

動画像からの病虫害の検出・識別に関する検討 Detection and Identification of Pests from Video Based on Convolutional Neural Network

葛原 宏章[†] 滝本 裕則^{††} 佐藤 安弘^{†††} 金川 明弘^{††}

Hiroaki Kuzuhara[†] Hironori Takimoto^{††} Yasuhiro Sato^{†††} Akihiro Kanagawa^{††}

[†] 岡山県立大学大学院 情報系工学研究科システム工学専攻 ^{††} 岡山県立大学 情報工学部情報通信工学科

^{†††} 科学技術振興機構さきがけ・龍谷大学 農学部

1 はじめに

近年、農業の現場では農業の高齢化と後継者不足が深刻さを増している。特に農業経営の一大リスクとして病虫害の被害が挙げられる。微小昆虫による食害は病気の媒介や品質の低下を招く深刻な問題であるにも関わらず、確実な防除には現場作業者の熟練度に強く依存しており、新規農業者にとっての障壁となっている。この課題に対する対策のひとつに ICT に基づくスマートアグリ (スマート農業) が注目されており、高齢化する農業従事者や新規農業者を支援する技術の実現が期待されている。そこで、図 1 のように、農業における病虫害駆除を支援する基盤の技術としてカメラと深層学習を用いて動画像から病虫害の検出と診断をすることを考える。虫の検出、識別を自動で行うことができれば熟練度の低い新規農業従事者でも病虫害の適切な防除を行うことができる。

物体検出・識別に関する技術としては街中の監視カメラで撮影した映像から人や車を検出する技術等が既にいくつか提案されている [1]。しかしながら、今回対象とした病虫害は微小であり外観が類似しているため、既存技術の転用ではこれら的高精度かつ高速な検出・識別を行うことは困難であると考えられる。実際に、先行研究 [2] において、深層学習に基づく物体検出・識別手法の 1 つである YOLO v2 を用いて、動画像からダイコンサルハムシの成虫と幼虫の検出・識別を行った。しかしながら、過検出が多く、特に農作物の葉の食害の後を虫領域であると多く判断された。よって実環境下で撮影された病虫害画像に対して検出・識別を行う際には、対象である領域が微小である、かつ病虫害と背景が類似または病虫害同士が類似していることが、課題であることを確認した。

本稿では、上記の課題に対して、1 つの CNN のみでの検出・識別には限界があると考え、検出と識別を 2 段階に分けて行うことを提案する。本研究では、基盤技術獲得のための第一歩として、YOLO v3 で検出した虫候補領域に対して、Xception で病虫害であるキスジノミハムシ、ムモンキスジノミハムシ、背景の 3 クラスで識別を行う。

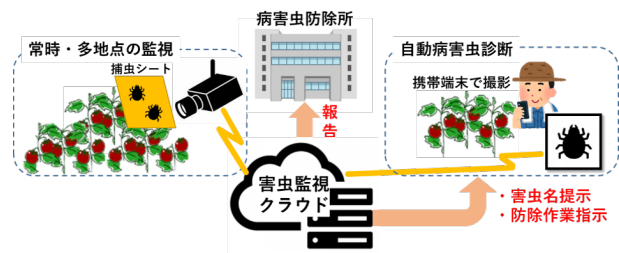


図 1: 本研究を基盤としたスマートアグリ例

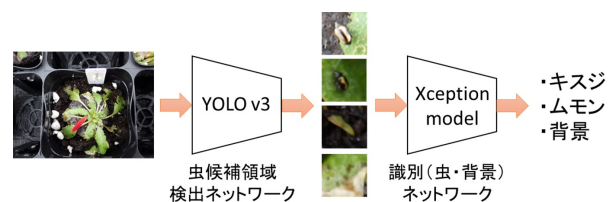


図 2: 提案手法の概略図

2 提案手法

先行研究 [2] の結果を受け、YOLO 単体での検出・識別精度向上には限界があると考えたため、YOLO v3 の過検出を減らすことを目的として、図 2 に示すように虫候補領域の検出を行う検出器に YOLO v3、候補領域をキスジノミハムシ、ムモンキスジノミハムシ、背景の 3 クラスに識別する識別器に Xception を用いた構成を提案する。

