画像解析技術である。トマトといった植物に対して、茎や枝の節といった Keypoint を定義し、植物に対する Keypoint 検出の精度を検証する。ここで、Keypoint 検出技術を選択した理由は2つある。1 つ目はピクセル間の距離を容易に算出可能であることで、2 つ目は画像から認識できない隠れた部分を検出できることである。主に人を対象に適用されてきた Keypoint 検出技術を、植物に対して適用することで有効性の検証及び精度評価を行う。また、画像処理によるデータセットの増幅や既存の学習済みモデルを用いた際の精度を検証し、植物に対する Keypoint 検出の有効性を確認する。

以降、第 2 章では関連研究について述べる。第 3 章では提案手法ならびに植物に対する Keypoint の定義について説明し、第 4 章で植物の Keypoint 検出精度について実験結果を整理し、第 5 章で本稿をまとめる。

2. 関連研究

2.1 画像解析を用いた節間距離測定

トマトの節間距離の推定を行う手法として、画像処理とルールベースの教師あり機械学習を用いた手法[1]がある。この手法では、節検出、節位推定、節間距離推定の3段階の処理によって節間距離を算出している。特に、節検出の

1 静岡大学情報学部
Faculty of Informatics, Shizuoka University
2 (株) Plant Data
PLANT DATA CO., LTD.
3 NTT 社会情報研究所
NTT Social Informatics Laboratories

4 静岡大学大学院情報領域/グリーン科学技術研究所
College of Informatics, Academic Institute, Shizuoka University /
Research Institute of Green Science and Technology

ために植物画像の全画素に対して「葉」,「茎」,「背景」のラベリングを手作業で行い, 決定木ベースのセグメンテーションアルゴリズム DTSM[2]を用いて, 画素単位での画像分割によって茎領域を抽出し, 節の検出を行っている. さらに節位推定では, Bag of Features[3]と呼ばれる一般物体認識手法を用いて誤推定の節位置の削除を行う. このようなルールベースの教師あり機械学習によって, 節間距離の推定を行っている. しかし, 想定している画像は, トマトの苗が成長するにつれて葉の量が増加し, 茎や節が成長した葉で隠れてしまった場合, 適切に検出できないという課題がある.

2.2 Keypoint 検出

Keypoint 検出とは, 入力された画像から Keypoint と呼ばれる複数の座標点を出力し, 物体の全体像を Keypoint から推定する技術である. Keypoint 検出を利用した研究として, 図 1 (a)のように人間の骨格を推定する OpenPose[4]がある. これは, 画像から抽出された人体の関節を Keypoint とし, 予め Keypoint 同士を対応付けて繋げる事で骨格の推定を行い, 画像から人間の姿勢を推定している. ただし, OpenPose は人間の骨格に着目したものであり, 人間でない対象物には対応していない.

一方, 人間ではないものに着目した Keypoint 検出の研究も行われている. 例えば, 魚の魚体長の推定に Keypoint 検出を適用した研究[5]を図 1(b)に示す. この研究では, 魚体長の推定のために, 口, 胸びれ, 尾へ Keypoint を定義し, それらを結んだ線分を作成することで魚体長を推定している. つまり, 人間ではない物体に対しても新たに Keypoint を定義し, 長さを推定可能なことを示しており, 植物に対しても新たに Keypoint を定義し, 長さを推定可能であるといえる. さらに, 動画画像から車両の簡易的な 3D ポリゴンを再生成する CarFusion[6]を図 1 (c)に示す. CarFusion では, 車両の角やタイヤに Keypoint を定義することで, 街中の道路を走行する車両の検出を実現している. 特に Keypoint 検出によって, 車長・車幅・車高といった 3次元の情報から 2次元から算出可能としている. つまり, 入力画像からは見えない位置にある車体で隠れたタイヤに対しても, Keypoint 検出による特徴点の推定を行えることを示している.

以上のことから, 対象物に対し適切な Keypoint を定義し, その位置関係を対応付けて学習させることで, 隠れた位置にある Keypoint の検出だけでなく, 対象物の長さも適切に推定できると考える.

