

植物病害自動診断に向けた葉領域検出および 病変認識手法の提案

川崎, 雄介 / KAWASAKI, Yusuke

(発行年 / Year)

2017-03-24

(学位授与年月日 / Date of Granted)

2017-03-24

(学位名 / Degree Name)

修士(工学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

平成28年度

修士論文

植物病害自動診断に向けた葉領域検出
および病変認識手法の提案

指導教員

彌富 仁

法政大学大学院 理工学研究科 応用情報工学専攻

15R4112 川崎 雄介

目次

論文要旨	3
第1章 背景	4
1.1 研究の背景	4
1.2 研究目的	5
1.3 本論文の構成	7
第2章 関連研究	8
2.1 はじめに	8
2.2 物体認識・検出に関する先行研究	8
2.2.1 物体認識の概要とその難しさ	8
特徴量ベースの物体認識	9
近年の物体認識手法	10
2.2.2 物体検出の概要	11
top-down approach	11
bottom-up approach	12
integrated approach	13
2.3 植物病害自動診断に関する先行研究	13
2.3.1 旧来の手法	13
2.3.2 近年の潮流	14
2.4 本研究の新規性	15
第3章 物体検出部の開発と実験	16
3.1 integral channel feature	16
3.2 葉検出機構	19
3.2.1 AdaBoost の概要	19
3.2.2 AdaBoost の学習	20
3.2.3 attentional cascade	21
3.3 評価実験	25
3.3.1 データセットの取得	25
3.3.2 実験条件	27

3.3.3	実験結果	27
第4章	病変識別部の開発と実験	29
4.1	畳み込みニューラルネットワーク (CNN)	29
4.1.1	CNN の各層における処理	29
	畳み込み層	30
	活性化関数	30
	プーリング層	31
	局所正規化層 (local contrast normalization)	31
	出力層	32
4.1.2	CNN の学習方法	33
4.2	植物病害分類実験	34
4.2.1	データセットの取得	34
	病害一覧	34
	撮影環境の差異について	36
4.2.2	CNN の設計	36
4.2.3	評価実験	38
	前処理および評価方法	38
	実験 A. CNN の識別能力の検証	40
	実験 B. 好条件下/悪条件下での画像における検証	40
第5章	考察	42
5.1	検出部に関する考察	42
5.2	葉識別部に関する考察	43
第6章	まとめ	45
	謝辞	46
	参考文献	47

論文要旨

農業現場において、植物病害の早期発見は特に重要な問題である。しかしながら、そういった病害発見は専門知識を要する人間の目視によって行われる場合がほとんどであるため、人的、金銭的コストがかかっているのが現状である。このような問題を解決するため、様々な工学的アプローチによる病害診断システムが提案されてきた。しかしながら、いずれの先行研究も、特定の病変にしか対応できない、実際に農場現場において活用するという観点で考慮した場合応用が難しい等の問題点があった。そこで本研究では、農業現場において、植物病害の早期発見を、農業現場に実際に活用できる形で実現するために、画像認識技術を利用し各種検討を行った。具体的には、植物病害の早期発見というタスクを、「畑から葉領域の自動検出」「植物病の自動診断」という2つのパートに分けて、各種研究開発を実施した。前者については、実際に農場現場に赴き、定点カメラをビニールハウスへ取り付けた場合を想定した画像を用いて、葉領域検出器を作成した。この検出器は多段構造を持ち、葉ではない箇所領域の早い段階で除去を行うことによる検出性能の高速化や、葉同士の重複に対する頑健性を向上させるような工夫を取り入れた。後者については、実際の農場現場において問題となっている植物病害に着目し、きゅうり葉を対象とした7種類+健全葉計8種類の識別問題について、光や影の影響のある画像に対しての平均識別率80.6%を達成した。

第1章 背景

1.1 研究の背景

農作物の品質管理は、農業産出額 950 兆円を誇る我が国にとって、経済活動を支えるための重要な課題の一つである [1]。農作物を市場に安定して供給するためには、農業従事者による圃場の適切な管理が必要であり、その管理領域は圃場環境の確認、気候や湿度などに基づいた水や農薬の適切な使用、農具の衛生管理など多岐にわたる [2]。その管理領域の中でも、作物の商品価値低下の直接的な原因となる、植物病害の早期発見は特に重要である。とりわけ植物ウイルス病害については、同じ圃場にある他の作物への蔓延による 2 次被害を引き起こしてしまう他、感染株への治療が行えないことから、現場において非常に憂慮されているものである。大島によると、そうしたウイルス病害による被害額は年間 1,000 億円を超えるとも推定されている [3]。

以上のことから、農業生産の現場において、植物病害を早期発見する方法が模索されてきたが、現状その方法は確立されているとは言い難い。最も一般的な方法は、農業従事者や植物病理に関する専門知識を持った人間による目視診断である。これは、診断者が感染の疑いのある植物の株をひとつずつ目視により観察し、植物の葉や茎の様子を勘案して病気の種類を決定する手法である。しかしながらこの方法は、圃場全体を診断者がひとつずつ判断しなければならないため、非常に手間を要する。また目視診断には、葉や茎、根等に発現する異変を早期段階で察知するための高度な知識が要求される。植物に出現する多くの病徴は、ある程度の経験が無いとその有無や種類を見分けることが困難であり、とりわけ初期段階の病徴ではほとんど健全なものとの区別がつかない。こうした事実から、一般の人間が病気の有無を判断できるようになる頃には、すでに病徴が末期段階まで進行し、手の施しようがない状態となっているケースが多い。このような実例を以下に Fig. 1.1 として示す。



Fig. 1.1: ウイルス病害 (MYSV) 感染葉の変遷 (左から初期, 典型, 末期病徴)

目視診断以外の方法として、植物から取得できる遺伝子情報を利用した方法や、顕微鏡を用いた方法などが提案されている。遺伝子診断法の代表的な手法に、LAMP法を利用した方法がある [4]。これはウリ科、トマト科の植物に対して有効な手法であるが、この診断には専用のキットが必要であるほか、診断できるウイルスの種類に限りがある。さらに、このキットも農業従事経験の浅い農家、もしくは専門知識の無い一般の人間にとっては使用が難しい。顕微鏡を利用した方法に関しても、ある程度の専門知識を以って成り立つ手法であるため、根本的な解決とはなりえない。

こういった背景から、専門知識を有する農家の人間や植物病の専門医が、広い圃場の中を目視で見て回り、一株一株異変の有無を確認していくという作業が行われているのが現状である。

1.2 研究目的

前項で述べた通り、現在の農業現場における植物病害の発見から診断にわたる一連のプロセスは、すべて農業従事者および専門家の勘や経験に拠って実施されているのが現状である。この事実は農業従事者の大きな負担となるのみならず、ウイルス病のような病害が伝播する病害であった場合、見落としによる被害額の拡大という可能性も孕んでいる。そこで本研究では、これらの問題に対し画像処理および機械学習という工学的観点からアプローチし、一連のプロセスを画像処理・機械学習技術を用い自動化することによって、農業従事者の負担を軽減することを大きな目的とする。

植物株の中で、病害が発現する箇所は葉、茎、根等多岐にわたる。その中でも葉は病徴診断の手がかりとしてしばしば用いられている他、カメラ等での観測が茎や根と比較して容易である。そこで本研究では、病徴が葉に出現する作物を対象とし、農場全体を撮影した画像から、ある葉に発現した病害を検出するというシステムを提案する。

本システムは、「葉検出部」「葉認識部」の2点から構成される。具体的には、農場全体を撮影した画像中から、葉の領域を推定し、葉領域と推定された箇所に対して、その葉の病変の有無を検知するという枠組みになっている。

以下に、本研究の入力から出力に至る一連の流れを Fig. 1.2 として示す。

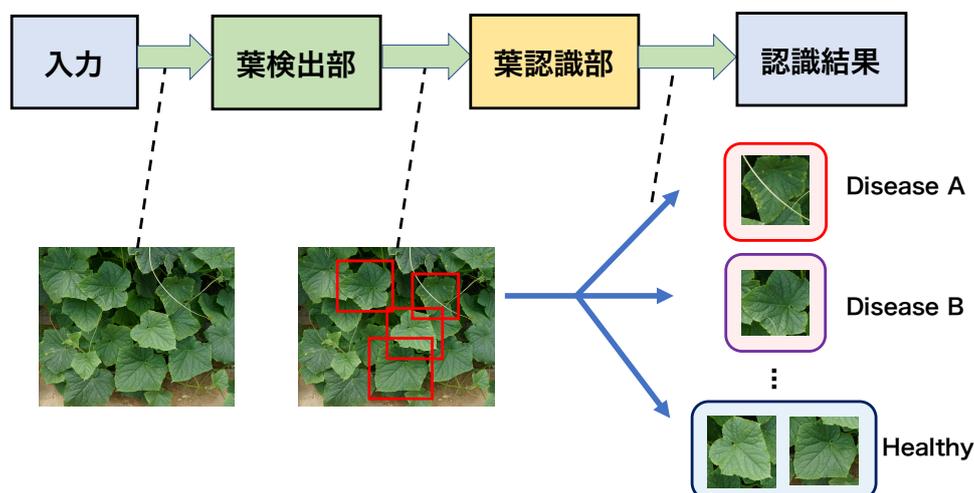


Fig. 1.2: 病害診断までの一連のプロセス

本研究では、対象作物をキュウリと定めて種々の実験を行った。日本の農業現場において、キュウリの植物病害はとりわけ顕著な問題となっており、キュウリ農家は対応に苦慮している。この大きな理由として、キュウリに感染する病害の種類が幅広く、その感染源によって農家の対策が異なることが挙げられる。

例として、非ウイルス病害とウイルス病害の対策の違いを示す。あるキュウリの株に感染した病害が、べと病などの非ウイルス病害である場合、その株に対して農薬を散布することで一定の対策を講じることが可能である。一方で、キュウリウイルス病害の場合は早急に感染株を除去しなければならない。これは、感染後の株に対する根治が不可能である上、他の株へ病害が伝播するためである。即ち、農業従事者は広いキュウリ農場を定期的に観察し、病変と思しき株を発見した場合、その原因を速やかに特定し対応しなければならない。キュウリ植物病害の対応が困難である背景にはこうした事実があり、この問題点が解消されることによる恩恵は非常に大きい。

またキュウリは、日本における産出額がトマト、イチゴに次ぐ第3位となっており、日本の総農業産出額という観点から俯瞰しても、影響の大きな野菜であると言える [5]。

こうした事実から、キュウリは日本で栽培される農作物の中でも、本システムを用いて診断プロセスを自動化する意義が大きいと判断できるため、キュウリを対象の作物として選択した。

本システムが実現することによって、

- 農業従事者の負担軽減
- 病変への迅速な対策の実現
- 収量増加に伴う収益向上

といった効果が期待できる。

1.3 本論文の構成

本論文の構成について大まかに述べる。本章では、本論文の背景および目的に関して述べた。第2章では、本研究に関連する先行研究について、物体認識および検出系、植物病害診断系の2つの系譜から述べ、本研究の特徴や利点を述べる。第3章では、今回提案する葉検出器の詳細および、実験結果を示す。第4章では、先行研究の認識システムの仕組みを述べ、その実験結果を提示する。第5章で全体を俯瞰した実験および考察を行い、第6章で本論文の結論と今後の展望について述べる。

第2章 関連研究

2.1 はじめに

2012年の一般物体認識コンテストにおいて、convolutional neural network(以下 CNN)が他の手法を凌駕する認識精度を達成して以降、物体検出、物体認識、またその応用に関する論文は激増し、2017年現在も日進月歩で新たな技術が研究・開発されている。

