大豆の生育情報を自動取得する画像センシング手法の開発: Single Shot MultiBox Detectorの導入

Development of Image Sensing Method to Automatically Obtain Soybean Growth Condition: Introduction of Single Shot MultiBox Detector

大村 和暉 *1	八幡 壮 *1	小澤 誠一*1	大川 剛直 *1	村上 則幸 * ²	辻 博之* ²
Kazuki Omura	So Yahata	Seiichi Ozawa	Takenao Ohkawa	Noriyuki Murakami	Hiroyuki Tsuji
	*1神戸フ	大学 * ² 北	海道農業研究セ	ンター	
	Kobe Univ	ersity Hokkaido N	ational Agricultural I	Research Center	

Observing growth state automatically is a very important task in smart farming. In this paper, we propose an image sensing method to detect soy flowers and seeds by using a state-of-art deep learning architecture called Single Shot MultiBox Detector (SSD). We also suggest a new method for counting seeds from a video. We used SSD, fast to compute and has high accuracy. We collected images at Hokkaido Agricultural Research Center from 2015 to 2017. Seed and flower detection performed 0.586 (F-measure) and 0.646(F-measure), respectively.

1. はじめに

近年の国内農業では、農業従事者の減少と高齢化が問題に なっている。その中で農家が蓄積してきた農業技術や知識が失 われることが懸念されており、生産力向上のために熟練者の/ ウハウを継承させることが課題となっている。

そこで、農業に ICT を導入して自動化・効率化を図った「ス マート農業」が注目を浴びている [E-Agriculture]. スマート 農業では、農作物の生育情報や環境情報をセンサやカメラなど から取得し、適切な栽培条件を見つけ出して自動制御すること で、収量の増加と農作業の効率化を目的としている. これによ り生産者の技術や経験に依存しない生産活動が期待されてい る. その中で、花や子実の検出データは農作物の収量につなが る重要な情報であり、スマート農業に有用であるといえる.

本論文では,生育情報として大豆の花,子実に着目し,屋外 圃場で得られた画像からこれらを認識するシステムを提案する. 2016 年度以前では静止画で株を撮影していたのに対し,2017 年度のデータセットでは動画で撮影を行っていた.以前のシス テムでは処理に時間がかかることが課題であり,動画に適用す ることが困難だった.しかし,今後動画に対して適用可能な検 知システムを構築するためにも,高速な処理が必要となる.そ こで,Convolutional Neural Network (CNN)[Krizhevsky 12] の一種である Single Shot MultiBox Detector (SSD)[Liu 15] を利用した.本論文では,この SSD を使用した高速な検出が 可能なフェノタイピングシステムを提案する.

本論文では、2節で子実検知について、3節で花の検知手法 について述べる.4節では実際の大豆画像に対して検知性能の 評価を行い、5節でまとめを述べる.

2. 子実の検知

大豆の子実は緑色をしているため,蜜柑やトマトなどと異 なり色相情報のみで葉や茎と区別するのは困難である.また, 部分的な形状も茎や葉と似通っており,勾配情報でも抽出しに くい.そこで,茎や葉との位置関係や背景情報などの情報を総 合的に捉えて物体検出することが必要である.SSDは,画像 全体に畳み込み処理を施して特徴マップを生成し,マップ上で

連絡先: 小澤 誠一, 神戸大学数理・データサイエンスセンター, 〒 657-8501 神戸市灘区六甲台町 1-1, TEL: 078-803-6466, E-MAIL: ozawasei@kobe-u.ac.jp 物体の検知を行う手法である.よって,特徴マップ内には物体 およびその周辺から抽出した情報が含まれている.そのため, 全体的な特徴量を捉えながらの検知が可能である.

2.1 SSD による子実検知

深層学習モデルの一つである SSD は、画像内の物体の位置 検出とクラス分類を同時に高速に行うことができるのが特徴で ある. この SSD を本研究で導入する主な理由は、まず、子実 の周囲にある物体や背景の情報を取り入れつつ子実の検知を行 えるため、あきらかに子実が存在しない場所での誤検知の抑制 が期待される点が挙げられる.また、圃場で花数や子実数のリ アルタイム計測が可能となることが望ましいため、動画に対し て高速に物体を検知・分類できる SSD が有望であると考えた からである.