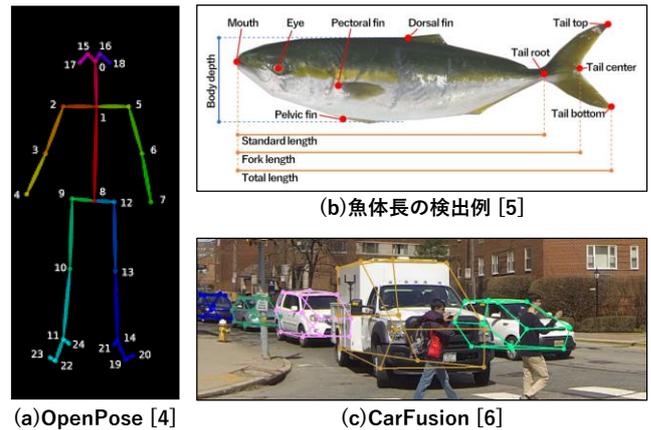


図 1 Keypoint 検出の例

3. 提案手法

3.1 想定する条件

本研究では, モデル植物として一般的であるトマトの長段ハウス栽培を対象とする. ハウス内の栽培ベッド間を自走する吊り下げ型ロボットで撮影された動画からデータセットを構築する. 苗が成長するにつれて葉量が増えた際に適切に検出できないというルールベースの手法[1]の課題を解決するため, Deep Learning ベースの Keypoint 検出を採用し, 検出対象となる茎や枝が隣の株の葉によって隠れている場合でも適切に Keypoint 検出が可能で, 節間距離や葉長の算出に利用できることを確認する. ここで, 骨格や剛体を対象とした Keypoint 検出の研究では, 対象とする物体が気温や天候, 撮影された日付によって大きく変化しない物体であって, 植物とは物体の性質が異なるといえる. 本研究では, 定期的な生育調査を想定した上で, トマトに対応可能な Keypoint の定義を行い, 理想的な画角の画像ではない環境下の画像でも, 正確に節間距離や葉長の推定を実現することを目指す.

3.2 概要

本研究では, 魚体長の推定にも採用された Mask Keypoint R-CNN[7]の後継である Detectron2[8]を用いて, 節間距離の自動計測を行う. Detectron2 は, Facebook AI Research が開発した次世代の物体検出・セグメンテーションアルゴリズムのフレームワークであり, Facebook AI Research がこれまでに研究開発してきた Detectron と Mask R-CNN のアルゴリズムを継承したフレームワークである. Detectron2 では, 先行の Detectron や Mask R-CNN と比べ, 学習速度と検出精度の向上を両立している.

本研究では, 公開されている Detectron2 の学習済みモデルを利用してファインチューニングを行い, 植物用に定義した独自の Keypoint 検出を試みる. 学習済みモデルは, COCO の人体骨格検知用のデータセットを用いてトレーニング・評価されたモデルを採用する. COCO データセット

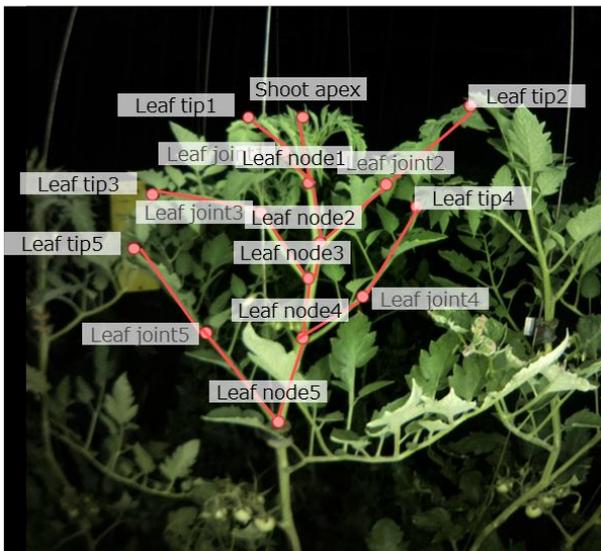


図 2 トマトの茎と節の Keypoint 定義

は、トレーニング画像約 20 万枚以上の 25 万人分の人物に対して、Keypoint アノテーションが行われている大規模データセットである。学習済みモデルは、人体骨格検知用のモデルではあるが、本研究で想定する植物の茎や枝の節を独自アノテーションした大規模なデータセットを用意することは困難である。そのため、少量の独自アノテーションデータセットで学習できるように、人体骨格検知用の学習済みモデルをファインチューニングすることで高精度に検出可能か検証することとした。

以下に、本研究の進め方を示す。

STEP1. Keypoint 検出の有効性検討

トマトの節間距離の推定を対象とした Keypoint の定義を行う。定義した Keypoint に従って、動画から切り出した画像へアノテーションを行ってデータセットを作成する。データセットを学習用とテスト用に分け、学習用データセットを用いて Detectron2 の学習済みモデルへ Keypoint の学習を行う。テスト用データセットを用いて、Keypoint 検出の有効性を評価する。

STEP2. 見えない部分に対する節の検出の検討

データセットの追加・改善や前処理を加えることで、画像には映らない隠れた部分にある節の検出が行えるようモデルの精度向上を図る。ただし、STEP1 同様に、まずは節間距離の推定を対象とする。