本項では、一般/特定物体認識の先行研究および、それらに追隨して実施されてきた植物病害自動診断に関する先行研究を紹介する。また、それら多種多様な研究の中での本研究の立ち位置を明確化し、特徴や有用性を示す。

2.2 物体認識・検出に関する先行研究

2.2.1 物体認識の概要とその難しさ

物体が写った画像をコンピュータに提示し、それが何であるか、あるいはどういった状況であるかなどの一定の判断を返させる問題を物体認識問題と呼ぶ。

物体認識問題は、大きく分けて一般物体認識と特定物体認識に大別される。前者は、計算機が実世界シーンの画像に対し、その中に含まれる物体が何であることを一般的な名称で認識するタスクである [6]。一方後者は、インスタンスレベルで同一物体であるかどうかを判定するタスクである。具体例を示す為に、Fig. 2.1 に示す画像について、ある領域内に何の果物が存在するかという判断を返す問題を考える。



Fig. 2.1: 例画像

一般物体認識では、画像中に存在するすべての果物に対して、各々「りんご」「キウイフルーツ」といったラベルを正確に付与することを目標としている。一方特定物体認識では、Fig. 2.1におけるクエリとして「キウイフルーツ」が与えられた場合、領域内を見て「この領域内に存在する物体はキウイフルーツか否か」を返すことを目標としている。

これら物体認識問題の一般的な処理は、以下の一連の流れに沿って実施される。

1. 前処理

入力された画像に対して、ノイズ除去、正規化などの処理を施すことによって、輝度値の差による影響などを補正する処理である。また、画像中から物体の領域を抽出する場合などもここに含まれる。

2. 特徴量の抽出

前処理後の画像から、その物体を表現する数値データを取得する処理である。一般的にデータはベクトルとして表現され、画像中の物体の色、形、テクスチャ、画像全体にわたる統計特徴など様々な種類があり、中でも統計特徴は今日に至るまで非常に多くの抽出法が提案されている [7]。

3. 識別

抽出された特徴量を、任意の識別器に入力することによって、一定の判断を得る処理である。物体認識問題の場合、その物体の同定や所属カテゴリ、その物体の状況を説明した文章などが返値となる。

物体認識の難しさは、その見た目 (appearance) の変動によるところが大きい。画像中に存在する物体はその撮影環境によって、大きさ、形状、照明条件などが様々に変動する。果物の例であれば、Fig. 2.1のように複数の果物同士が重なり合って、果物の一部が見えなくなってしまうたり、同じ画像中に存在する別の物体の影の影響を受けてしまったりするケースが考えられる。極端に難しい例としては、カットフルーツのように、対象の果物があたかも全く別の状態で提示されているようなケースも考え得る。

人間であれば、対象の物体がどんな形状をしていようと、過去の膨大な経験則に基づいてその物体を同定することが可能であるが、計算機の場合それは困難である。換言すれば、カットされた果物の種類が何であるか人間が判断できるのは、人間が知識や経験に基づいて果物の断面の様子を知っているからであり、そういった事前知識のない計算機が判断することは難しい。こういった点に物体認識の難しさがある。

特徴量ベースの物体認識

CNN 登場以前は、物体認識を行うために「いかに優れた物体表現を作成するか」ということに重点を置いた研究が多く行われてきた。前述の通り、計算機に入力される物体は、そ

の撮影環境に応じて大きさ、形、明度条件などが大幅に異なる。この問題を解決するためには、物体の本質的な情報を数値的に表現する手法が必要となる。

Lowe が提案した SIFT (scale-invariant feature transform) [8] は、その頑健性から物体認識分野において波及した特徴量記述法である。この手法は、エッジやコーナーなど、物体と背景を分離しうるような箇所が大きな輝度勾配を持つことを利用している。Lowe はそのような物体毎に特徴的な点を keypoint と呼び、keypoint 周囲の輝度強度や輝度方向を 128 次元のベクトルとして記述した。こうしたステップを踏むことによって、物体の拡大縮小、回転に対して頑健な特徴記述子を実現した。

SIFT が持つ位置情報を捨て、統計学的な手法へと昇華させた手法が、Csurka によって提案された bag-of-visual-words (bag-of-keypoints, bag-of-words などとも呼ぶ、以下 BoVW) [9] である。彼の提案したこの特徴を用いた実験では、顔、車、電話など 7 クラス分類問題に対し平均分類性能 78.0% を達成している。この特徴量は、SIFT で得た局所特徴をベクトル量子化し、ヒストグラム化したものである。ベクトル量子化は、局所特徴の類似性に基づくクラスタリングによって実施される。クラスタリングを行った際、最終的にセントロイドとなったベクトルは “Visual Words” と呼ばれ、画像中の物体、あるいはその物体の背景などを端的に表現したベクトルと見なされる。画像ごとに、そのベクトルに類する成分をヒストグラムとして表現することで、その物体にエッジやコーナーなどの情報がどれだけあるか、という情報を含んだ特徴量が算出できる。

BoVW は、その汎用性から多くの理論的研究や応用がなされ、一般物体認識の特徴量記述子として広く利用されてきた。BoVW を取得する際に、クラスタ周辺の 1 次、2 次モーメントを用いた手法が Fisher vector (以下 FV) である。この手法は、BoVW でベクトル量子化を行う代わりに、クラスタリング結果の分布を gaussian mixture model として解釈し、特徴量設計を行う手法である。本手法は 2007 年に提案され、2010 年に L2-normalization 等による性能向上が報告されている [10, 11]。メモリ削減の為に、FV をよりコンパクトに表現した手法が Jégou らの vector of locally aggregated descriptors (以下 VLAD) である。彼らは特徴ベクトルの表現に、最近傍クラスタのセントロイドとの誤差の総和を求めることによって、コンパクトな表現を実現している [12]。

こうした特徴量群は、人間の視覚特性に学んだものや、物体が持つテクスチャなどの特性を数式に表現することによって実現されたものであり、その理論背景は、現代の一般物体認識手法を解析するための礎となっている。

近年の物体認識手法

コンピュータビジョン界隈において、大きなパラダイムシフトが起こったのは 2012 年である。この年に開催された大規模一般物体認識コンテスト [13] において、Krizhevsky らの提案したモデルが旧来の手法を大幅に達成する認識性能を達成した。Krizhevsky らは、5 層の畳

み込み層，3層の全結合層から成る CNN を識別器として提案し，過学習抑制のための画像枚数の水増し（data augmentation）等の工夫を加える事によって，top-5 error rate 15.3%を達成した [14].

この物体認識コンテスト以降，CNN を用いた物体認識モデルが多数提案されるようになり，またその応用研究も加速度的に増えていった．同物体認識コンテストではそれまで下火であった neural network ベースのアプローチが SVM に代わって数多く検討され，毎年のように error rate が更新されていった [15–17]. 2015 年のコンテストでは，人間の error rate を超えるという結果も報告されている．CNN の応用研究も，動画像への応用，自然言語処理と組み合わせた状況説明など数多い．また，CNN は識別に有効な特徴量を学習過程で獲得するようなモデルであるため，特徴抽出器としても非常に優秀な性能を誇る．定量的な議論は，Dosovitskiy らが実施している [18]. 彼らは CNN で得られた特徴を逆変換することで，CNN が学習過程で得られている特徴を可視化し，その有益性について議論した．

しかしながら，こういった識別性能を実現するためには，高級な計算資源 (GPU) や多量のデータセットが必要不可欠となっており，一般家庭等で利用するためには物理的制約がいまだ大きいことが現状である．その為，如何にコンパクトなモデル，限られたデータで高い識別性能を実現するかが，今後のコンピュータビジョンにおける課題となると考えられる．

2.2.2 物体検出の概要

コンピュータにある物体が写った画像を提示し，あらかじめ与えた情報をもとにその物体を探し出す問題を物体検出問題と呼ぶ．現在，物体検出問題には大きく分けて2つの方法論が存在する．一方は，sliding window を用いて総当り的に対象画像中から目的の物体を探し出すトップダウン的な手法，もう一方は対象画像をあらかじめ小領域に分割し，その小領域の Texture 類似度や Saliency 等を用いて，物体領域とそうでない領域に分けるボトムアップ的な手法である．この小項目では，この2種類の手法から先行研究を概観する．

top-down approach

物体検出で一般的なアプローチの一つが，検出対象の物体が写った画像中で sliding window と呼ばれる一定サイズの矩形領域を走査させ，その window 内の物体を判定する方法である．具体的には，window 内から取得した特徴量を，あらかじめ同一の特徴量を用いて学習させておいた識別器に入力することで判断を返す処理を逐一行うようなアプローチとなる．

物体検出において，顔をターゲットとしたものは数多く，とりわけ Viola らの Haar-like 特徴は有名である [19]. これは，人の顔のパーツが作り出す局所的な明暗差を利用した特徴である．彼らはそれを，単純な白黒の矩形を用いて定量化し，さらに候補領域の早期棄却を実施する枠組みを利用することで，高速な顔検出を実現した．この Viola らの検出モデルは，長らくの間顔認識に関するベンチマークとなっている．

人の検出に関して大きな影響を与えた特徴量は、Dalalらによって提案されたHoG(histogram of oriented gradient)特徴量である [20]. HoG特徴量は、ある局所領域(フィルタ)ごとに学習用画像の輝度強度と輝度勾配を算出し、その輝度勾配のヒストグラムを特徴量とする手法である。

本手法は、勾配差に着目した特徴量であるため、照明変動や物体のオクルージョンなどに対してロバスト性がある。それ故、荷物や群衆等で一部が隠れている画像に対しても検出が可能というメリットがある。彼らはHoG特徴量とガウシアンカーネルを用いたSVMを用いて、高精度な人検出を実現している。

前述のHoGをさらに発展させたモデルとして、deformable part model(以下DPM)がある。このモデルは、画像に写る物体が、物体の大まかな形のみならず、そのパーツの部分集合でも表現できるという点に着目している。例えば、人の全身をHoG特徴量を用いて取得することを考える。人は頭、腕、胴体、足、といったパーツの集合からなるため、同サイズのフィルタを用いて特徴を取得すると、高解像度の画像では各パーツの、低解像度の画像では人の全身のHoG特徴量を取得できる。彼らはこのことを応用し、マルチスケールなフィルタと、その位置や解像度情報を用いた評価関数を提案し、強力な物体検出を実現した [21].

bottom-up approach

このアプローチは、画像を画素集合としてみなし、一定の基準をもとに画素をグルーピングしていくことで、物体の位置をpixel単位で定めるアプローチである。画素のグルーピング結果(superpixel)の外接矩形を一般的には物体候補領域(object hypotheses)と呼び、検出結果とは区別される。すなわち、あくまで「候補領域」であるため、最終的な検出結果はNNやSVM等の任意の識別器に入力し、クラス分類問題を解いた結果となる。

グルーピングする方法には、領域内類似度などをベースにした分割型の手法、クラスタリング等を用いる凝集型の併合手法等があるが、本項ではグラフ理論に基づくアプローチに着目する。

このグラフ理論に基づくアプローチは、画像を無向グラフとして解釈する点が肝である。具体的には、画素集合をノード、画素間の輝度値の差をエッジに持つ無向グラフとして解釈し、各アルゴリズムでさだめられた条件の下、エッジの分割や併合を繰り返し行うような手法である。Felzenszwalbらの手法 [22] では、領域内類似度および領域間類似度という指標を用いて、画素を小領域群へと併合している。

