CH-015

果樹栽培の摘果作業支援を目的とした着果密度･空間配置の認識
Identification of Fruiting Position for Fruit Picking Operations with Enhanced Precision in Fruit Cultivation

佐藤 響† 石井 雅樹† 伊東 嗣功† 堂坂 浩二†
Hibiki Sato Masaki Ishii Hidekatsu Ito Kohji Dohsaka

1. はじめに

近年，日本の農業従事者数は減少傾向にあり，基幹的農業従事者の数は平成 23 年から令和 3 年までに約 30% 減少している．加えて，労働力の高齢化も深刻な課題であり，平成 23 年から令和 3 年までの間で平均年齢は約 2 歳上昇している[1]．こうした背景から，農業分野では慢性的な労働力不足が課題として挙げられ，秋田県内においてもそれらの影響は顕著である．

秋田県内陸地域で栽培が盛んなリンゴをはじめとする園芸作物の場合，継続的に高品質な果実を得るため，適度につぼみや花や幼果を間引く摘蕾・摘花・摘果とよばれる作業が欠かせない．しかし，こうした作業の多くが熟練者の経験や勘によるものであり，隣接している果実との間隔を把握しながら着果密度・果枝長を考慮した間引きが求められるため技術に差が生じやすい．一方で，熟練した雇用労働力の確保も困難であるため，着果管理技術のばらつきや新規就農者への栽培技術継承が課題となっている．

本研究では，果樹試験場や篤農家の方々が持つ栽培技術を画像処理によって情報化し，摘花・摘果作業の平易化を可能とするシステムの開発を目的としている．本稿では，果樹の着果密度や空間配置を認識する手法を提案する．

2. リンゴ樹の生態と管理作業の現状

リンゴ樹の管理を行うにあたり，10 a あたりに必要な作業時間を表 1 に示す．農林水産省の調査[2]では，管理作業の合計作業時間 272.89 時間のうち，全体の約 25%を授粉・摘果の作業が占めている．

加えて，こうした管理作業を怠ってしまった場合，その年はもちろん翌年以降にも糖度や着色の低下といった商品の価値に直結する影響が大きく生じてしまう．現状では熟練者の経験や勘に頼っているこうした技術を見える化することで，新規就農者はもちろん，既存農業者や雇用労働者の作業能力向上が期待される．

表 1　リンゴ樹における 10 a あたりの作業時間割合

作業内容	作業時間(h)	作業内容	作業時間(h)
基　肥	1.71	管　理	63.21
整枝･せん定	35.56	袋かけ･除袋	17.20
追　肥	0.49	収穫・調製	47.13
除草・防除	17.42	出　荷	19.66
授粉・摘果	67.16	管理･間接労働	3.35

†秋田県立大学　Akita Prefectural University

3. 先行技術

3.1 一般物体検出アルゴリズム

深層学習（ディープラーニング）を用いた一般物体検出アルゴリズムは，教師データを元に動画の中から学習済みの物体を認識するアルゴリズムである．これらは認識の過程によって大きく二つに分類される．Faster R-CNN に代表される Two-Stage 法は物体領域の検出と領域候補のカテゴリー識別を直列に実行するものである．一方，YOLO（You Only Look Once）[3]や SSD（Single Shot Detector）[4]は Single-Stage 法に分類され，領域の抽出とカテゴリー識別を同時に行う手法である．

今回は，従来の Faster R-CNN などと比較して高速でありながらほぼ同等の精度が得られる YOLOv4 [5]を使用した．これは，$S \times S$ のグリッドに分割した局所領域ごとに矩形の中心座標と大きさ，信頼度をカテゴリー識別と同時に算出する手法[6]であり，畳み込み層，プーリング層および全結合層で構成されている．また，従来の YOLOv3 と比較して大幅に検出精度が向上している．

3.2 評価指標

学習モデルの性能を評価する指標について以下で簡単に述べる．

3.2.1 TP および FP

TP（True Positive）とは，画像上で検出すべき物体を正しく検出できた数を示しており，FP（False Positive）は検出すべきでないものを誤って検出してしまった数である．このほか，TN（True Negative）は検出すべきでないものを画像上にはないと正しく判断できた数を示し，FN（False Negative）は検出すべきものを検出できなかった数を示している．