STEP3. 葉長の推定に対応

STEP2 までの結果を通して、葉長推定のための Keypoint を定義し、検出が可能であるか検証する。

STEP4. 動画へのリアルタイム推定の検討

入力画像の解像度の調整や、モデルの改良等を通じて高速に Keypoint を検出し、距離推定可能か検証する。



(a)主茎に生える側枝

(b)葉に対するアノテーション定義

図 3 葉の定義

表 1 データセットの概要

取得日	2020年8月, 2021年5月, 2021年6月, 2021年7月のうちに撮影 (計10日間分)
レンズ画角	165度 (魚眼)
枚数	計568枚
画像サイズ	幅960, 高さ1280
ファイル形式	JPG
ファイルサイズ	平均220KB (128KB~350KB)

3.3 節間距離・葉長推定のための Keypoint 定義

Keypoint 検出は主に人間に対して利用され、その定義は人間の各関節に定義されることが多い。その他にも、Deep Learning を用いた Keypoint 検出は様々な分野で応用されており、人間以外の物体にも定義が行われている。しかし、植物に対する Deep Learning を用いた Keypoint 検出を行った事例はこれまでに見当たらない。そのため、本研究では共同研究者との議論を経て、節間距離や葉長推定のための Keypoint を新たに定義することとした。図 2 に、定義したトマトの茎と節に対する Keypoint を示す。

ここで、本研究では図 3 に示すように、主茎に対して分岐する側枝の全体を指すものを「葉」と定義する。一般的な「葉」の定義と異なるが、今回想定するトマトの生育調査業務において、側枝の全体を「葉長」として計測が重要だからである。

4. 評価実験

本稿では、前述の STEP2 までの進捗状況について報告する。まずは、節間距離の推定に必要な Keypoint 定義に従ってアノテーションされたデータのみを用いて実験を行い、葉長の検出に必要な Keypoint は未設定である。つまり、節間推定のための Keypoint は主茎部分の 6 点となる。

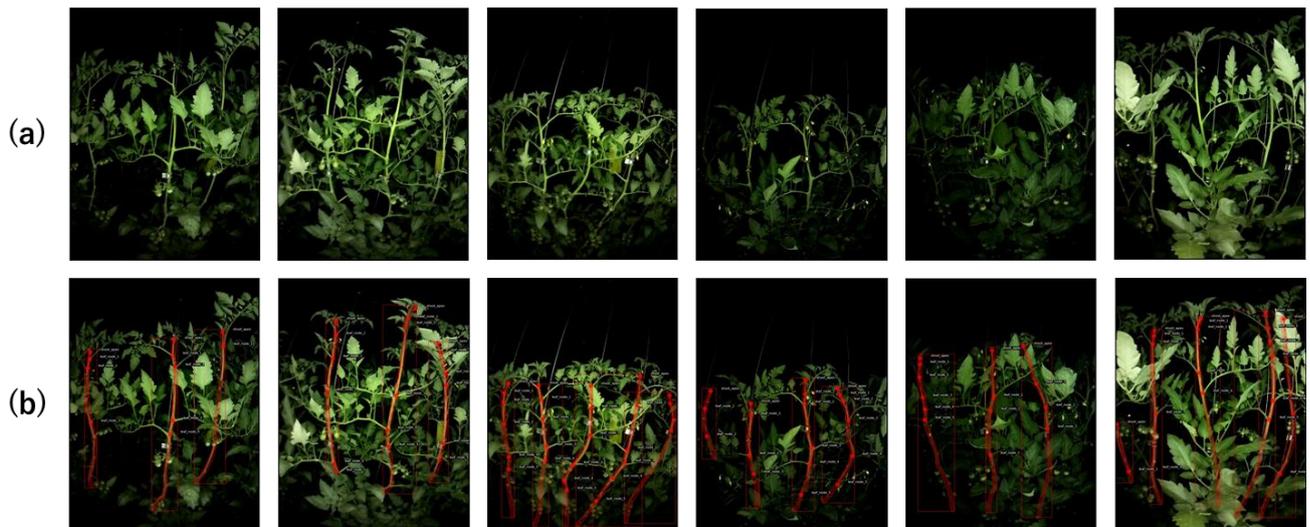


図 5 データセットのサンプル

(a) 動画から切り出した元画像 (b)アノテーションデータを描画した画像

4.1 データセット

植物の茎や節の推定のために定義した Keypoint に従ってアノテーションを行う。データセットは動画データから画像データを切り出して利用した。元となる動画は、レールで吊り下げられたロボット撮影機材を利用して一定の高さで、トマトの長段ハウス栽培の様子を撮影した。動画を約 2m 間隔 (50 ピクセル間隔) で切り出して画像データセットを準備した。図 4 に、撮影された動画データの一例としてパノラマ化したものを示す。