SSDでは、図1のように画像全体を一定サイズの正方形にリ サイズしてから、ベースネットワークに入力する.ベースネッ トワークは、事前に一般物体認識データセットで学習させた CNNであり、SSDの原論文に従って VGG16[Simonyan 14] を使用している.ベースネットワークの中間層における特徴 マップを取り出し、これを1つ目の特徴マップとする.それを さらに畳み込んでいくことで、大きさが異なる複数の特徴マッ プを生成する.これらのマップ上で物体の探索と分類を行う.

続いて、特徴マップ上での位置検出と分類の方法について 述べる。旧来のように原画像全体を移動窓で探索する方法で は、アスペクト比とサイズを変えながらの探索が必要であるた め計算量が非常に大きくなる. そこで SSD では,特徴マップ 上で移動窓を使用することで、原画像上の場合よりも大幅に 探索回数を減らして高速化している。例えば4×4の特徴マッ プの上で,図2(b)のように、あらかじめ設定しておいたサイ ズ・アスペクト比の移動窓で1ピクセルごとに探索していく とする. これは、入力画像を4×4のグリッドに分割して、各 グリッドを中心に図 2(a) のように探索していくのと同じこと になる.よって、大きな特徴マップにおいては小さなグリッド に区切って小物体検知を、小さな特徴マップにおいては大きな グリッドで大物体検知をすることが可能である.したがって, 図3のように複数の特徴マップを用いて探索をしていくこと で、マルチスケールの物体検知が可能になる。移動窓のサイズ を調整し、小物体を検出できるように調整した.

さらに、子実検知においては性能を高めるためにファイン チューニングを行っている.ファインチューニングとは、あら



図 1: SSD のアーキテクチャ[Liu 15]



(a)入力画像上での窓

図 2: 特徴マップ上での探索 [Liu 15]

かじめ別のデータセットで学習しておいたモデルを用意して おき,そのモデルのパラメータを初期値として,本来学習した いデータセットで再学習させることである。一般的には、別の データセットには大量の画像を含むデータセットを使用する. ファインチューニングによって、学習データが不十分でも十分 な性能を発揮できる場合がある.また、学習時間を短縮する効 果がある.ベースネットワークのパラメータは初期値のまま固 定し、図1の conv4_3 以降のパラメータをファインチューニ ングした.

花の検知 3.

花検知においても、子実検知と同様に SSD を用いる。2.1節 でも述べたように、SSD はマルチスケールの物体検知が可能 であり、小物体の探索は大きな特徴マップ上で行われている. 花は子実に比べて小さな物体であることから、大きな特徴マッ プがあれば花検知に対してより効果的であるといえる.

そこで、花検知では SSD1024 を実装した。SSD1024 とは、 入力画像のサイズを 1024×1024 にした SSD である。単に入 カサイズを拡大しただけでなく、サイズの変更に合わせてモデ ルの構造も微調整し、既存手法よりスケールの大きな特徴マッ プを得られるようになった.表1に入力サイズと特徴マップの サイズを示す. 表の通り, 特徴マップが拡大されていることが 分かる.



図 3: 複数マップを用いたマルチスケールな位置検出 [Liu 15]

表 1: 各特徴マップにおける	1辺あたりのピクセル数 (一部の
み抜粋)	

画像サイズ	conv4_3	conv7	$conv8_2$	$conv9_2$
300	38	19	9	5
512	64	32	16	8
1024	128	64	32	16



図 4: SSD と CNN による 2 段階検知

花検知では、クラス分類で出力されるスコアのしきい値を上 げればより花らしい領域のみを残すことができるが,花の見落 としも増えやすくなる。逆にしきい値を下げると、花の見落と しは減るものの誤検知が増える危険がある。そこで、図4のよ うに、SSD と CNN による 2 段階の手順で花検知を行う。つ まり、まず SSD のしきい値を低めに設定して花の検知を行う。 続いて,SSD が検出した矩形領域を覆うような正方形で画像 を切り抜き、CNN に入力して本当に花かどうか精査する。花 でないと判定された領域は棄却し、残った領域のみを花領域と する.以上の2段階の手順によって,見落としと誤検知の両 方を減らすようにしている. 今回は精査用の CNN として、学 習が早く性能も高いことから Inception[Szegedy 15] に Batch Normalization[Ioffe 15] を追加したモデルを使用した.