3.2.2 適合率と再現率

適合率 P（Precision）とは検出結果の中に適合しないものが入っていない割合のことを示し，式 3.1 で表される．

$$P = \frac{TP}{TP \cup FP} \tag{3.1}$$

また，再現率 R（Recall）とは，すべてのデータのうち，どの程度のデータを拾うことができたかを示し，式 3.2 で表される．

$$R = \frac{TP}{TP \cup FN} \tag{3.2}$$

3.2.3 IoU

認識された物体の領域を示すバウンディングボックスの重なりを評価するための指標として IoU（Intersection over Union）がある．IoU は式 3.3 で表される．

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \tag{3.3}$$

ここで，B^{gt}は正解領域を示し，Bは予測領域を示している．つまり，正解領域と予測領域との重なりが大きいほど高い値を示す．

3.2.4 平均適合率および mAP

平均適合率 AP（Average Precision）は，先述の適合率 P および再現率 R から算出することができる．再現率が r のときの適合率の値を $P(r)$ とすると，平均適合率 AP は PR 曲線の面積として式 3.4 で表すことができる．

$$AP = \int_0^1 P(r)dr \tag{3.4}$$

なお，実際には式 3.4 で得られる PR 曲線の面積を矩形に近似している．再現率が $r_1, r_2, \cdots, r_N$，そのときの適合率を $P(r_1), P(r_2), \cdots, P(r_N)$ としたとき，式 3.5 のように求められる．

$$AP = \sum_{i=2}^{N} (r_i - r_{i-1}) P(r_i) \tag{3.5}$$

mAP（mean Average Precision）とは各カテゴリーにおける平均適合率の学習モデル全体での平均値を示し，式 3.6 で表される．つまり，すべての時点 i における平均適合率を示し，この値が大きいほど学習モデルの精度は高いといえる．なお，式 3.6 において C はカテゴリーの総数を示している．

$$mAP = \frac{1}{C} \sum_{i=1}^{C} AP_i \tag{3.6}$$

本研究で用いた YOLOv4 では mAP@50 が用いられている．これは，IoU=50，つまり重なりが 5 割以上を示したときに正解とみなす指標である．

3.2.5 損失関数

損失関数（Loss）は，予測された矩形領域（バウンディングボックス）と正解ラベルとの誤差を示す関数である．一般的な物体検出アルゴリズムでは，平均二乗誤差（MSE, Mean Square Error）を用いることが多いが，YOLOv4 では，IoU（Intersection over Union）をもとにした CIoU-loss（Complete IoU Loss）と呼ばれる手法を用いている．

CIoU-loss では，矩形領域の重なっている部分の面積や中心点の距離，アスペクト比などを考慮した上で，従来の方法に比べ高速な収束と性能向上を実現している[7].

CIoU-Loss は式 3.7 で示される．

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{3.7}$$

ここで，$\rho(\cdot)$はユークリッド距離を表し，正解領域B^{gt}の中心点b^{gt}と予測領域Bの中心点bの 2 点間の距離を意味している．また，cは正解領域B^{gt}と予測領域Bの両方を囲む最小の矩形の対角線長を表している．αは式 3.8 に示す通り正のトレードオフの関係にあるパラメーターで，式 3.9 に示すvはアスペクト比の整合性を測定する関数である．

なおw，w^{gt}とh，h^{gt}はそれぞれ正解領域B^{gt}と予測領域Bの縦と横の長さを示している．

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{3.8}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{3.9}$$

このとき，YOLOv4 全体での損失関数は式 3.10 のようになる[8].

$$\begin{aligned} Loss = {} & 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v - \\ & \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \left[\hat{C}_i \log(C_i) + \left(1 - \hat{C}_i\right) \log(1 - C_i) \right] - \\ & \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{noobj} \left[\hat{C}_i \log(C_i) \right. \\ & \left. + \left(1 - \hat{C}_i\right) \log(1 - C_i) \right] - \\ & \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} \left[\hat{p}_i(c) \log\left(p_i(c)\right) \right. \\ & \left. + \left(1 - \hat{p}_i(c)\right) \log\left(1 - p_i(c)\right) \right] \end{aligned} \tag{3.10}$$