2.1 YOLO v3 を用いた虫候補領域の検出

YOLO v3 は single shot の物体検出手法の 1 つである YOLO (You Only Look Once) の最新モデルであり、YOLO は Redmon らによって提案された高速な物体検出・認識アルゴリズムである [5]。通常の CNN では、最終層に全結合層を導入し、softmax 関数などを用いることによって画像の識別を行うが、FCN では最初から最後まで全てが Convolution 層のみで構築されており、特徴マップの精確な位置情報を保持したまま最終層まで伝播されるという特徴がある。YOLO, YOLO v2[4], YOLO v3 とモデルの改良が行われている。YOLO v2 が 19 層でモデルが構成されていたのに

対し、YOLO v3 は 53 層の Neural Network モデルを使用している。また、soft max 関数の禁止やアップサンプリングを行う等の変更が加えられている。これにより検出時間を若干落とし、精度の向上がなされている。

YOLO v3 が要求する入力画像サイズは 416×416 pix. であるが、本研究では 512×512 pix. に変更している。入力された画像に対して病虫害の候補となる領域を出力する。

2.2 Xception による虫候補領域の識別

Xception[3] はベンチマークテストである”Image Net”において VGG16 や InceptionV3 よりも精度の高い CNN モデルである。Xception では、InceptionV3 の Inception models の代わりに、Depthwise Separable Convolution を用いる。InceptionV3 等で用いられる通常の畳み込みでは特徴マップの空間方向とチャネル方向に対して同時に畳み込みを行う。これに対して、Depthwise Separable Convolution では Depthwise Convolution と pointwise Convolution を用いることで、特徴マップの空間方向の畳み込みとチャネル方向の畳み込みを分離させている。これにより Xception は、パラメータ数と計算時間を削減しており、その分モデルの深さや幅を大きくすることでモデルの性能を向上させている。図 3 に Xception の概略図を示す。本研究では学習の際、初期重みに”Image Net”を使用しファインチューニングを行っている。Xception の入力画像サイズは 299×299 pix. である。これに対して YOLO v3 が出力する虫領域候補画像はサイズが様々である。本研究では画像をアスペクト比を無視して 299×299 pix. にリサイズして入力する。入力画像に対して、キスジノミハムシ、ムモンキスジノミハムシ、背景を識別し、対応するクラスを出力する。

2.3 学習用画像生成

深層学習を用いた物体検出・識別を実現するためには、学習用画像が大量に必要となる。しかし、今回対象とする病虫害には大規模画像データセットが存在しないため、学習用画像を収集する必要がある。その収集コストが課題である。第三者によって撮影された画像に加え、インターネットを介して画像を収集したところ、学習に利用できる解像度の画像はわずかに 32 枚 (キスジノミハムシ 20 枚、ムモンキスジノミハムシ 12 枚) であった。

本研究では、収集した少数の画像をもとに画像処理技術を用いて大規模な画像データを生成する。具体的には、収集した病虫害の画像群から、個体領域のみを抽出することで前景画像を生成する。次に、虫が含まれない自然画像 (32 枚、平均画像サイズは約 4000×3000 pix.) を準備し、ランダムに選ばれた 1 枚の背景画像から YOLO v3 学習用画像生成の場合は、 512×512 pix., Xception 学習用画像生成の場合は 299×299 pix. でそれぞれ領域を切り出し、それを背景画像とする。背景画像に対して前景画像を画像処理により合成することで学習用画像を生成する。合成の際、以下に示す処理を適用する。