性能評価 **4**.

画像の撮影環境 4.1

本研究では、北海道農業研究センターの協力のもと、芽室 研究拠点の試験圃場内で撮影された大豆の画像を用いた。簡 易な撮影を目指すために撮影手法の改良を進めており、年度ご との撮影方法は2のようになっている。撮影期間は2015年か ら 2017 年の間である。なお、2017 年については動画から一 定のフレーム間隔で静止画を取得しており、それを撮影枚数と している. 2015年と 2017年は、図 5(a) のようにカメラを下 から上に等速で上昇させながら撮影を行った。2015年は一定 間隔での静止画撮影,2017年は動画モードによる撮影である. 2016年は、図5(b)のようにカメラ5台を垂直方向に並べ、リ モコンで同時撮影することで連続画像を取得した。

試験圃場内の複数ブロックで撮影が行われ、一部のブロック では多収阻害要因を特定するため、生育初期に除草剤が意図的 に多く散布されている.

表 2: 大豆画像の撮影方法

撮影年度	撮影方式	カメラ台数	撮影枚数 (枚/株)
2015	静止画	1	31
2016	静止画	5	5
2017	動画	1	7



図 5: 撮影装置概念図 . (a)2015 · 2017 年, (b)2016 年

表	3.	子実椅	資知の	性能	評価	Ŧ
1	.	1 10	ミハロッノ	1.1.1164	U I I IIII	ц

24.0						
モデル	適合率	再現率	F 値			
SSD300	0.602	0.571	0.586			
SSD512	0.454	0.491	0.472			

4.2 子実検知結果

子実の検知は、北海道芽室町のテスト用区画において、2017 年8月に撮影された231枚の画像を評価に使用した.訓練画 像は、芽室において2015年および2016年8月、2017年8月 (テスト用区画以外)に撮影された合計636枚を使用した.本 実験では、SSD300と512を使用した.どちらも一般物体認 識データセット21493枚で学習させたパラメータを初期値と してファインチューニングを行った。事前にラベル付けした子 実領域と、検知した領域のオーバーラップの割合がしきい値以 上であればTrue Positive(検知成功)とする。また、すでに検 知された子実領域を他の検知結果が検知したとしても、それは False Positive(検知失敗)とする。

性能の指標として、テスト区画全体にシステムを適用した ときの再現率,適合率、F値、ROC曲線を求めた.ROC曲 線とは、横軸に False Positive の割合、縦軸に True Positive の割合をとり、識別器のしきい値を変えた時のそれぞれの割合 をプロットすることで描いた曲線である.曲線下部の面積が大 きければ大きいほど高性能であるといえる.

評価結果は表3のようになった.表より,SSD300の方が高 性能であると分かる.SSD512では物体探索時の移動窓を小さ めに設定しており,それが子実の大きさに対してふさわしいパ ラメータではなかったことが原因であると考えられる.また, SSDの適用例を図6に示す.画像を見ても,SSD512の方が 検出している領域が小さく,子実を捉えきれていない場合が多 い一方で,SSD300は子実に近い大きさの矩形で検出している ことが分かる.今後,パラメータを改善することでSSDをよ り高性能にできる可能性がある.

子実検知における SSD300 と SSD512 の ROC 曲線は図 7 のようになった. この図を見ても SSD300 の方が曲線が上方向に広く, SSD300 が優れていることがわかる.