Felzenszwalbの手法を拡張し、後述するRegion CNN [23]にも用いられた手法がselective searchと呼ばれる手法である [24]. 具体的には、色、テクスチャ、面積、外接矩形のサイズの計4指標を隣接した小領域同士で計算することによって、その出力が一定のしきい値を超えた場合のみ併合する、という処理を行っている。Uijlingらは、selective searchに加え、物体検出への応用方法も提案している。これは、ground truthとなるbounding boxの位置と、

候補領域の中で物体の領域と一部だけ重なった bounding box の 2 クラス分類問題を解くことで実現している。

integrated approach

近年は、CNN を用いて物体の検出から認識まで一律の枠組みで実施する手法が非常に増えている。

その枠組みのベースとなる研究が、Sermanet らによる OverFeat [25] と、Girshick らによる Region CNN である。Sermanet らによる手法は、CNN ネットワークひとつで、物体検出～認識を行う手法である。彼らは与えられた画像に対し、ラスタスキャンによって多数の物体候補領域を生成することで物体認識精度を、bounding box の座標の回帰を用いることによって検出精度を向上させ、一律のアーキテクチャによる物体検出を実現した。なおこの手法は、2013 年度の ILSVRC localization 部門で優勝している。Girshick らによって提案されたパイプラインは、入力画像に対し前述した selective search を施し、作成された物体候補領域を CNN に入力して、取得した特徴を SVM の学習に用いるというものである。この手法は、一般物体認識に対する CNN の特徴の有益性を示唆している。なお、selective search で算出された領域は一般に物体の大きさに依存するため、スケールや領域の形状に対して些か問題があった。しかしながら、後にこの問題は He らの提案した spatial pyramid pooling という手法によって解消される [26]。この手法をベースに提案されたのが、Girshick による Fast R-CNN、Ren らによる Faster R-CNN である [27, 28]。前者は RoI pooling というレイヤーの導入と、識別用ネットワークと回帰用ネットワークを分けることによる高速化、Faster R-CNN は、region proposal network による Fast R-CNN の作り出す候補領域の総数を削減による更なる高速化を実現している。

最近では region proposal network などの枠組みをなくし、画像入力のみで bounding box の回帰、内部の物体の認識等を実施するネットワークの提案も見られる。

2.3 植物病害自動診断に関する先行研究

画像認識技術を利用した、植物病害の診断方法も数多く提案されている [29, 30]。植物病害の診断を工学的に解釈すると、病変の種類によるクラス分類問題と捉えることが出来る。多くの手法では、2.2.1 で示した流れに従った、前処理、特徴量の抽出、識別という 3 つのステップで病害の診断が行われている。

2.3.1 旧来の手法

工学的観点からの植物病害自動診断は、濃度共起行列 (gray level co-occurrence matrix, 以下 GLCM) や、空間グレーレベル依存行列 (spatial gray-level dependence matrix, 以下

SGDM) を用いて特徴量の設計を行い、任意の識別器を用いて識別を試みる、というフローで実施されているものが多く見られる。

Huang は、GLCM を特徴量として学習させた neural network による識別を試みた。彼はランの葉に出現した軟腐病等合計 3 種類の病変識別を実施し、平均識別性能 89.6% を達成した [31]。Kai らは、トウモロコシの葉に対して病変識別を試みた。彼らはトウモロコシの病変葉画像を用いて、紋枯病など 3 種類の病変を neural network を用いて 98% 以上の精度で診断することに成功した [32]。Bashish らが提案したものは、植物の種類によらない診断フレームワークである。彼らは HSV 色空間上で取得した SGDM を利用して特徴量を抽出し、3 層の neural network を識別器として利用することによって、延べ 5 種類の病変に対する平均識別性能 89.5% を達成した [33]。Arivazhagan らは、SVM を用いた識別手法を提案している。彼らは葉の HSI 画像から SGDM を算出し、各々の植物に対しその特徴を用いて学習させた SVM を試作した。この識別器はバナナやグアバなどの多種類の植物病に対して、平均識別性能 87.6% を達成した [34]。

2.3.2 近年の潮流

近年は、計算機の高性能化および低価格化、画像認識技術の飛躍的進歩、参入障壁の低下などの背景から、農業分野にも広く CNN などの先鋭的な手法が使用されるようになっていく。植物葉の「同定」という観点では、Lee らの手法が高性能である [35]。彼らは 44 種類の植物葉から CNN を用いて特徴抽出を行い、multi-layer perceptron を識別部に据えることによって 99.5% の精度での識別に成功している。Hall らの研究では、部分的に隠された植物葉の同定問題に対し、SIFT 等の hand-crafted 特徴や、CNN の学習過程で作成された特徴量を用いて random forest を学習させ、97.3% の識別性能を達成した [36]。

また、植物病診断に CNN を用いる方法も近年提案されている。藤田らは、農場で撮影されたきゅうり葉の病変識別に対し、CNN を利用することによって、8 クラスの病変識別問題に対して平均識別率 83.2% を達成した [37]。Mohanty らは、CNN の植物病診断への有用性を、複数の CNN アーキテクチャを用いて検証している [38]。彼らは 14 種類、26 病徴、50,000 枚以上の植物病感染葉が収録されたデータセットを用いて検証実験を行い、学習枚数が 10,000 枚程度でも 98.2% 以上の精度が出ることを実証した。

2015 年末には、クラウドソーシングによって収集された植物病害写真の大規模データセット [39] も展開されており、農業分野、とりわけ植物病害分野に対する画像処理・機械学習技術の応用に対する市場への期待感は大きいと言える。

2.4 本研究の新規性

本章では、現在に至るまでの画像認識、検出アルゴリズムの発展やその応用に関して論じた。また、農業分野においても、旧来の手法のみならず、CNNなどの先鋭的な手法が利用され始めていることを述べた。

本研究の新規性は、「実際の農業現場での活用」という点である。前述したCNNによる植物病害の手法 [38] は、好条件下で撮影された大規模画像データセットを用いていた。しかしながら、農業現場で撮影される画像が、必ずしも好条件とは限らない点が問題点として挙げられる。例えば、実際の農業現場では、多くの場合多数の株が狭い範囲に植えられている関係上、葉の重なりの影響を非常に受けやすい。また、太陽による光や影の影響、風による葉の揺れなど、撮影環境が非常に多くの外的要因に左右される。

本研究では、そういった外的要因に依存しない植物病害自動診断システムの実現に向けた実験、議論を行う。

第3章 物体検出部の開発と実験

本研究では, integral channel features(以下 ICF) を学習に用いた, attentional cascade による葉領域抽出を提案する.

本手法の特長を以下に述べる.

- 高性能な計算資源が必須でない点. 2.2.1. 項でも触れたが, 近年の CNN を用いた物体検出技術は, GPU 等を用いた高火力コンピューティングを前提としたアーキテクチャを用いている. 本手法は基本的に CPU を用いれば計算可能なほどモデルがコンパクトである点が一つ目の特長である.
- パラメータ設計の容易さ. 一般に CNN を用いた物体検出は多数のパラメータおよびハイパラメータを用いて実現されることに對し, 本手法は学習時に設定するパラメータ数が少数で済む.
- 解釈可能性. attentional cascade の学習過程で使用される AdaBoost は, 一つの弱識別器に対して一つの特徴が割り当てられる. それ故, 識別に有効な特徴量がどれであったかの示唆を定量的に得ることが容易である点が挙げられる.

3.1 integral channel feature

integral channel features(以下 ICF) [40] は, 画像に對し線形・非線形の様々な変換を施し, 変換後の画像チャンネル内部の矩形領域の総和を計算することによって作成される特徴量である.

画像チャンネルの組み合わせによる特徴取得は, Tu が最初に議論した [41]. Tu は, 画像に對して Gabor filter および Canny edge filter を畳み込むことによって2通りのエッジ検出画像を作成し, その画像パッチを学習データに用いるという手法を提案している. ICF を用いた Dollár らの手法は, この Tu らの手法を拡張し, 物体検出タスクへ適用したものと解釈することができる.

ICF の大きな特長として, 特徴量計算時に, ほぼ制約条件なく様々な画像チャンネルから特徴抽出を行える点が挙げられる. これは, 検出対象の物体が類似した色, 形状を持つ場合, その特徴が強く現れるような任意のチャンネルが選択可能であることを意味しており, 実際に歩行者検出や顔検出において広く活用されている [40, 42]. ICF の制約条件については後述する.

本研究で検出対象としている植物葉もまた，“緑成分を強く持つ”，“植物の種類によってほぼ葉の形状が決定づけられる”等のアドホックな特徴を持つため，葉検出に対しても有効に活用であると予想できる。

以下本項では，原著論文 [40] に従い，記号を以下のように定義する。

- I : 入力画像
- C : 1次元のチャンネル画像
- $\Omega(\cdot)$: 任意のチャンネル生成関数
- $f(\cdot)$: 特徴ベクトル

これまで述べてきたとおり，画像中に存在するエッジやコーナーなどは，物体もしくは物体の境界を記述する上で重要な意味を持つ。本特徴量は，そういった画像中に含まれるエッジ，コーナー，色，輝度勾配といった様々な成分を一律の枠組みで取得することを目指している。

ICF は，

- 入力画像の変換
- 矩形領域内の合計輝度値の計算

という2つのステップで取得することが出来る。

- 入力画像の変換

ICF 取得までの枠組みでは，まず様々な種類の関数を用いて，色や形などを抽出，および強調したチャンネル画像を複数取得する。

入力画像 I に対して，変換関数 Ω を定義する時，あるチャンネルに属する画像 C は Eq. (3.1) に示す形で一般的に表すことが出来る。

$$C = \Omega(I) \quad (3.1)$$

例えば，画像 I に対しサイズ $h \times w$ のフィルタを畳み込む処理 $filter$ は，Eq. (3.2) で示せる。

$$C(x, y) = \sum_{j=0}^{h-1} \sum_{i=0}^{w-1} filter(x-i, y-j) \cdot I(i, j) \quad (3.2)$$

この Ω に対する制約条件は， Ω がシフト不変性 (shift invariance) を持つという点のみである。すなわち関数 Ω は，画像 I および I' が存在し，それらの内部にある物体 (=画素の集合) に位置的な関係性があれば，変換後の $C = \Omega(I)$ ， $C' = \Omega(I')$ の間にも同様の関係性があることを要求する。直観的には，画像内に存在する物体の位置や大きさ，および画像そのもののサイズを変化させることなく，違う空間に変換させるような関数であれば良いという

ことになる。ICFの強みは、この制約条件の緩さにある。シフト不変性を持つ変換には様々な物があり、色空間の変換、エッジの取得、二値化処理などが挙げられる。さらにこうした基本的な変換処理は、様々な画像処理ライブラリで実装されているため、数行のコードでこういった処理を行うことが出来る点もメリットとなる。