表 1: 学習用画像生成のパラメータ

生成枚数	3000 枚 (12000 枚)
前景物体のサイズ	10~70 pix. (200~250 pix.)
1 枚に含まれる虫の最大個体数	3(1)
拡大縮小率	80~120% (アスペクト比無視)
ガウシアンノイズのパラメータ σ	0~9
ガンマ変換のパラメータ γ	0.5~1.5
ガウシアン平滑化のパラメータ σ	1~3
ガウシアン平滑化のフィルタサイズ k	3~7

- 前景画像に対して拡大縮小を適用する
- 前景画像に対して左右上下反転、回転処理を加える
- 複数の前景画像を合成する場合、前景画像の重なりを許可する
- 合成した画像に対してガウシアンノイズを加える
- 合成した画像に対してガンマ変換により明るさを変化させる
- 合成した画像に対してガウシアン平滑化を適用する (50%の確率で適用する)

表 1 に使用した画像処理のパラメータを示す。表の数値のうち括弧内の数値は Xception の学習用画像生成のパラメータを表す。実際に作成した画像を図 4 に示す。



図 4: 学習用画像の例 (左:YOLO v3 用, 右:Xception 用)

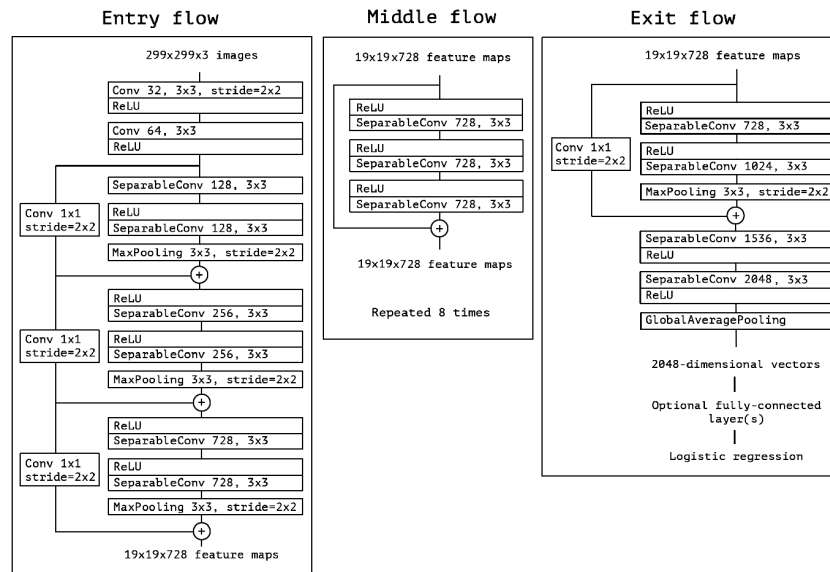


図 3: Xception 概略図 [3]



図 5: 原画像を重複部分有りで 8 分割した例

3 実験・考察

評価実験として実際に動画からキスジノミハムシとムモンキスジサルハムシを検出・識別する。実験に用いる動画は、実環境で栽培されている農作物に付着した病害虫を撮影したものである。動画のフレーム数は 478 フレームで、延べ 808 個体 (キスジノミハムシ 458 個体, ムモンキスジノミハムシ 350 個体) の病害虫が撮影されている。取得された画像を YOLO v3 の要求サイズに合わせてリサイズを行うと虫領域が縮小されることから、縮小率を軽減するためにスライディングウィンドウ方式を用いる。図 5 に示すように、画像を重複ありの 8 分割して 1 枚あたり 800×800 pix. にしたのち、YOLO v3 の要求サイズにリサイズし、入力する。本研究では、次に示す 2 つの実験を行い提案手法の有用性を確認する。

- 実験 1: YOLO v3 を用いた検出・識別
- 実験 2: 実験 1 で検出された全領域について Xception を用いた再識別

表 2 に実験 1 の結果を示す。YOLO v3 が出力した 1088 個の結果の内、正しく検出・識別が行われた結果が 191 個、背景領域を虫領域と誤検出した結果が 818 個、虫の誤識別が 79 個であった。実験結果の約 8 割