式 3.10 において，S^2 は，$S \times S$ グリッドを表し，各グリッドが B 個の候補ボックスを生成する．信頼度損失関数および分類損失関数の算出には交差エントロピー誤差が用いられている．なお，ボックス内に候補がない場合については信頼度損失のみが測定され，重み係数 λ を増加させる．

4. 提案手法

4.1 概要

秋田県果樹試験場（秋田県横手市）で栽培しているリンゴ樹の摘花・摘果作業を，作業者の頭部につけたアクションカメラ（GoPro）によって画像や映像として記録し，つぼみや花，幼果の識別器を作成する．識別器の作成には深層学習を用いる．

加えて，深度情報を測定することが可能な RGB-D カメラで記録し，これらのデータを組み合わせることで隣接する果実との間隔や着果密度を解析する．

4.2 花と果実の認識方法

花と幼果の認識には一般物体検出アルゴリズムを使用した．撮影した画像をもとに，以下に示す手順で深層学習を行った．

4.2.1 教師データの作成

果樹試験場において撮影した作業映像から数フレーム毎に静止画を作成した．静止画は 2020 年 5 月 3 日撮影の動画から切り出したものが 102 枚，同年 6 月 3 日撮影のものが 815 枚，6 月 4 日撮影のものが 575 枚の計 1,492 枚である．撮影時期とリンゴ樹の関係としては，5 月 3 日撮影の画像にはつぼみや花が多く，6 月 3 日および 4 日撮影の画像には幼果が主に記録されている．

これらに対して，アノテーションツール「labelImg」を用いて矩形領域を設定し，「Bud（つぼみ）」「Flower（花）」「YoungApple（幼果）」の 3 種類をタグ情報として付与した．計 1,492 枚の全データ上に付与されたタグの内訳を表 2 に示す．

4.2.2 学習モデルの作成

学習データと評価データの内訳は 8:2 に設定し，学習データを 1,194 枚，評価データを 298 枚とした．それぞれに付与されたタグの内訳を表 3 に示す．これらのデータについて，YOLO を用いた学習を行った．

表 2　付与されたタグの内訳（全データ）

種　類	つぼみ	花	幼果
タグ数	4,412	600	10,630

表 3　付与されたタグの個数内訳（学習データ・評価データ）

種　類	つぼみ	花	幼果
学習データ	3,591	488	8,539
評価データ	821	112	2,091

4.3 リンゴ樹の空間把握

花や幼果を動画像上から認識した上で，深度情報を取得することでリンゴ樹の空間把握を行う．

先述の識別器作成ではアクションカメラによる映像をもとにした．一方で，リンゴ樹の着果密度を算出するためには，リンゴ樹を三次元的に測定する必要がある．そのため本研究では，深度情報を取得することのできる RGB-D カメラを用いた．

4.3.1 使用機材

深度情報の取得には Intel 社の RealSense シリーズと呼ばれる RGB-D カメラを用いた．これは，通常の RGB 画像に加え，備え付けられた 2 台の赤外線カメラの視差を用いることで，形状や深度（Depth）情報を取得することが可能なデバイスである．

本研究に用いた RGB-D カメラは 3 機種（RealSense D415 および D435i，D455）である．3 機種ともに，深度の測定手法は同一のアクティブ IR ステレオ法を用いているものの，撮影可能な画角や測定可能距離が異なっている．

4.3.2 3 次元ユークリッド距離の算出

着果密度を測定するため，YOLO を用いて画像から認識された矩形領域間の距離を算出する．この算出は，取得した深度情報を利用することで，3 次元ユークリッド距離を計算するものである．YOLO によって認識された 2 点の物体矩形がある場合，それぞれの 3 次元座標 (x_1, y_1, z_1) および (x_2, y_2, z_2) を取得し，結ぶ線分の長さ d を式 4.1 のように計算している．

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \qquad (4.1)$$

また，算出されたユークリッド距離から，果樹試験場によって摘果の目安とされている 25 cm 以内に果実が隣接している場合は線分を出力として表示する．

5. 結果と考察

5.1 学習モデルの作成結果

前述の YOLO を用いた学習結果を以下に示す．本研究では，学習回数は 6,000 回とした．学習回数と損失関数および mAP の推移を図 1 に示す．作成した学習モデルの mAP は 65.2%となった．