表 1 に、準備した画像データセットの概要を示す。画像データセットを COCO-Annotator [9]を用いて手作業でセグメンテーション及び Keypoint のアノテーションを行った。最終的にアノテーションを行った画像は計 568 枚である。図 5 に、画像データセットのサンプルとアノテーションを行ったものを示す。

学習には、訓練データ、検証データ、評価データごとに 8 : 1 : 1 にデータセットを分割して利用した。

4.1.1 実験方法

作成したデータセットについて節間距離推定のため、節の検出を Detectron2 の Keypoint 検出を用いて実験を行った。本稿では、「画像増幅による精度比較」「学習済みモデルによる精度比較」を行う。

Detectron2 での学習には、バッチサイズ 128、学習率 0.0004 とし、それぞれのデータセットで 1000[iter]の学習を行った。また、推論の信頼度の閾値を 0.80 する。

4.1.2 評価方法

Keypoint の推論結果における評価には、COCO データセットの評価に用いられる object keypoint similarity (OKS)

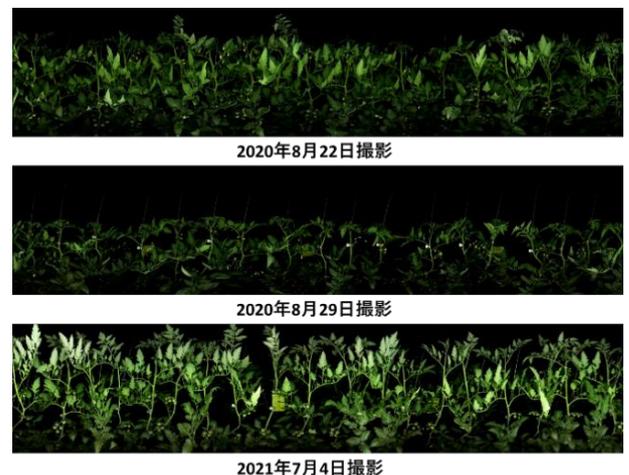


図 4 取得される動画データのパノラマ画像

$$OKS = \frac{\sum_i \exp\left(\frac{-d_i^2}{2s^2k_i^2}\right) \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}$$

d_i	: i 番目のKeypointに対応する推論Keypointと正解Keypointのユークリッド距離
s	: 正解BBoxで定められた物体の矩形領域面積
k_i	: Keypointごとに設定される定数
v_i	: Keypointのvisibilityラベル(0, 1, 2)
$\delta(v_i > 0)$: $v_i > 0$ ならば1, それ以外なら0となる関数

図 6 OKS の定義

[10]を利用した。OKS は、図 6 のように定義される。OKS は正解座標と推論座標の類似度を表現する評価指標で、値が大きいかほど類似度が大きくなり、推定精度が高いといえる。

図 6 の OKS の数式に定義される変数 k_i は Keypoint ごとに設定される定数である。 k_i は 0~1 の定数であり、 k_i の値が大きいかほど Keypoint アノテーションが実際の座標よ