以下で Fig. 3.1 として図解する。

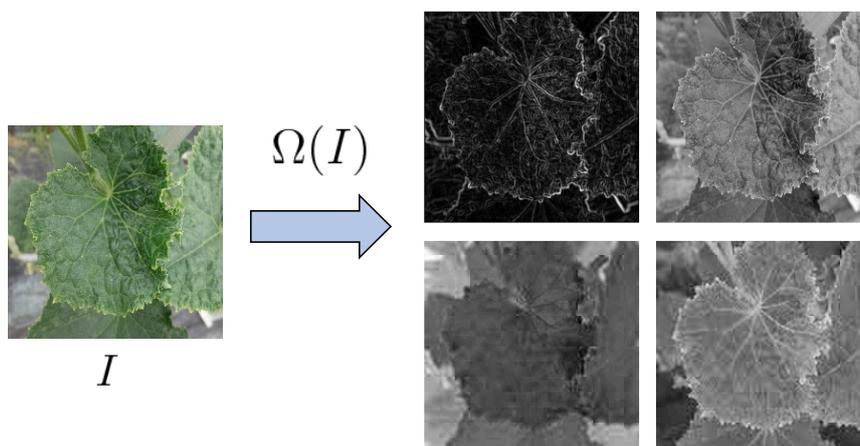


Fig. 3.1: 画像チャンネルの変換

- ・ 矩形領域内の合計輝度値の計算

このステップでは、チャンネル画像 C から、任意の矩形領域の合計輝度値 f_C を取得する。この合計輝度値が最終的に出力される特徴ベクトルの要素となる。

特徴ベクトルの要素を形成する矩形領域の取得位置、取得サイズなどはランダムに選択される。(著者らは、矩形領域のサイズは 25×25 px. 以上という制約を設けたが、 1×1 px. であっても性能に変化がなかったことが著者らの文献補遺にて示されている)

矩形領域の合計値は、積分画像によって高速に計算できることが知られている。積分画像は、各画素値を元画像の原点からの総和で定義した画像であり、以下 Eq. (3.3) で示される。

$$S(x, y) = \sum_{j=0}^{y-1} \sum_{i=0}^{x-1} C(i, j) \quad (3.3)$$

ただし、 C はグレースケール画像、 S は積分画像である。通常の方法で $M \times N$ の大きさを持つ領域内部の輝度値の総和を算出する為には、 $M \times N$ 回の加算が必要になる。しかし積分画像をあらかじめ算出しておくことによって、僅か3回の加減算で同じ結果を得ることが可能となる。以下に、積分画像で矩形領域内の総和を計算する際の概念図を Fig. 3.2 として示す。

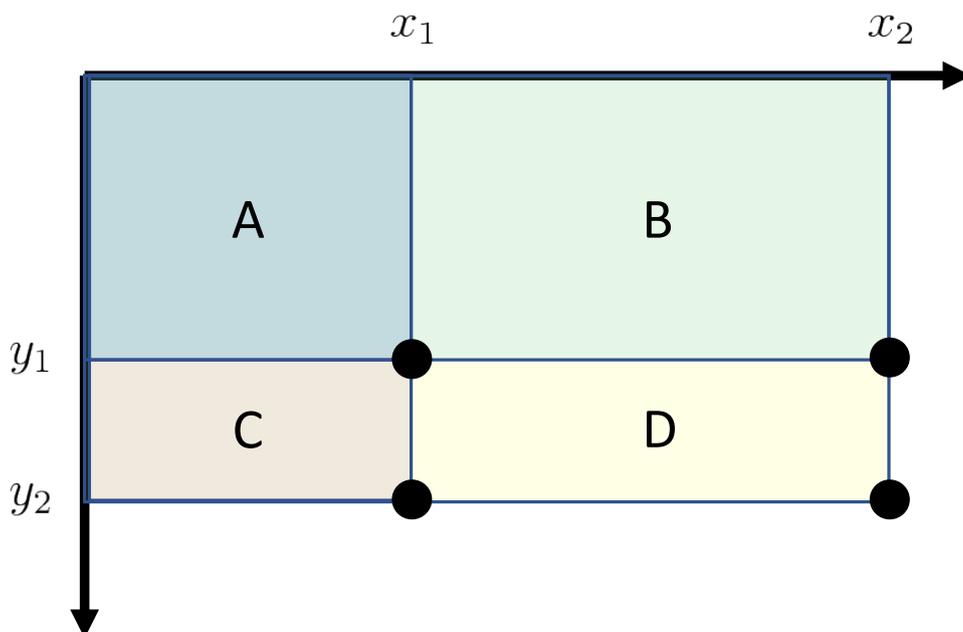


Fig. 3.2: 積分画像からの輝度総和の計算方法

例えば、上図積分画像において、領域 D の輝度値の総和を計算する際は、Eq. (3.4) を計算すれば良い。

$$D = S(x_2, y_2) + S(x_1, y_1) - S(x_2, y_1) - S(x_1, y_2) \quad (3.4)$$

以上の手順で計算された ICF 特徴は、後述する AdaBoost 検出器の学習に用いられる。

3.2 葉検出機構

本研究では、viola らが使用した attentional cascade を葉領域検出部として使用する。attentional cascade は、AdaBoost によって学習された弱識別器を直列に連結させることによって、非物体領域の早期棄却を実現している。

本項では、AdaBoost の学習アルゴリズムおよび特色、attentional cascade の学習過程について説明する。

3.2.1 AdaBoost の概要

AdaBoost (adaptive boosting) は、集団学習の枠組みの一つである。集団学習とは、識別性能の弱い識別器（以下、弱識別器）を複数組み合わせ、性能の良い識別器（以下、強識別器）を実現する手法である。任意の識別器を適用することが出来るが、多くの場合決定木を識別器として設定する。

代表的な手法として、bagging, boosting, random forest などがある。bagging は、学習用データ集合から重複を許してランダムに一定数サンプルを取得し、複数の弱識別器の多数決

をその識別結果の多数決を最終的な出力とする手法である。この時、複数の弱識別器は並列的に、独立に学習されるため、サンプルの取得の仕方によって識別器間の相関が高くなる可能性がある。random forestはこの識別器同士の相関を除去するために、決定木の学習後、識別時にランダムに決定木の枝刈りを行う手法である。本手法は汎化性が高い上、識別時にその識別基準が陽に示されることから、データ分析などにもしばしば活用される。

boostingは、識別器を独立に学習するのではなく、直列に学習するという点が上記2つとの本質的な違いである。具体的には、それぞれの学習サンプルに重みパラメータを設定し、前段の弱識別器が誤った結果を後段の弱識別器に反映させる枠組みである。AdaBoostはこのboostingの発展系である。アルゴリズムは以下で詳説する。

3.2.2 AdaBoostの学習

AdaBoostは、boostingの中でも、入力する特徴それぞれに重みを付与し、学習過程で適応的にサンプルの重みを変動させるような集団学習法である。具体的には、前の弱識別器（弱識別器群）の出力結果を受けて、正しく学習したサンプルの重みを小さく、後から学習したサンプルの重みを大きく変更する。この処理によって、識別器全体の学習プロセスが進行するほど、出力として誤り率の高いサンプルに対する影響の高い識別境界が得られる。

AdaBoostのアルゴリズムを以下に Alg. (3) として示す。

Algorithm 1 AdaBoost Algorithm

- 1: 部分画像ベクトル \mathbf{x} と教師ラベル c の対を準備 $(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \dots, (\mathbf{x}_n, c_n)$
 - 2: 特徴の重みを初期化 $w_i^1 = \frac{1}{n}, (i = 1, 2, \dots, n)$
 - 3: **for** $t = 1$ to T **do**
 - 4: 分類器 $h_t(\mathbf{x})$ を、以下訓練誤差が最小になるように学習

$$e_t = \sum_{i=1}^n w_i^t I(h_t(\mathbf{x}_i) \neq c_i) \quad (I(\cdot) \text{ は指示関数})$$
 - 5: $\alpha_t \leftarrow \frac{1}{2} \left(\frac{1 - e_t}{e_t} \right)$
 - 6: $w_i^{(t+1)} \leftarrow w_i^t \exp(-\alpha_t c_i h_t(\mathbf{x}))$
 - 7: w_i を正規化
 - 8: **end for**
 - 9: $H(\mathbf{x}) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(\mathbf{x}))$
-

ここで、部分画像ベクトル \mathbf{x} とは入力画像の矩形領域から計算される値で構成されたベクトルである。使用する弱識別器 $h(\cdot)$ は、識別率が50%を超えるという条件のもと、しばしば決定株 (decision stump) もしくは決定木 (decision tree) が用いられる。決定株とは、特徴量で形成されたベクトルの中で、ある一つの値 (属性) が設定した閾値を超えているか否かのみで分類を行うような識別器である。決定木は決定株を深くしたもの、すなわち複数のルールを組み合わせることによって最終的な識別結果を返すものである。単純なルールに基づい

て行われるため、高速な演算を行うことができる利点がある一方で、特徴ベクトルのほとんどの情報を切り捨ててしまうため高い精度が望めないという欠点がある。

AdaBoost は、この欠点を解消するために、“前の識別器によって得られた識別結果”を利用するという工夫を行っている。

学習過程の概略を図で表すと以下 Fig. 3.3 のようになる。

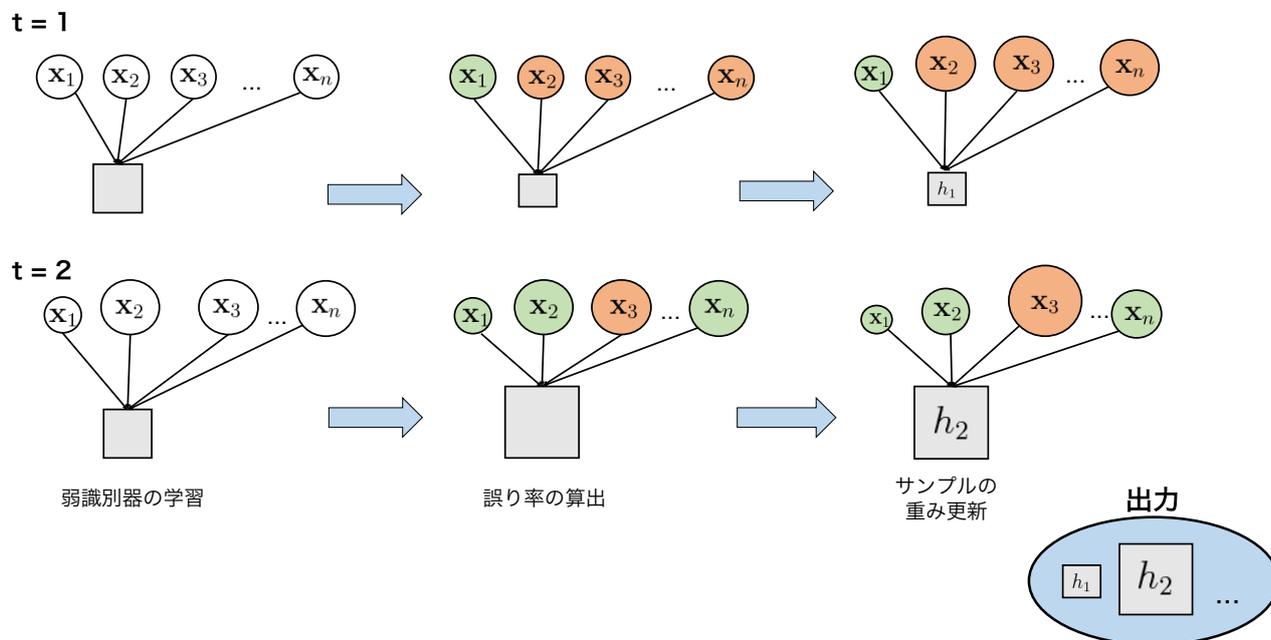


Fig. 3.3: AdaBoost の学習過程模式図

Fig. 3.3 において、円は特徴ベクトル、その大きさは特徴ベクトルの重要度を示している。更に、学習時 (Fig. 3.3 左) 正しく判定した特徴ベクトルを緑色、誤って判定したものを橙色で模式的に表している。一方で、四角形は弱識別器、その大きさは信頼性を表しており、識別器の添字 h_t は、特徴ベクトル x_t を用いて学習させた識別器であることを示している。上図で、1 回目の弱識別器の学習終了時 (Fig. 3.3 右上)、特徴量に対する重要度が増加、それが 2 回目の学習開始時に引き継がれている点に注目されたい。