4.3 花検知結果

花の検知は、北海道芽室町のテスト用区画において、2017 年7月に撮影された74枚の画像を評価に使用した。訓練画像 は、芽室において2015年および2016年の7~8月、2017年7 月(テスト用区画以外)に撮影された合計2323枚を使用した。 なお、テスト区画は子実検知の場合と同じである。本実験では、 SSD512とSSD1024を使用した。SSD512は、入力画像を拡 大することによって性能が向上するかどうかを比較するために 使用している。SSD512は子実検知と同じ事前学習済みモデル



図 6: 子実の検知結果 (青色:子実ラベル,水色: True Positive, 黄色: False Positive)



図 7: 子実検知の ROC 曲線

を用いてファインチューニングしているが,SSD1024の方は公開されているライブラリに含まれておらず,事前学習済みモデ ルが存在しない.そこで、ベースネットワークのみをSSD512 の事前学習済みモデルと同じパラメータで固定し、それ以外 は xavier 初期化 [Glorot 10] によって初期値をとってファイ ンチューニングなしで学習させた.精査用の Inception モデル には事前学習済みモデルが存在するので、そのパラメータを 初期値とし、2015 年と 2016 年の花画像を切り取ったデータ セットを用いてファインチューニングを行っている.SSD512, SSD1024 のどちらの場合も、検出結果の精査に用いた CNN はこの同じモデルである.

性能評価では、子実検知と同様に再現率、適合率、F値、ROC 曲線を用いた.しかし、花検知において問題となるのは、開花 途中、枯れかけであったり、ボケが極めて強いなど、ラベル付 けの判断が困難な花が多く存在することである.そこで、撮 影条件によってラベル付けに検知の重要度を考慮することに する.鮮明に写っており、かつ満開の花は検知するべきなので 「検知必須」とし、判断がしにくいものについては検知の成否 を問わず「検知任意」とした.検知の成否判定は子実と同様 に、ラベルとのオーバーラップの割合で判定する.また、再現 率と適合率の計算式は、ラベル付けを変えたことに合わせて次 のように変更している.

再現率	=	検知された検知必須の花数 検知必須の花数
適合率	=	検知必須または検知任意を検知した数 全検知数

評価結果は表4のようになった.表より,SSD1024の方が

表 4:	花検知の性能評価	

モデル	適合率	再現率	F 値
SSD512	0.155	0.571	0.244
SSD1024	0.585	0.722	0.646



図 8: 花の検知結果 (赤色:検知必須,青色:検知任意,紫色: 検知した領域)

高性能であると分かる.また,SSD の適用例を図 8 に示す. SSD512 は誤検知が多いのに対し,SSD1024 は誤検知が少な い上に,花領域をより密接に捉えていることが分かる.

ROC 曲線は図 9 のようになった. これは, CNN による 精査なしで SSD 単体の出力の検知結果を用いて描いている. SSD512 よりも SSD1024 の方がわずかに曲線下部の面積が大 きく,入力画像を拡大したことによって性能が向上したことが 確認できる.

5. さいごに

本研究では、大豆の生育情報を自動観測する画像センシン グシステムとして、SSD をベースとした子実と花の検知シス テムを提案した.子実検知,花検知では、それぞれ 0.586 (F 値), 0.646(F 値) という性能を示した.また、処理の高速化に よって、今後動画にも適用していくことが可能になった.

今後さらに検知性能を高めていくとともに、動画に対しての 検知を進めてより詳細な生育情報を取得し、データ解析によっ て人手では得られなかったノウハウの取得を目指したい.



図 9: 花検知の ROC 曲線

謝辞

本研究は、平成27年度農林水産省委託プロジェクト研究「多 収阻害要因の診断法及び対策技術の開発」の助成を得て行わ れた.

参考文献

- [E-Agriculture] E-Agriculture, : Agricultural Value Chains and ICT, http://e-agriculture.org/ agricultural-value-chains-and-ict
- [Glorot 10] Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, In Proc. of the Thirteenth International Conference on Artificial Intelligence and Statistics, Vol. 9, pp. 249–256 (2010)
- [Ioffe 15] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, In Proc. of the 32Nd International Conference on International Conference on Machine Learning, Vol. 37, pp. 448–456 (2015)
- [Krizhevsky 12] Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, In Proc. of the 25th International Conference on Neural Information Processing Systems, Vol. 1, pp. 1097–1105 (2012)
- [Liu 15] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C.-Y., and Berg, A. C.: SSD: Single Shot MultiBox Detector., *CoRR*, Vol. abs/1512.02325, (2015)
- [Simonyan 14] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, CoRR, Vol. abs/1409.1556, (2014)
- [Szegedy 15] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, *In Proc.* of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)