表 2 画像増幅の比較結果

画像増幅手法	枚数 [枚]		keypoints_AP	keypoints_AP50	keypoints_AP75	keypoints_APm	keypoints_API
	学習	テスト					
オリジナル	454	57	2.13(-0.79~+0.67)	8.77(-1.99~+2.22)	0.47(-0.21~+0.3)	3.74(-2.26~+2.5)	2.26(-0.9~+0.7)
NOP(何もしない)	4994	57	1.32(-0.48~+0.3)	6.4(-2.29~+1.84)	0.12(-0.05~+0.05)	3.03(-2.53~+3.01)	1.41(-0.52~+0.31)
DP (ドロップアウト)	4994	57	1.25(-0.49~+0.49)	6.39(-1.98~+2.21)	0.13(-0.07~+0.09)	2.5(-1.6~+3.64)	1.33(-0.53~+0.53)
LR (左右反転)	908	57	2.04(-0.48~+0.65)	7.86(-1.53~+1.53)	0.45(-0.32~+0.27)	4.87(-3.88~+4.44)	2.16(-0.47~+0.72)
AGN (ガウシアンノイズ)	4994	57	1.43(-0.73~+1.08)	6.93(-3.68~+4.2)	0.16(-0.15~+0.24)	3.37(-0.59~+0.99)	1.53(-0.83~+1.16)
AC (自動コントラスト調節)	4994	57	1.65(-0.56~+0.61)	7.09(-1.7~+1.76)	0.36(-0.2~+0.32)	1.6(-1.6~+0.97)	1.77(-0.6~+0.69)
AF (幾何学変形)	4994	57	1.55(-0.4~+0.2)	6.72(-1.16~+0.69)	0.26(-0.07~+0.07)	4.81(-2.83~+4.69)	1.61(-0.38~+0.24)
DP+LR+AGN	4994	57	1.41(-0.41~+0.27)	6.4(-2~+1.15)	0.2(-0.09~+0.13)	3.74(-1.86~+1.5)	1.52(-0.44~+0.3)
DP+LR+AC	4994	57	0.93(-0.16~+0.16)	4.35(-1.22~+1.01)	0.09(-0.04~+0.05)	1.43(-0.53~+1.35)	0.98(-0.22~+0.18)
DP+LR+AF	4994	57	1.68(-0.44~+0.85)	7.66(-1.65~+2.3)	0.31(-0.14~+0.33)	3.98(-2.99~+1.56)	1.77(-0.5~+0.95)
DP+AGN+AC	4994	57	0.9(-0.19~+0.13)	3.99(-0.41~+0.59)	0.25(-0.06~+0.16)	2.57(-1.09~+1.88)	0.94(-0.17~+0.18)
DP+AGN+AF	4994	57	1.12(-0.23~+0.25)	5.33(-0.78~+1.46)	0.12(-0.07~+0.07)	2.42(-1.43~+1.25)	1.19(-0.26~+0.28)
DP+AC+AF	4994	57	0.8(-0.36~+0.39)	3.86(-2.01~+1.7)	0.08(-0.03~+0.06)	1.5(-1.01~+1.27)	0.85(-0.38~+0.42)
LR+AGN+AC	4994	57	1.27(-0.34~+0.37)	5.91(-1.62~+2.31)	0.16(-0.07~+0.13)	4(-1.43~+2.04)	1.35(-0.37~+0.39)
LR+AGN+AF	4994	57	1.8(-0.46~+0.37)	8.38(-2.22~+1.19)	0.14(-0.12~+0.15)	4.04(-2.26~+1.6)	1.9(-0.51~+0.44)
LR+AC+AF	4994	57	1.76(-0.36~+0.26)	7.63(-1.34~+2.19)	0.24(-0.16~+0.22)	3.41(-2.02~+2.24)	1.83(-0.48~+0.27)
AGN+AC+AF	4994	57	1.77(-0.63~+0.52)	7.57(-2.32~+1.62)	0.43(-0.25~+0.28)	3.05(-1.37~+2.99)	1.92(-0.68~+0.58)
DP+LR+AGN+AC+AF	4994	57	1.34(-0.27~+0.69)	5.92(-1.2~+2.23)	0.14(-0.06~+0.05)	2.06(-1.56~+1.7)	1.46(-0.29~+0.73)

りどれだけブレやすいかを示す。COCO データセットでは、約 5000 枚の人間の骨格に対するアノテーション付き画像によって k_i が算出され設定されるため、植物に対する評価指標に OKS を利用するには k_i を再設定する必要がある。しかし、本研究で用意したアノテーションデータセットは約 500 枚程度であり、 k_i のチューニングには十分な量のデータセットではないと判断し、本研究において k_i は一定値 0.1 とした。そのため、一般的な OKS の評価基準ではなく、得られた OKS の値から相対的にモデルの精度を比較する。

OKS は COCO のデータセットにおいて、一般的な物体検出タスクで用いる Intersection over Union (IoU) と同じ役割として用い、OKS の閾値を変化しながら Average Precision (AP) を算出する。本研究の評価指標は、COCO データセット API の物体検出用 Metrics を利用して算出した AP を利用する。算出する AP を以下に示す。

- mAP^{kp} (AP)

OKS の閾値を 0.50 から 0.95 まで 0.05 ずつ変化させた際の平均 AP を示す。COCO データセットで Keypoint 評価を行う際の主要な評価基準となる。
- $AP_{OKS=50}$ (AP50)

OKS の閾値が 0.50 以上の場合の AP を示す。
- $AP_{OKS=75}$ (AP75)

OKS の閾値が 0.75 以上の場合の AP を示す。
- AP_{medium} (APm)

検出対象となるオブジェクトの大きさが 32^2 ピクセルから 96^2 ピクセルのものを対象とした AP を示す。
- AP_{large} (API)