AdaBoost は、有効な特徴量の選択機構としても解釈することが出来る。これは、弱識別器 (決定株) の分割基準が、そのデータ空間を最もよく分離するような特徴量によって計算されるためである。(=不純度) ひとつの弱識別器に対してひとつの特徴量が割り当てられるため、AdaBoost を形成する弱識別器の個数に応じて、識別時間および精度が変動する。

後述する attentional cascade は、この仕組みを巧みに活用している。

3.2.3 attentional cascade

attentional cascade(以下、cascade 検出器と記述する) は、複数の識別器を連結することによって、物体検出の速度を大幅に向上させた検出器である。

具体的には、各段で用いる特徴を変更して学習させた AdaBoost を複数連結させることによって、検出対象の物体が内部に存在する画像（以下、単に「検出対象画像」と表記する）の中から、明らかに対象の物体でない箇所を早い段階で棄却する。以下に cascade 検出器の物体検出までの流れを Fig. 3.4 として示す。

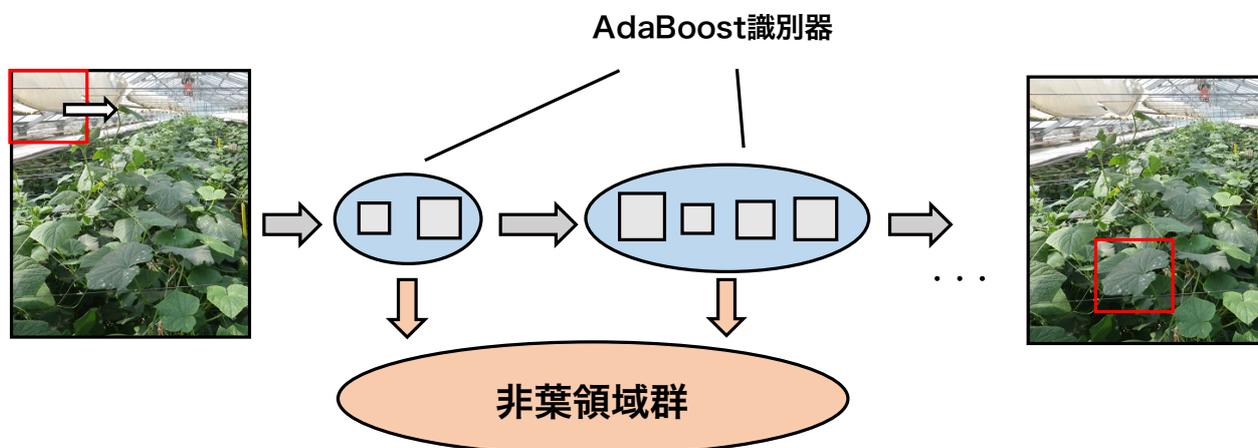


Fig. 3.4: cascade 識別器の検出までの流れ

物体の早期棄却を実現するために、cascade 識別器は coarse-to-fine な手法をとっている。検出器前半部は少量の特徴から学習された AdaBoost 識別器の集合を用いた検出を実施し、高速に物体でない領域を除去する。この時、多少の誤検出 (false positive) を許容してでも、未検出 (false negative / miss rate) を防ぐようにパラメータを調整する点がポイントである。そして、検出器後半部では、前半で除去することができなかった識別が困難なサンプルのみを多数の特徴から学習された AdaBoost 識別器を用いて評価する。

ここで後の説明の為に、用語を以下のように定義する。

- 正例, Positive : 葉がある領域
- 負例, Negative : 葉が存在しない領域 (非葉領域)
- True Positive(TP) : 葉領域のうち、検出器が正しく葉と判断した領域
- False Negative(FN): 葉領域のうち、検出器が誤って非葉と判断した領域 (= 葉を検出できなかった領域)
- True Negative(TN) : 非葉領域のうち、検出器が正しく非葉と判断した領域
- False Positive(FP): 非葉領域のうち、検出器が誤って葉と判断した領域

以下、偽陽性率 (= false positive rate, $\frac{\#offP}{\#ofN}$) を γ , 未検出率 (= $\frac{\#offN}{\#ofP}$) を β と定義する。¹ L 段の AdaBoost 識別器からなる cascade 検出器の最終偽陽性率 γ_0 や未検出率 β_0 は、各段の

¹N: TN+FP, P: TP+FN

識別器が返した偽陽性率や検出率の総乗として表すことが出来る．以下 Eq. (3.5), Eq. (3.6) として示す．

$$\gamma_o = \prod_{l=1}^L \gamma_l \quad (3.5)$$

$$\beta_o = \prod_{l=1}^L \beta_l \quad (3.6)$$

例として、各段の弱識別器が独立であると仮定できる AdaBoost 識別器について、 $\gamma_l = 0.7$, $\beta_l = 0.1$ と設計した場合を検討する．Eq. (3.5), Eq. (3.6) より、cascade 検出器の段数が 10 個 ($L = 10$) である時、 $\gamma_o = 0.7^{10} \approx 0.03$, $\beta_o = 0.1^{10} = 1.0 \times 10^{-10}$ となり、検出器の性能が cascade 構造によって十分に向上することが陽に示される．

Cascade 検出器を構築する上でのポイントは以下 2 点である．

1. 調整パラメータ s_l の導入
2. 各段での学習サンプルの変更

1 つ目の調整パラメータとは、Cascade 検出器を構成する AdaBoost 識別器に順応性を持たせるための変数である．偽陽性率および未検出率は一般的にトレードオフの関係にあるため、AdaBoost 識別器のみでは双方の値を効果的に減少させることは困難である．そこで、各識別器の出力を求める Alg. (3), 式 (9) に調整パラメータ $s \in [-1, 1]$ を導入し、Eq. (3.7) を計算する．

$$H(\cdot) = \text{sign}\left(\sum_{t=1}^T \alpha_t (h_t(\cdot) + s)\right) \quad (3.7)$$

パラメータ s を適用することによって、対象の段までの偽陽性率もしくは未検出率を調整することができる．具体的には、 s が 1 に近づくほど、入力特徴ベクトルから得る識別結果が正に傾くため、検出率の向上を見込むことができる．同様に -1 に近づくと、偽陽性率の低減を見込むことができる．

2 つ目の学習サンプルの変更とは、各段の AdaBoost の学習により有意性を持たせるためのものである．前述したとおり、Cascade 検出器後半に入力されるサンプルは positive/negative の識別がより困難なものである．したがって、後段の AdaBoost 識別器の学習においては、こうしたサンプルに対する識別能を確保する必要がある．一般的には、簡単なサンプル (= 現段階までの Cascade 検出器で正しく negative と判定されるようなサンプル) を削除し、より検出が困難なサンプルを同数分注入することによって検出器を有意なものとする事が多い．

以下、cascade 検出器の学習アルゴリズムを Alg. (2) として示す．

Algorithm 2 attentional cascade algorithm

Input: 正例の特徴量集合 $\mathcal{P} = \{\mathcal{P}_t, \mathcal{P}_v\}$, 負例の特徴量集合 $\mathcal{N} = \{\mathcal{N}_t, \mathcal{N}_v\}$,

目標未検出率 γ_o , 各段ごとの目標偽陽性率 γ_l , 検出率 β_l ,

各段の弱識別器の最大個数 N_l

- 1: $\hat{\gamma}_o \leftarrow 1$, カウンタ $l \leftarrow 0$,
- 2: **while** $\hat{\gamma}_o > \gamma_o$ **do**
- 3: $l \leftarrow l + 1$, $u \leftarrow 10^{-2}$, $s_l \leftarrow 0$, $T_l \leftarrow 1$,
- 4: $\mathcal{P}_t, \mathcal{N}_t$ で AdaBoost 識別器 $H(\mathbf{x}) = \text{sign}[\sum_{t=1}^{T_l} \alpha_t h_t]$ の学習
- 5: $\mathcal{P}_t, \mathcal{N}_t$ および $\mathcal{P}_v, \mathcal{N}_v$, で AdaBoost 識別器 $H(\mathbf{x}) = \text{sign}[\sum_{t=1}^{T_l} \alpha_t (h_t + s_l)]$ を実行
学習データ, 評価データそれぞれの偽陽性率 γ_t, γ_v と検出率 β_t, β_v を算出
- 6: $\hat{\gamma}_l \leftarrow \max(\gamma_t, \gamma_v)$, $\hat{\beta}_l \leftarrow \max(\beta_t, \beta_v)$
- 7: **if** $\hat{\gamma}_l \leq \gamma_l$ and $\hat{\beta}_l \leq \beta_l$ **then**
- 8: $\hat{\gamma}_o \leftarrow \hat{\gamma}_o \times \hat{\gamma}_l$
- 9: **else if** $\hat{\gamma}_l \leq \gamma_l$ and $\hat{\beta}_l > \beta_l$ **then**
- 10: $s_l \leftarrow s_l + u$
- 11: 5へ戻る
- 12: **else if** $\hat{\gamma}_l > \gamma_l$ and $\hat{\beta}_l \leq \beta_l$ **then**
- 13: $s_l \leftarrow s_l - u$
- 14: 5へ戻る
- 15: **else**
- 16: **if** $T_l > N_l$ **then**
- 17: $s_l \leftarrow -1$
- 18: **while** $\hat{\beta}_l > 0.01$ **do**
- 19: 5,6を実施
- 20: **end while**
- 21: $\hat{\gamma}_o \leftarrow \hat{\gamma}_o \times \hat{\gamma}_l$
- 22: **else**
- 23: $T_l \leftarrow T_l + 1$
- 24: 4へ戻る
- 25: **end if**
- 26: **end if**
- 27: 現在までに作成された得られた cascade 検出器で偽陽性サンプル (false positive), 真陰性サンプル (true negative) の除去
- 28: 取り除いたサンプル数分新たなサンプルを注入
- 29: **end while**

Output: cascade 検出器

3.3 評価実験

実際の農場において運用するためには、実際の農場の状況を鑑みた撮影条件における実証実験が必要不可欠である。本項では、これまで述べてきた技術で実現された葉検出機構を用いて行った計算機実験の詳細、および結果について述べる。

3.3.1 データセットの取得

検出器に対する性能実験において、使用したデータセットは合計で3種類ある。

まず、cascade 検出器の学習に用いたデータセットについて述べる。正例として用いた画像は、埼玉県農林技術研究センターにおいて撮影されたキュウリ葉画像である。これらのキュウリ葉画像は、葉が1枚画像の中央部に配置されるように撮影されたものであり、光や影による影響を受けていないものである。

負例として用いた画像は、McGill Calibrated Colour Image Database [43] に収録されている画像群である。この画像データベースには、鳥、動物、人工物等を撮影対象とした画像が、色のキャリブレーション及び一律のサイズへの縮小処理を施した状態で収録されている。一般に、cascade 検出器の学習を行う際、ネガティブデータとして注入する画像群は対象のデータと大きく異なるものを入力することが多い [44]。今回、「非葉」と判定される領域を早期に棄却するため（具体的な非葉領域としては、ビニールハウスの天井部分、畑の土部分、畝と畝の間の通り道部分等が考えられる）こうしたデータを用いた。