また，学習 1,000 回ごとの適合率，再現率の推移を図 2 に示す．

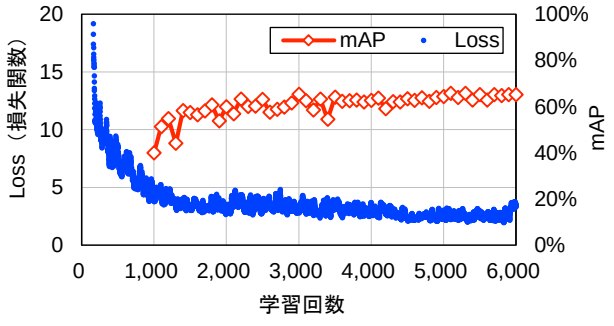

図 1　学習回数と損失関数，mAP の推移

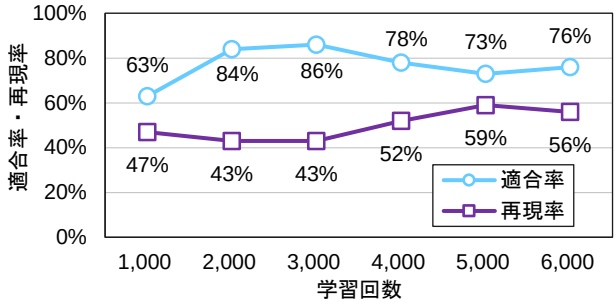

図 2　学習回数と適合率，再現率の関係

加えて，各カテゴリーでの TP，FP および AP を算出した．表 4 に「Bud（つぼみ）」について，表 5 に「Flower（幼果）」についてのデータを示している．

表 4 に示した通り，幼果の AP が 70% 程度であり，花についても幼果と同等の水準となった．一方，蕾の AP は表 5 の通り 50% 程度となった．これは教師データの質（撮影画質や撮影日の偏りなど）に起因するものではないかと考えられる．そのため，撮影時の天候条件やリンゴ樹の生育段階により一層合わせた教師データの作成が必要である．

5.2 深度情報の取得結果

3 種類の RGB-D カメラを用いて，2021 年 7 月 8 日に秋田県果樹試験場において深度情報の撮影を行った．なお，撮影時の天候は曇りである．

RGB-D カメラでの撮影は，脚立上で摘果作業を行うときの目線を想定し，一脚上に装着した RGB-D カメラをノートパソコンに接続する形で行った．

RGB-D カメラを用いて取得された RGB 画像の一例を図 3 に，Depth 画像は図 4 にそれぞれ示す．ここで示す画像の撮影デバイスは当日最も良好にデータを取得することのできた Intel RealSense D435i である．なお，図 4 および図 5 の Depth 画像において，青色（寒色）側がカメラから近い領域を，赤色（暖色）側が遠い領域を示している．濃紺色で示された領域は物体の影や測定可能距離よりも近距離に物体が存在していることによって深度情報が欠損している部分である．

表 4　各カテゴリーにおける TP，FP および AP（つぼみ）

学習回数	TP	FP	AP
1,000	379	506	32.55%
2,000	183	42	48.93%
3,000	229	71	50.74%
4,000	291	167	45.99%
5,000	360	226	49.66%
6,000	331	175	50.76%

表 5　各カテゴリーにおける TP，FP および AP（幼果）

学習回数	TP	FP	AP
1,000	1,016	287	62.33%
2,000	1,078	191	71.84%
3,000	996	118	73.28%
4,000	1,231	281	73.47%
5,000	1,360	431	74.10%
6,000	1,296	346	74.27%