検出の対象となるオブジェクトの大きさが 96^2 ピクセル以上のものを対象とした AP を示す。

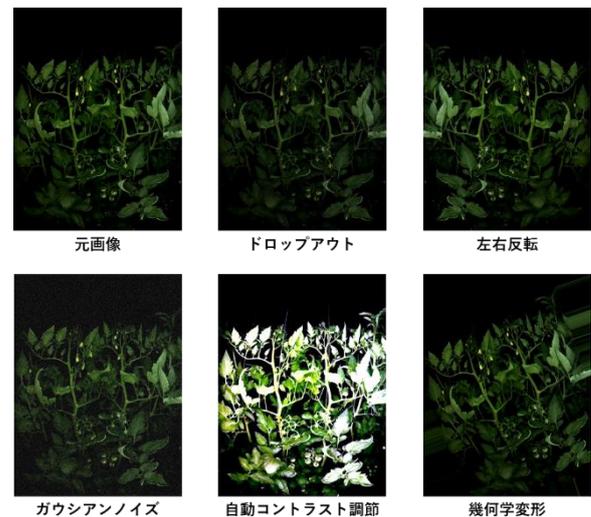


図 7 適用した画像増幅処理の例

4.2 画像増幅による精度

作成したデータセットを画像処理によって増幅させて Keypoint 推定精度を比較し、画像増幅手法の有効性について検証を行う。画像増幅は `imgaug[11]` を使って行い、画像データの加工に伴う Keypoint アノテーションデータの座標変換も行った。

本検証には「元データセットのみ (オリジナル)」「無加工 (NOP)」及び「ドロップアウト (DP)」「左右反転 (LR)」「ガウシアンノイズ (AGN)」「自動コントラスト調節 (AC)」「幾何学変形(回転, 平行移動, 拡大など) (AF)」の 5 種類の画像加工を行ったデータセットの組み合わせについて精度の比較を行った。

検証には、作成された元のデータセットの画像に対して

表 3 学習済みモデルの比較結果

学習済みモデル	keypoints_AP	keypoints_AP50	keypoints_AP75	keypoints_APm	keypoints_API
R_50_FPN_x1	1.08(-0.42~+0.65)	4.93(-1.98~+3.02)	0.14(-0.12~+0.19)	2(-1.5~+1.07)	1.15(-0.48~+0.74)
R_50_FPN_x3	1.76(-0.63~+0.72)	7.39(-2.76~+3.47)	0.34(-0.28~+0.19)	6.57(-1.72~+1.84)	1.83(-0.69~+0.81)
R_101_FPN_x3	1.4(-0.68~+0.57)	7.13(-3.28~+3.5)	0.15(-0.08~+0.12)	3.6(-1.72~+1.84)	1.5(-0.74~+0.59)
X_101_32x8d_FPN_x3	2.29(-0.79~+0.88)	10.2(-2.11~+2.92)	0.35(-0.18~+0.24)	4.75(-2.57~+4.85)	2.42(-0.84~+1.01)

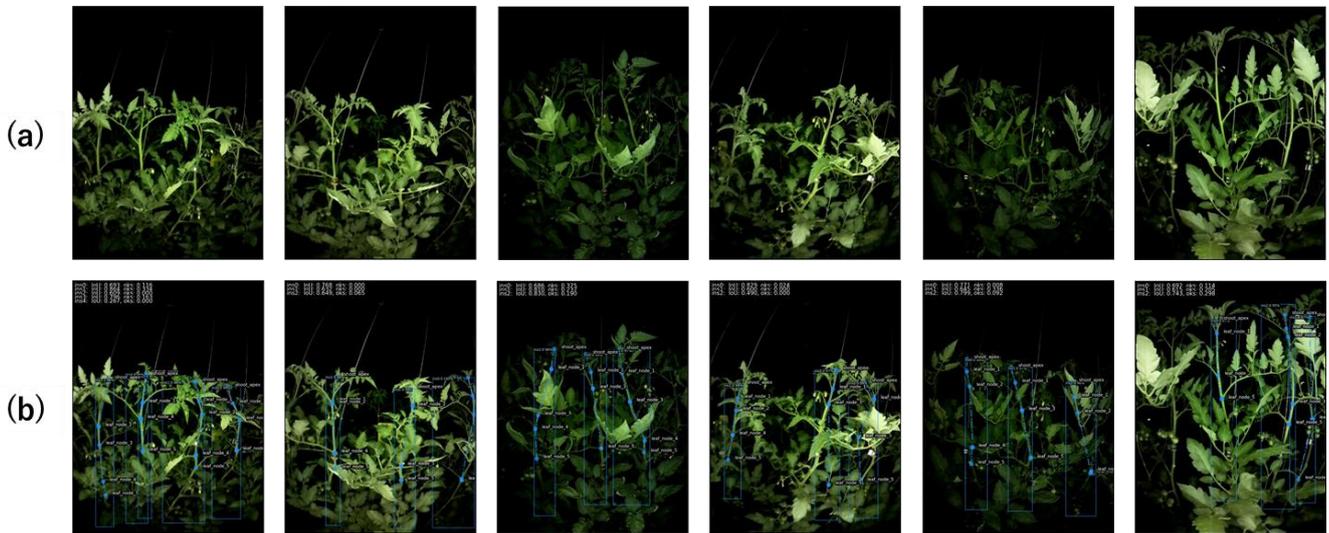


図 8 最も精度の良いモデルの推論結果例

(a)テスト画像 (b)Keypoint 検出結果

各画像加工を行い、画像数を 10 倍に増幅させて検証を行った。画像増幅の実施例を図 7 に示す。また、Detectron2 の学習済みモデルには「R_50_FPN_3x」を用いて学習を行った。モデルのファインチューニングを同じデータセットを用いて 5 試行し、各評価値の平均値を算出した結果を表 2 に示す。