これらの画像群は、すべてICFを取得する直前に一律のサイズへと変更を行った。具体的には、正例画像は64×64サイズへ、アスペクト比を無視してリサイズを行った。これは、64×64pixelという小規模の領域全体に、「葉の輪郭」という情報を残すためである。負例画像に対しては、負例画像1枚につき5箇所の64×64pixelのパッチをランダムに取得することによって取得した。

以降、この条件で取得した学習用画像群を、4章で用いるものと明確に区別するために「検出部検証用学習データセット」と呼ぶ。

casacade 検出器の評価に用いたデータセットは、同じく埼玉県農林技術研究センターにおいて撮影されたキュウリ農場の画像である。撮影に際しては実際に農場で運用することを想定し、撮影状況は「ビニールハウスに定点カメラを取り付けた場合」というFig. 3.5のようなケースを想定して行った。具体的には、撮影者が脚立に登り、株が密集している箇所を様々な角度で、株の上から撮影したものとなる。

以下に想定している運用例を Fig. 3.5 として、撮影の様子を Fig. 3.6 として示す。

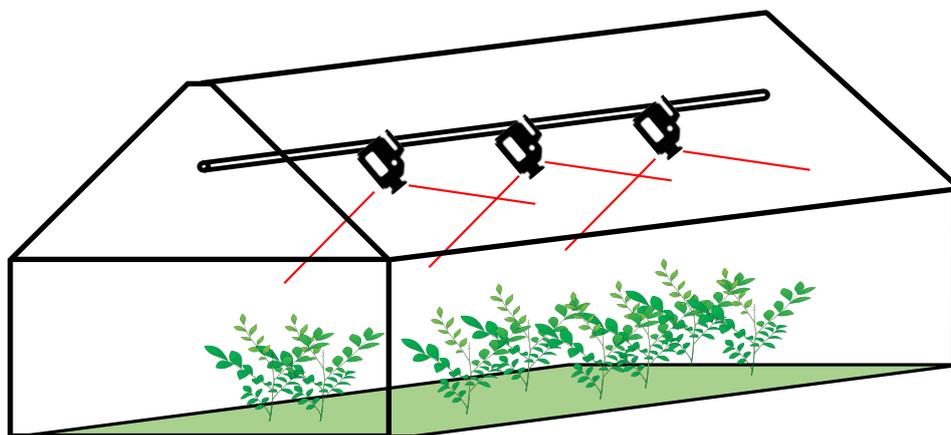


Fig. 3.5: 運用の想定例



Fig. 3.6: 撮影の様子

こうして収集された写真に対し、dlib のパッケージ群の中に存在する imglab というアノテーションツールを用いて、葉領域の bounding box を付与した [45]. 作業画面を以下に Fig. 3.7 として示す.

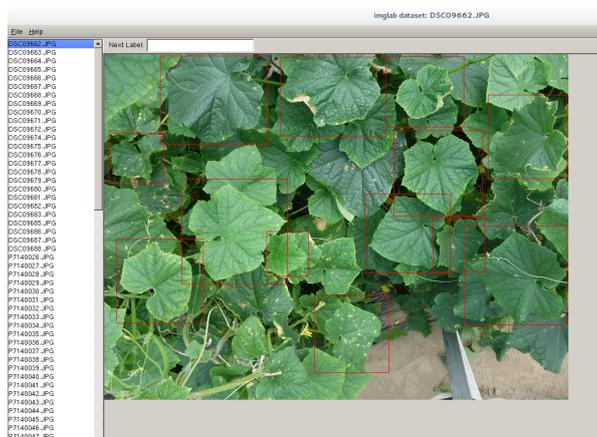


Fig. 3.7: バウンディングボックス付与の様子

本アノテーション作業は、著者含む 2 名の手作業によって行われた。以下、このアノテーション済みの画像を「検出部評価用データセット」と呼ぶ。

なお、評価用データセット撮影に使用されたデジタルカメラは、SONY NEX-5N, SONY DSC-W1, SONY DSC-RX100, および OLYMPUS μ 9010 の4種類であり、いずれも一般向けに市販されているものである。

3.3.2 実験条件

前項で述べた検出部検証用データセットを用いて、以下のように実験を実施した。

検出部検証用データセットにおいて使用した画像パッチ枚数は、正例（葉画像）1,000枚、負例（非葉画像）3,000枚である。このうち、正例750枚、負例2,250枚を training 用として、正例250枚、負例750枚を validation 用として用いた。

ICF を取得したチャンネル数は、以下に示すチャンネル群合計10種類である。

- 画像 I の LUV 色空間における各チャンネル (3種類)
- 画像 I の勾配強度 (1種類)
- 画像 I のグレースケール画像に対して15度165度の直線フィルタを畳み込んだもの (6種類)

これらは原著論文 [40] に従った。

特徴量を取得する矩形領域のサイズは 16×16 pixel とし、矩形領域は各チャンネル、各位置ランダムに50個取得した。

3.3.3 実験結果

3.4.2 で示した条件の下 cascade 検出器を学習した結果、合計2層の cascade が作成された。弱識別器の個数は各レイヤー10個、20個となっている。

以上の条件下で検出実験を行った結果、検出箇所は0箇所となり、検出を行うことができなかった。検出実験に使用した画像を以下に Fig. 3.8, Fig. 3.9 として示す。



Fig. 3.8: 検出対象画像1



Fig. 3.9: 検出対象画像 2

また，検出アルゴリズムを以下に示す．

Algorithm 3 Detection Algorithm

Input: 検出対象の RGB カラー画像 $Z(W \times H\text{px.})$ ，学習済み cascade 検出器 (L 段)，

探索窓のサイズ sz ，負領域集合 $\mathcal{R}_{neg} \in \phi$ ，正領域集合 $\mathcal{R}_{pos} \in \phi$

1: $\mathcal{Q} \in \{[h, h + (m \times sz)], [w, w + (m \times sz)] \subset Z, m \in \mathbb{N}\}$

2: **for** $l = 1$ to L **do**

3: **for** $\mathcal{Q} \setminus \mathcal{R}_{neg}$ **do**

4: $Z_{roi} = \exists \mathcal{Q} \setminus \mathcal{R}_{neg}$

5: Z_{roi} を l 段目の AdaBoost 識別器で選択されたチャンネルへ変換

6: AdaBoost 学習時と同じ次元数の特徴量を取得

7: **if** Z_{roi} 内で得た特徴が l 段目の AdaBoost によって正と判断された時 **then**

8: **if** $l = L$ **then**

9: \mathcal{R}_{pos} に Z_{roi} を追加

10: **else**

11: $l = l + 1$

12: **end if**

13: **else**

14: \mathcal{R}_{neg} に Z_{roi} を追加

15: **end if**

16: **end for**

17: **end for**

Output: 検出結果 \mathcal{R}_{pos}

第4章 病変識別部の開発と実験

本章では、抽出した葉候補領域の中から病変識別を行う各手法を検討する。

4.1 畳み込みニューラルネットワーク (CNN)

本項では、本研究の識別部として使用した CNN の構造について詳説する。

4.1.1 CNN の各層における処理

これまでに CNN のアーキテクチャは多数提案されてきたが、ほとんどのアーキテクチャで採用されている最も基本的な処理は「畳み込み」「プーリング」と呼ばれる2つの処理である。前者は、画像（もしくは特徴マップ）とフィルタとの畳み込み積分を行い、物体や特徴マップ内に存在する局所特徴を抽出する処理であり、後者はそうした局所特徴を集約するような処理を行う。これらの処理を行うユニットの集合を、慣用的に「畳み込み層 (convolution layer)」「プーリング層 (pooling layer)」と呼ぶ。以下、畳み込みとプーリング、およびその間に介在する活性化関数について述べる。

本論文の説明では、以下のように文字を定義する。なお、本章で「ベクトル」と表記したものはすべて「縦ベクトル」とする。

- \mathbf{x}_n n 次元学習データベクトル
- \mathbf{d}_n n 次元教師データベクトル
- $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{d}_1), (\mathbf{x}_2, \mathbf{d}_2), \dots, (\mathbf{x}_N, \mathbf{d}_N)\}$ 学習ベクトルと教師ベクトルの集合
- x : $W \times W$ サイズを持つ対象画像内の任意の位置にあるユニットの値
- h : サイズ $H \times H$ を持つフィルタ
- z : 畳み込み層における任意の位置にあるユニットの出力
- a : 活性化関数を通した後の任意の位置にあるユニットの出力
- u : プーリング層における任意の位置にあるユニットの出力

畳み込み層

畳み込み層では、入力画像もしくは特徴マップ（後述する）と、フィルタとの畳み込み演算が行われる。グレースケール画像に対する畳み込み演算は以下 Eq. (4.1) で定義される¹。

$$z_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p, j+q} h_{p,q} \quad (4.1)$$

画像とフィルタを畳み込む処理は、画像の局所領域とフィルタとの類似性を発見することと等しい。グレースケール画像の場合、フィルタとの濃度値の例えば、斜線上のフィルタを畳み込めば、画像内から斜線成分を持つ箇所が抽出される。これは、一般的なエッジ検出フィルタと同等の働きを持つ。

畳み込みは、任意のチャンネル数を持つ画像に対して適用することが可能である。以下 Eq. (4.2) として示す。

$$z_{i,j,k} = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p, j+q, k} h_{p,q,k} + b_k \quad (4.2)$$

ここで、 b_k はバイアス項であり、チャンネル毎に共通の値を持つ。

畳み込みは、プーリングと合わせて複数回実施されることが一般的だが、CNN の中間層以降ではその出力結果の画像を特徴マップと呼ぶ。その為、2 回目の畳み込み処理では、特徴マップとフィルタの畳み込み積分となり、更に局所的な成分が抽出されていくことになる。

上記畳み込み演算は、厳密にフィルタを 1px. ずつ走査させると、その分演算回数が増えることによる学習時間の増加を招く。その為、数 px. 間隔で画像ないしは特徴マップを走査するケースが多い。この時のフィルタの走査幅のことをストライドと呼ぶ。また、畳み込みの際、画像（特徴マップ）の隅などを対象ユニット (= x) とした場合、フィルタが画像ユニット群が定義する範囲をはみ出してしまふ。この時、定義されていないユニットの部分は、適当な値（多くの場合 0、特にこの時 zero-padding と呼ぶ）を用いて補完される。

活性化関数

一般的なニューラルネットワークで学習される関数は、高々結合荷重と入力値の線形結合であるため、そのまま出力結果の値を利用するだけでは非線形な識別を実現することはできない。このため、通常 CNN を含むニューラルネットワークでは、活性化関数と呼ばれる入力を非線形空間へ写像させる関数が利用される。CNN においてしばしば利用される活性化関数が、Eq. (4.3) で示される Rectified Linear Unit（以下 ReLU）である。

$$\text{ReLU}(x) = \max(0, x) \quad (4.3)$$

¹厳密にはこの式は「相関」を算出する式だが、出力結果に実質的な差異はない。本論文では、慣習に従い“畳み込み”という表現を使用する。

式から明らかな通り、入力されるスカラ量が負であるときは0を、正である時はその入力値をそのまま返すような関数である。特長として、単純な演算であるため計算量が少なく、後述する学習が早く収束することが知られている。

プーリング層

プーリング層の目的は、前段で得た特徴マップを、ある局所領域毎に集約することによって、得た特徴にシフト不変性を持たせることである。具体的には、畳み込み層での出力内の一定の矩形領域を指定し、その矩形領域内での最大値や平均値を取得し、その値を次の層へと伝搬させる。この処理は一般式で表すことができる。一定の矩形領域を P とした時の一般式を Eq. (4.4) として示す。

$$u_{i,j,k} = \left(\frac{1}{H^2} \sum_{(p,q) \in P} z_{p,q,k} \right)^{\frac{1}{P}} \quad (4.4)$$