が背景領域を虫であると判断していることから、過検出であるといえる。

表 3 に実験 2 の結果を示す。実験 1 と同様の 1088 個の結果の内、正しく識別が行われた結果は 585 個 (病害虫 143 個, 背景 442 個) であった。一方で、虫を背景と誤識別した結果が 85 件、背景を虫と誤識別した結果が 376 個、虫の誤識別が 42 個であった。

また、実験 2 における成功例を図 6 から図 13 に、失敗例を図 14 と図 15 示す。成功例では、背景領域を YOLO v3 が病害虫であると誤識別した結果に対して、2 段階目の Xception では背景領域であると正しく再識別できた例を示す、失敗例では、YOLO v3 が病害虫領域を正しく識別したにもかかわらず、Xception が背景領域であると誤識別した例を示す。図における青色矩形はキスジノミハムシ領域、赤色矩形はムモンキスジノミハムシ領域、緑色矩形は背景領域であることを示している。

表 2, 表 3 から、Xception の推論は、YOLO v3 が推論した虫候補領域の内、実際は背景であった領域を半数以上正しく識別しているため、本手法は有用であるといえる。

一方で、虫領域を背景と誤識別する結果も確認し、今後の課題である。特にムモンキスジノミハムシと背景の誤識別が多い結果となった。これは、ムモンキスジノミハムシの全体的に黒いという特徴が、背景の農業用の土と特徴が著しく類似していることが原因であると考えられる。

表 2: 正解データと YOLO v3 の推論結果の比較
(縦: 正解クラス, 横: 予測クラス)

	キスジ	ムモン	背景
キスジ	114	40	0
ムモン	39	77	0
背景	396	422	0

表 3: 正解データと Xception 推論結果の比較
(縦: 正解クラス, 横: 予測クラス)

	キスジ	ムモン	背景
キスジ	120	14	20
ムモン	28	23	65
背景	255	121	442

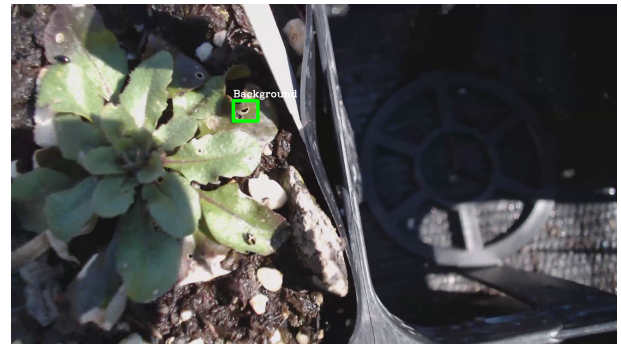
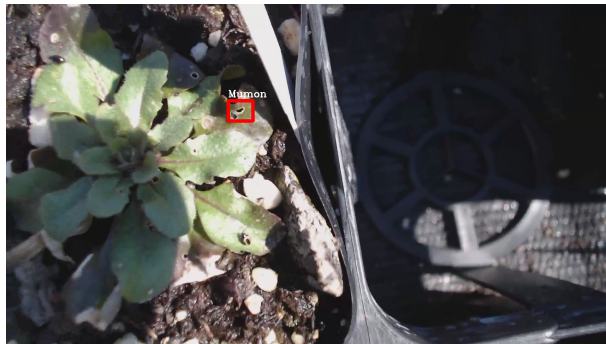


図 6: YOLO の過検出を Xception で訂正できた例 (左: 動画の 32 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 7: YOLO の過検出を Xception で訂正できた例 (左: 動画の 41 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 8: YOLO の過検出を Xception で訂正できた例 (左: 動画の 216 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 9: YOLO の過検出を Xception で訂正できた例 (左: 動画の 257 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)