表 2 の結果より、データ増幅を行っていないオリジナルデータが一番高い精度を示した。一般的に、画像増幅を行うことで学習の精度は向上が期待されるが、本実験では精度が低下した。学習のプログラム実装上に何らかの不具合が残っている可能性を考慮もあるが、画像増幅を行ったことによってモデルのオーバーフィッティングが発生していると考えられる。今回用意した植物のデータセットは、似たような画角の画像が多く、容易に区別できるようなオブジェクトではない。そのため、今回のような少ないアノテーションデータセットでは、画像増幅を行ったことにより類似した画像が大量に生成され、オーバーフィッティングが生じた可能性が懸念される。そのため、画像増幅によってモデルの精度向上を実現するためには、アノテーション済みデータセットを増加させたり、様々な画角の画像を学習させたりする手法が有効な可能性がある。

4.3 学習済みモデルによる精度比較

Model Zoo で公開されている Detectron2 の Keypoint 学習

済みモデルについて比較を行う。これらの学習済みモデルは 2019 年の 9~10 月に Detectron2 によって COCO のデータセットを用いて学習された大規模モデルであり、概要については次の通りである。

- R_50_FPN_x1
バックボーンのアーキテクチャに ResNet-50[12]と Feature Pyramid Network(FPN)[13]を使用したモデルであり、一般的な COCO データセットの学習スケジュールで学習された学習済みモデル
- R_50_FPN_x3
R_50_FPN_x1 と同様に、バックボーンのアーキテクチャに ResNet-50 と FPN を使用したモデルであり、一般的な COCO データセットの学習スケジュールの 3 倍で学習された学習済みモデル
- R_101_FPN_x3
バックボーンのアーキテクチャに ResNet-101 と FPN を使用したモデルであり、一般的な COCO データセットの学習スケジュールの 3 倍で学習された学習済みモデル

● X_101_32x8d_FPN_x3

バックボーンのアーキテクチャに ResNeXt-101-32x8d [14]と FPN を使用したモデルであり、一般的な COCO データセットの学習スケジュールの 3 倍で学習された学習済みモデル

以上 4 つの学習済みモデルを用いて植物の Keypoint 検出のためファインチューニングを行う。データセットには画像増幅を適用しないデータセットを用いて実施した。モデルのファインチューニングを同じデータセットを用いて 5 試行し、評価値の平均値を算出した結果を表 3 に示す。

表 3 の結果より「X_101_32x8d_FPN_x3」の学習済みモデルが最も高い精度を示した。表 3 の結果より、植物の Keypoint 検出において、「X_101_32x8d_FPN_x3」の学習済みモデルの使用が有効だとわかった。

4.4 Keypoint 検出結果

「X_101_32x8d_FPN_x3」の学習済みモデルを使って実際にテスト画像（画像増幅なし）に対して Keypoint 検出を行った。図 8 に実際に推定された Keypoint 検出の結果を示し、図 9 に失敗事例を示す。図 8 より、少量の学習データセットながらも概ね良好に節の Keypoint 検出ができてることが確認できる。一方で、図 9 のように Keypoint 検出に失敗した事例が見られた。図 9 (a) では、今回の学習データセットには含めていない側枝の部分に対して誤検出を行っている。図 9 (b) では、同じ節に対して 2 つの Keypoint が検出され、図 9 (c) では、leaf_node4 が検出されていない。生育状態の記録においてある程度のズレは許容できるが、誤検出によって大幅に値が異なってしまうことは極力避けるべきである。誤検出をなくすためには、より多くの画角を含めた大量のアノテーションデータが必要になると考える。その他に、動画の前後フレームから TTA[15]を用いることで誤検出の防止を行うことが有効であるのではないかと考える。

5. おわりに

本研究では、トマト栽培において生育調査業務に含まれる「節間距離」「葉長」の計測のような長さを測る項目について、Keypoint 検出という観点から画像より自動で長さ推定を行う手法を検討した。提案手法では、多段栽培されるトマトの様子を自動走行する車両で撮影し、Detectron2 を用いた Keypoint 検出によって「節間距離」「葉長」を推定する。植物（トマト）に対して Keypoint の定義を行い、実際に手動でアノテーションを行って約 500 枚のデータセットの準備を行った。本稿では、まず「節間距離」推定のため節の Keypoint 検出に焦点を絞って、「画像増幅の有効性」「学習済みモデルの比較」という観点で実験を行った。実