Eq. (4.4) において $P = 1$ とした時、特に平均値プーリングと呼び、矩形領域内の平均値を算出する処理を指す。また、 $P = \infty$ の時は max プーリングと呼び、矩形領域内の最大値を算出する処理となる。

以上の処理を図解したものを以下に Fig. 4.1 として示す。

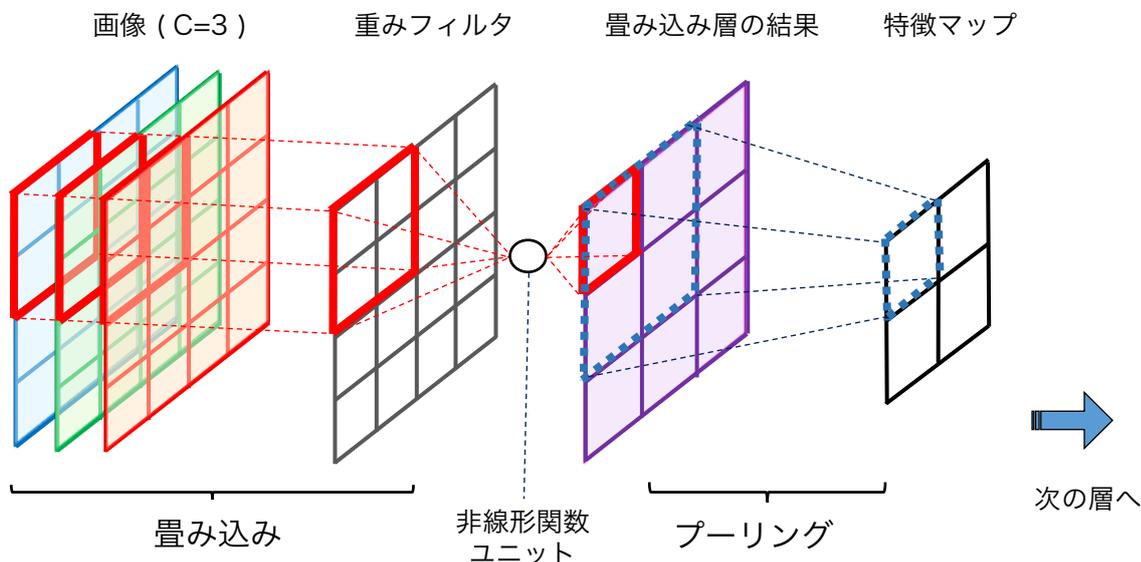


Fig. 4.1: CNN の各層での処理

局所正規化層 (local contrast normalization)

一般的に画像を対象として認識、特徴点抽出等を実施する際、撮影時の明暗差を軽減するため、画像内コントラストの正規化を行うことが一般的である。

この処理は、CNNの層として役割を持たせることができる。その場合、ある局所領域内の全チャンネルにわたって正規化処理を行うことができる。

$H \times H$ サイズを持つ小領域 Q を考える。小領域に所属する画素 Q_{ij} について、その位置に存在するユニットの加重平均 \bar{x}_{ij} および分散は以下 Eq. (4.5), Eq. (4.6) として定式化される。

$$\bar{x}_{ij} = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{(p,q) \in Q_{ij}} w_{pq} x_{i+p, j+q, k} \quad (4.5)$$

$$\bar{\sigma}_{ij}^2 = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{(p,q) \in Q_{ij}} w_{pqk} (x_{i+p, j+q, k} - \bar{x}_{ij}) \quad (4.6)$$

この値を用いて、減算、除算正規化を行う。減算正規化は Eq. (4.7)、除算正規化は Eq. (4.8) として定義される。

$$z_{ijk} = x_{ijk} - \bar{x}_{ij} \quad (4.7)$$

$$z_{ijk} = \frac{x_{ijk} - \bar{x}_{ij}}{\max(c, \sigma_{ij})} \quad (4.8)$$

ここで、 c は正規化の際の濃度を調整するためのパラメータである。この値を設定しないと、標準偏差が1より小さな値である場合、出力の値は入力値より大きな値を返すため、よりユニットごとの変動が大きくなってしまう。

出力層

複数回の畳み込み処理、プーリング処理の後、出力層と呼ばれるレイヤーが置かれる。出力層では、畳み込みおよびプーリング処理によって抽出された高次元特徴ベクトルを、教師ベクトルに対する確率分布へと写像させる役割を持つ。

K クラス分類の場合、以下 Eq. (4.9) で示されるソフトマックス関数が、出力層のユニットに対する活性化関数として用いられる。

$$y_k = \frac{\exp(u_k)}{\sum_{j=1}^K \exp(u_j)} \quad (4.9)$$

ソフトマックス関数の出力は、 $[0, 1]$ の間に正規化されて出力される。即ち、出力層の値は、与えられた入力があるクラスに所属する確率を表していると解釈できる。この出力層に存在するユニットのうち、最大値を取るユニットが与えられた入力に対するCNNの出力となる。

4.1.2 CNNの学習方法

CNNの学習プロセスは、巨大なデータ量を取り扱うための工夫が加えられている他は、通常のニューラルネットワークと大きな差はない。即ち、任意の誤差関数を定義し、学習器によって得られた出力と、実際の出力間に生じる誤差を最小にすることが目的となる。

学習の対象となるのは、畳み込み層で定義されるフィルタの各ユニットの結合重み \mathbf{w} と、バイアス b である。

今回の植物病害の種類分別のような多クラス分類問題である場合、実際の出力と理想出力との乖離は、交差エントロピーと呼ばれる誤差関数で測られる。これは、出力層の活性化関数にソフトマックス関数を選択した場合の最尤推定によって導くことができる。

$$E(\mathbf{w}) = - \sum_{n=1}^N \sum_{k=1}^K d_{nk} \log y_k(\mathbf{x}_n; \mathbf{w}) \quad (4.10)$$

以下、Eq. (4.10) を目的関数と呼ぶ。学習の目標は、Eq. (4.10) のような誤差関数が最小となる \mathbf{w} を発見すること、即ち $\mathbf{w} = \operatorname{argmin}_{\mathbf{w}} E(\mathbf{w})$ を得ることである。しかしながら、目的関数 $E(\mathbf{w})$ は一般的に非凸関数である。² その為、一般に関数を最小にする \mathbf{w} を厳密に求めるのではなく、勾配降下法を用いて準最適となる解を求める。

関数 E の勾配を $\nabla E = \frac{\partial E}{\partial \mathbf{w}}$ と定義する。勾配降下法では、あるステップ t におけるフィルタの結合重み \mathbf{w}_t を、負の勾配方向へと反復的に更新していく。式で表すと Eq. (4.11) のようになる。

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \epsilon \nabla E \quad (4.11)$$

ここで、 ϵ は学習係数である。

しかしながら、一度に計算する勾配の規模が大きいと、学習が非常に長時間に及んでしまう。そのため、CNNの学習では一般的にEq. (4.11) で表されるような全ての勾配によって \mathbf{w}^t を更新する方法ではなく、重み更新の際にある学習データに対する勾配のみをステップごとに乱択する確率的勾配降下法 (stochastic gradient descent) が用いられる。

また、その「ある学習データ」のサイズ (以下、 \mathcal{D}_t と記載する) をあらかじめCNN学習開始時に定めておき、各学習ステップごとにそのサイズに応じたサンプルを選択する手法もしばしば使用される。この学習データのサイズをミニバッチと呼び、通常学習前に定められる。³

以上2点を鑑みた更新式を Eq. (4.12)、Eq. (4.13) として示す。

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \epsilon \nabla E_n \quad (4.12)$$

²関数 f が凸であるとは、 f 上の任意の2点 $(x_1, f(x_1)), (x_2, f(x_2))$ を結んだ線分が常に f の上側 (下に凸)、もしくは下側 (上に凸) に存在することを表す。下に凸な関数の場合は関数の極小値が常に最小値となるため、二階微分によって容易に関数の最小値を導出することが可能となる。

³CNNアーキテクチャ毎の最適なミニバッチサイズ等の議論は [46] に詳しい。

$$E_t(\mathbf{w}) = \frac{1}{|\mathcal{D}_t|} \sum_{n \in |\mathcal{D}_t|} E_n(\mathbf{w}) \quad (4.13)$$

ただし $E(\mathbf{w}) = \sum_{n=1}^N E_n(\mathbf{w})$ である。また、Eq. (4.13) における添字 t は、各学習ステップごとに異なるサンプルが選択されることを示している。

4.2 植物病害分類実験

本章では、CNN を用いた植物病害自動診断実験について述べる。

4.2.1 データセットの取得

病害一覧

識別対象とした画像は、埼玉県農林技術研究センターで撮影されたキュウリウイルス病害感染葉である。病害の種類は、実際の農業現場で影響の高いものとして、Table. 4.1 に示す7種類の病害を選択した。

以下、本実験に使用するキュウリ葉に出現する病害の写真を Fig. 4.2 として、病害一覧とその特徴を Table. 4.1 として示す。なお、以降これらのキュウリウイルス病徴のうち特に限定して言及する場合、各病害の略称を用いる。

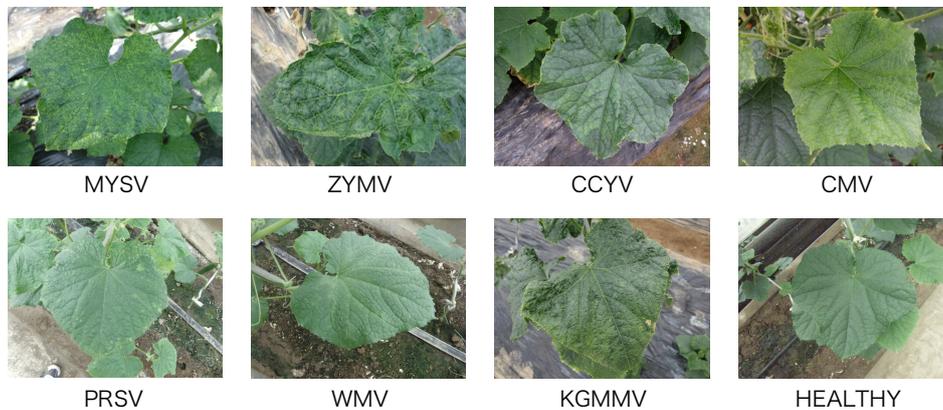


Fig. 4.2: 本研究で使用した病害写真の一覧

Table. 4.1: 本研究で使用した植物病害の主な症状（日本植物病害大辞典等を参考に作成）

病徴名	初期	末期	備考
MYSV (Melon Yellow Spot Virus)	葉脈透過症状 退緑斑紋を伴ったモザイク症状	葉が激しく黄化 葉脈間にえそ斑点が生じ、後に葉全体が黄化する	一部の品種でモザイク症状を生じる
ZYMV (Zucchini Yellow Mosaic Virus)	新葉に緑色濃淡の明瞭なモザイク その後葉脈間が黄化	葉脈の両側に農緑帯が出現 奇形を伴った激しいモザイク	WMVとよく似ているが 症状が強く出る
CCYV (Cucurbit Chlorotic Yellow Virus)	葉脈間に黄緑色の小斑点が多数発生	葉脈間は黄化するが、葉脈部の緑色は残る	
CMV (Cucumber Mosaic Virus)	上位葉、展開直後の新葉に出現 葉脈の透過や濃淡のある斑紋	葉は硬化し、下に巻く 斑点が集合して モザイク様になる	生育が緩慢となる
PRSV (Papaya Ring Spot Virus)	葉脈透過、モザイク症状	次第に黄化症状が強く現れる	果実は奇形となる場合が多い
WMV (Watermelon Mosaic Virus)	モザイク症状を呈するが 程度が極めて軽く識別が困難	葉脈の両側に農緑帯が発現 古い葉は健全葉と比較し光沢が出にくい	ZYMVとよく似ているが、症状が軽微
KGMV (Kruri Green Mottle Mosaic Virus)	新葉に黄色の小斑点 次第に明瞭なモザイクを呈する	緑色濃淡の明瞭なモザイク ときに農緑部が盛り上がり奇形葉となる	果実への症状が強く出る