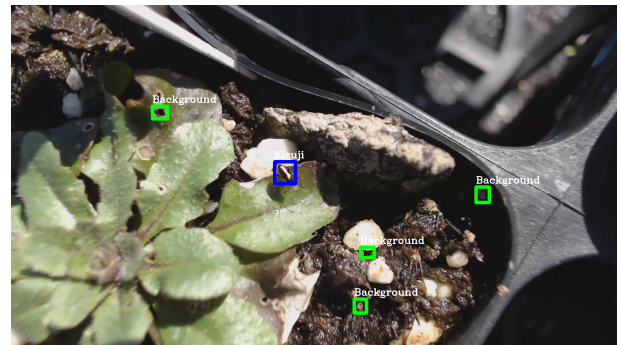


図 10: YOLO の過検出を Xception で訂正できた例 (左: 動画の 266 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 11: YOLO の過検出を Xception で訂正できた例 (左: 動画の 287 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 12: YOLO の過検出を Xception で訂正できた例 (左: 動画の 360 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)

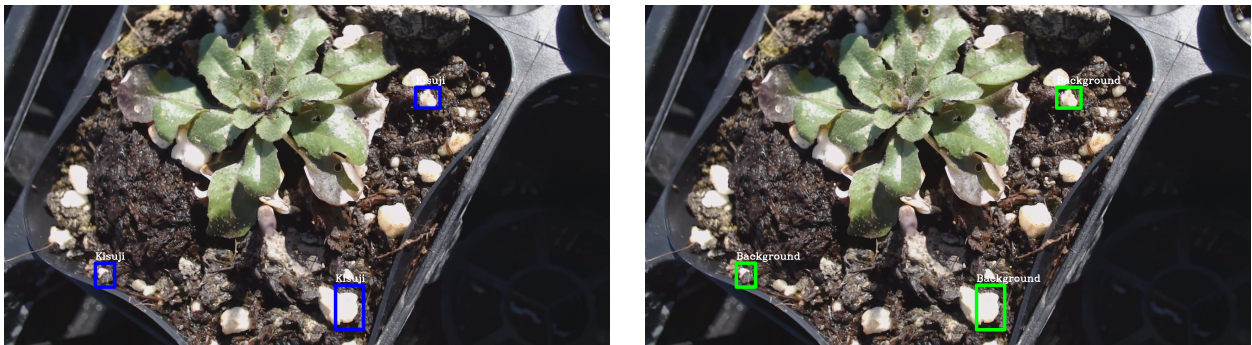


図 13: YOLO の過検出を Xception で訂正できた例 (左: 動画の 470 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 14: YOLOv3 が正しく識別した領域を Xception で誤識別した例 (左: 動画の 413 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)



図 15: YOLOv3 が正しく識別した領域を Xception で誤識別した例 (左: 動画の 423 フレーム目を YOLOv3 で検出・識別した結果, 右: YOLOv3 が出力した矩形領域を Xception で再識別した結果)

4 おわりに

本論文ではCNNを組み合わせて、カメラで撮影された病害虫の検出・識別を行った。微小かつ外見が類似する対象に対して、2段階の検出・識別を行うことで結果が改善した。今後は、改善されなかった部分を検討し、さらなる精度向上につなげたい。

参考文献

- [1] 岡部亜梨子, 安倍満, 小沢慎治, “AbS手法を用いた店舗内における人物の検出”, 電気学会論文誌C (電子・情報・システム部門誌), 127 巻 4 号 pp.506-512, 2007.
- [2] 葛原宏章, 滝本裕則, 金川明弘, 佐藤安弘, “深層学習に基づく動画像からの病害虫の検出”, 第20回IEEE広島支部学生シンポジウム (20th HISS) 論文集, A2-5 pp.105-107, 2018.
- [3] F.Chollet: “Xception: Deep Learning with Depth-wise Separable Convolutions”, arXiv:152.02357, 2017.
- [4] Redmon, Joseph, and Ali Farhadi: “YOLO9000: better, faster, stronger.”, arXiv preprint arXiv:1612.08242, 2016.
- [5] J Redmon, A Farhadi: “YOLO v3: An Incremental Improvement”, arXiv preprint arXiv:1804.02767, 2018.