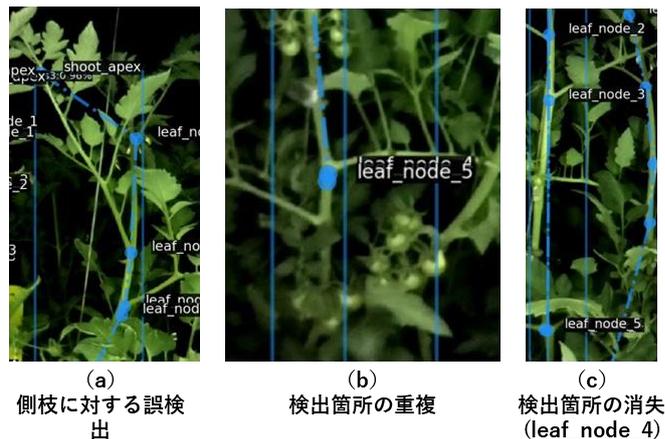


図 9 検出失敗事例

験の結果、学習済みモデルの比較においては、植物に対してもファインチューニングを行う学習済みモデルの違いによって精度が変化することを確認した。一方で、画像増幅の比較においては、増幅を適用していないオリジナルデータセットのほうが高い精度となり、期待通りの精度向上は見込めなかった。このことから、今後はデータセットの修正・追加によるデータセットの改善、増幅手法の改善により、本研究タスクにおいても画像増幅が有効な手法となるよう検証をすすめる。さらに、画像から葉長の推定を行い、実運用で想定される動画データから「節間距離」「葉長」の推定を高精度に行うため、検出速度向上や個体識別の検討を行っていく。

6. 謝辞

本研究の一部は、JST 創発的研究支援事業 (JPMJFR201B) の支援を受けたものである。また、データセットの提供及び農学知見を支援いただいた株式会社 PLANT DATA の皆様、NTT 社会情報研究所の皆様に深い感謝の意を表す。

参考文献

- [1] 山本恭輔, "画像解析と機械学習によるトマトの自動生育診断および高速フェノタイプングに関する研究", 東京大学大学院 農学生命科学研究科博士論文 (2015). <<https://irdb.nii.ac.jp/en/00926/0001754966>>(参照 2021-12-16).
- [2] Guo, W., Rage, U. K., and Ninomiya, S., "Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model", Computers and Electronics in Agriculture, Vol. 96, pp. 58-66 (2013).
- [3] Csurka, G., Dance, C. R., Fan, L., Willamowski, J., Bray, C., "Visual categorization with bags of keypoints", Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision, pp. 59-74 (2004).
- [4] Zhe, C., Gines H., Tomas, S., Shih-En, W., Yaser, S., "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186 (2019).
- [5] 長谷川達人, 田中基貴, "水産資源管理に向けた Mask Keypoint R-CNN による魚体長認識", 情処研報 Vol.2021-

- CDS-32, No.12 (2021).
- [6] N Dinesh Reddy, Minh Vo, Srinivasa G. Narasimhan, "CarFusion: Combining Point Tracking and Part Detection for Dynamic 3D", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1906-1915 (2018).
 - [7] He, K., Gkioxari, G., Dollár, P., Girshick, R, "Mask R-CNN", 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988 (2017).
 - [8] FaceBookAI, "Detectron2", (2021).
<<https://github.com/facebookresearch/detectron2/>>(参照 2021-12-16)
 - [9] Justin Brooks, "COCO Annotator", (2019).
<<https://github.com/jsbrooks/coco-annotator/>>(参照 2021-12-16)
 - [10] "Common Objects in Context(COCO) Keypoint Evaluation"
<<https://cocodataset.org/#keypoints-eval>>(参照 2021-12-16)
 - [11] "imgaug"(2020)
<<https://github.com/aleju/imgaug/>>(参照 2021-12-16).
 - [12] Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S., "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
 - [13] Tsung-Yi, L., Piotr, D., Ross, G., Kaiming, H., Bharath, H., Serge, B., "Feature Pyramid Networks for Object Detection", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
 - [14] Saining, X., Ross, G., Piotr, D., Zhuowen, T., Kaiming, H., "Aggregated Residual Transformations for Deep Neural Networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017).
 - [15] Angela Casado-García, Jonathan Heras, "Ensemble Methods for Object Detection", Frontiers in Artificial Intelligence and Applications, pp. 2688 - 2695 (2020).