図 3　RGB-D カメラによって撮影された RGB 画像

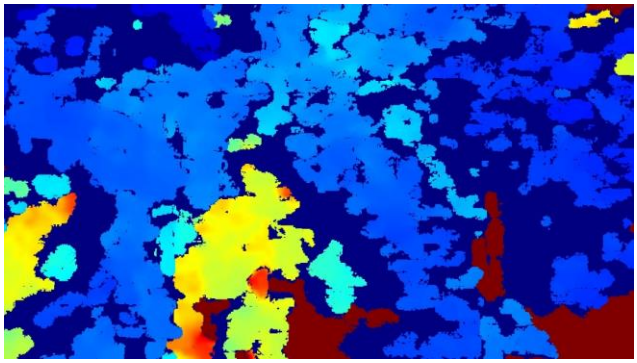

図 4　RGB-D カメラによって撮影された Depth 画像

得られた深度情報は，図 4 に示したように物体の影となる部分に欠損が生じている．そのため，次の手順に従ってフィルター処理による補完を行う．

はじめに，デシメーションフィルター（Decimation Filter）を用いて画像を圧縮し，再度拡大させることで深度情報の複雑さを軽減させる．続いて，空間的エッジ保存フィルター（Spatial Edge-Preserving Filter）[9]を用いて平滑化の処理を行う．

なお，深度情報の欠損した領域に対して，近傍の値を用いて穴埋めを行う．本研究では，隣接領域の最も値の大きい（遠い）部分の値を取得して穴埋めを行っている．フィルター処理後の Depth 画像を図 5 に示す．

5.3 予想収穫果の表示

動画像上から認識された幼果の位置に，予想収穫果をオーバーレイ表示する．取得された深度情報から，動画像上での予想収穫果のサイズを算出し，図 6 のようにオーバーレイ表示を行っている．

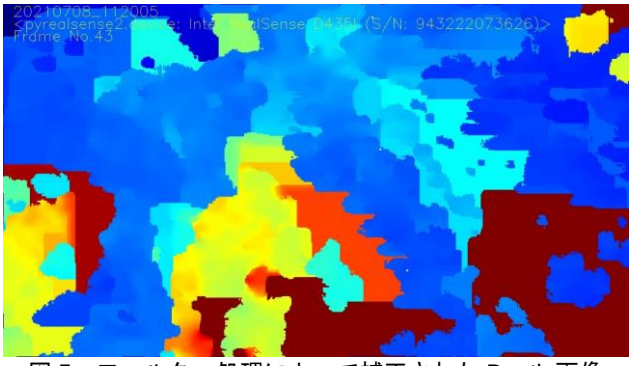

図 5　フィルター処理によって補正された Depth 画像

図 6　予想収穫果の表示

6. おわりに

本研究では，リンゴ樹のつぼみや花，幼果を動画像上から識別する識別器を作成した．加えて，深度情報を併用することで果樹の着果密度や空間配置を認識するシステムの基礎検討を行った．

実際の果樹管理作業の動画から各物体を識別できる学習モデルを開発し，花と幼果で 70%，つぼみで 50%の AP を達成した．奥行き情報を取得し，画像認識結果と組み合わせることで果実間の 3 次元ユークリッド距離の算出が可能となった．

今後の取り組みとして，物体認識精度および認識速度の向上，取得された深度情報の精度評価を検討している．

参考文献

[1] 農林水産省, “農業労働力に関する統計”, https://www.maff.go.jp/j/tokei/sihyo/data/08.html, (参照 2022-05-14).
[2] 農林水産省大臣官房統計部, “品目別経営統計（りんご）”, 品目別経営統計〈平成 19 年産〉―農業経営統計調査報告, (2010).
[3] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, IEEE-Conference on Computer Vision and Pattern Recognition (CVPR), pp.779-788 (2016).
[4] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu and Alexander C. Berg, “SSD: Single Shot MultiBox Detector”, Computer Vision – ECCV 2016, pp.21-37 (2016).
[5] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, arXiv:2004.10934 [cs.CV] (2020).
[6] 藤吉弘亘, 山下隆義, “深層学習による画像認識”, 日本ロボット学会誌, Vol.35, No.3 (2017).
[7] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, Dongwei Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression”, Proceedings of the AAAI Conference on Artificial Intelligence, Vol.34, No.07 (2020).
[8] Wu, Lin, Jie Ma, Yuehua Zhao, and Hong Liu, “Apple Detection in Complex Scene Using the Improved YOLOv4 Model”, Agronomy, Vol.11, No.3 (2021).
[9] Eduardo S. L. Gastel, Manuel M. Oliveira, “Domain Transform for Edge-Aware Image and Video Processing”, ACM Transactions on Graphics, Vol.30, No.4 (2011).