識別部の実験においては、これら7種類の病徴に、健全な葉を加えた計8クラスの分類を行った。以降の記述では、これらの病変データから成るデータセットを「識別部検証用データセット」と呼ぶ。(3章で用いたデータセットとは異なることに注意されたし)

撮影環境の差異について

本研究では、定量的な実験を行うにあたり、撮影条件の良いものと、撮影条件の悪いものの2種類を使用した。2.4.でも述べた通り農業現場においては、太陽光、自然風、葉と葉の重畳等の外的要因によって、撮影される画像が必ずしも好条件とならない。そこで、本研究ではCNNの病変識別能力を定量的に調査するため、好条件下で撮影された画像に対する実験、および悪条件下で撮影された画像に対する実験を実施した。

好条件下で撮影された画像は、日光を要因とした、葉領域における明度差がないものである。即ち、葉の一部に他の葉や人工物による影が入っていたり、日光による白飛びなどの影響が無いものを指す。悪条件下で撮影された画像とは、こうした悪影響を含むものを指す。好条件で撮影された画像、悪条件下で撮影された画像の例を以下に Fig. 4.3 として示す。



Fig. 4.3: 左：好条件下で撮影された画像，右：悪条件下で撮影された画像

4.2.2 CNN の設計

本研究で用いたCNNのアーキテクチャの図を Fig. 4.4 として、詳細構造を Table. 4.3 として示す。

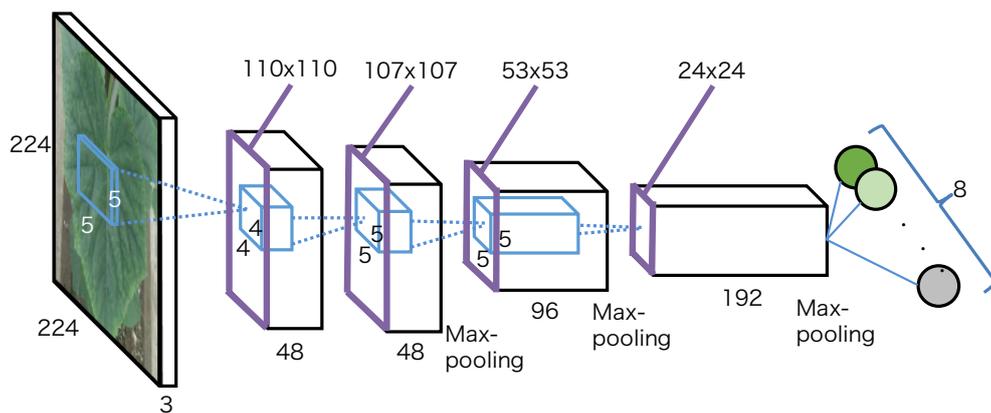


Fig. 4.4: 本研究で使した病害写真の一覧

なお本研究のアーキテクチャは AlexNet [14] を参考に決定した。

Table. 4.2: 本研究で用いた識別器の詳細構造

層の名称	動作	フィルタサイズ	出力時のデータ構造	その他備考
(入力)	-		$224 \times 224 \times 3$	
conv1	convolution	5×5	$107 \times 107 \times 3$	stride=2
conv2	convolution	4×4	$107 \times 107 \times 48$	
pool2	max pooling	3×3	$53 \times 53 \times 48$	
norm2	LCN		$53 \times 53 \times 48$	
conv3	convolution	5×5	$53 \times 53 \times 96$	padding=2
pool3	max pooling	3×3	$26 \times 26 \times 96$	stride=2
norm3	LCN	5×5	$26 \times 26 \times 96$	
conv4	convolution	5×5	$24 \times 24 \times 192$	padding=1
pool4	max pooling	5×5	$12 \times 12 \times 192$	
(出力)	full connection		$8 \times 1 \times 1$	

他、詳細なハイパラメータを以下に Table. ??として示す。これらのパラメータは予備実験によって決定している。

Table. 4.3: 本研究で用いた識別器の詳細構造

ハイパラメータ	内容
学習 epoch	30 epoch
バッチサイズ	10 枚
learning rate	8.0×10^{-5}
momentum	0.9
weight decay	4.0×10^{-6}

また、CNN の作成にあたっては Caffe [47] を利用した。

4.2.3 評価実験

前処理および評価方法

後述する全ての実験において、すべての画像サイズは 224×224 pixel へとリサイズしている。しかしながら、各実験において少しずつ前処理のフローが異なっている。これは、学習させる画像を擬似的に増加させることによって、CNN の識別性能が向上することが経験的に知られているためである [48]。この処理を data augmentation と呼ぶ。以降、data augmentation をしていない前処理をフロー (i)、した処理を (ii) と定義し明確に区別する。

まず、双方のフローにおいて元画像（矩形）から、画像の中央領域を取得した。⁴この時、フロー (ii) を施す画像に対しては、数 pixel 分ずらした画像を同時に作成した。この処理によって、対象の画像は 5 倍へ増加する。その後、フロー (i) においては、その領域をそのまま 224×224 pixel へと縮小し、学習用データとした。一方フロー (ii) においては、まず 316×316 pixel に縮小した。続いて、対象画像を角度 10 度刻みで 360 度ずつ回転させた画像も合わせて算出した。この処理によって、対象の画像は 36 倍へと増加する。そして、その中央領域を獲得し、 224×224 pixel へと縮小した。最後に、ここまでのフローで手に入れた画像の鏡面画像を算出し、学習用データとした。

以上の処理を Fig. 5.1 として図解する。フロー (ii) を施す画像に対しては、 $5 \times 36 \times 2 = 360$ 倍にデータが増大させたことになる。

⁴画像の解像度は撮影されたデジタルカメラの種類によってまちまちであるが、いずれも Quad-VGA (1280×960 pixel) 以上あることを確認している。

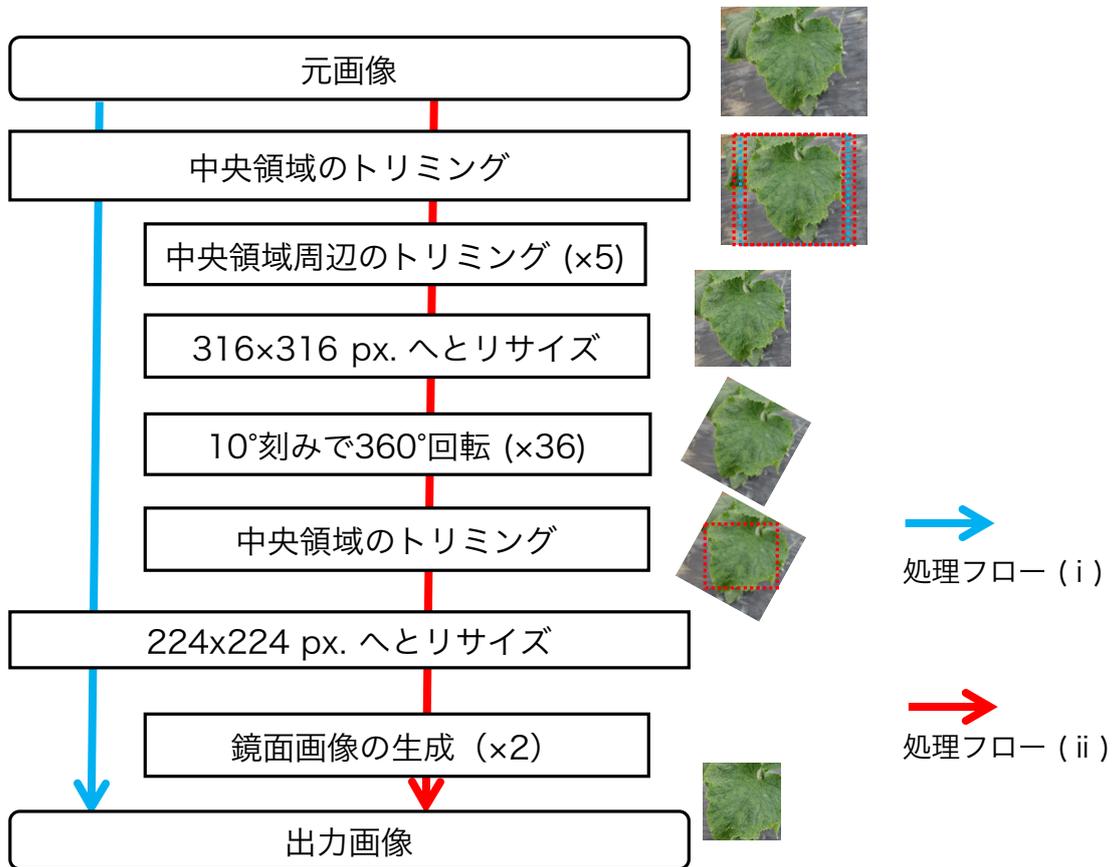


Fig. 4.5: 本研究で使用了した病害写真の一覧

評価方法には、4-fold cross validation を用いた。これは、全サンプル群を均等な数4つのグループに分割し、1つを評価 (validation) データとし、残りの3つを学習 (training) に用いる方式である。この手法を用いることで、学習サンプルの分布に依拠した学習が行われることを防ぐ。

また、評価指標には正解率 (Accuracy), 感度 (Sensitivity, SE.), 特異度 (Specificity, SP.) を用いた。定性的には、感度は「ある病変を正しく病変と識別できた割合」特異度は「健全な葉を正しく健全と識別できた割合」を表す。

これらの評価指標は、以下 Eq. (4.14)~Eq. (4.16) によって計算される。

$$Accuracy[\%] = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.14)$$

$$Sensitivity[\%] = \frac{TP}{TP + FN} \quad (4.15)$$

$$Specificity[\%] = \frac{TN}{FP + TN} \quad (4.16)$$

実験 A. CNN の識別能力の検証

植物病害分類問題に適した識別器を選択するため、BoVW [9] と SVM による比較実験を実施した。BoVW による識別法は、ILSVRC2012 以前にしばしば一般、特定物体認識に用いられていた手法であり、植物病害の診断に対しても有益である可能性がある。そこで、特徴量を BoVW、識別器を SVM とした 8 クラス分類問題を計算機実験によって解き、その精度を比較した。

使用した識別部検証用データセットは、好条件下で撮影された植物病害画像 8,000 枚（各クラス 1,000 枚）である。これらは全てフロー (i) に沿って前処理を実施した。

BoVW は、SIFT 特徴量 [8] のキーポイントを、k-means クラスタリングにより算出した。セントロイド数は $k = 64$ 、SVM のカーネルは RBF カーネルを用い、パラメータは $\gamma = 8.0 \times 10^{-3}$ 、 $C = 1.0$ とした。

以下に、4-fold cross validation によって得られた実験結果を Table. 4.2.3 として示す。

Table. 4.4: 識別器ごとの性能評価結果

手法	Accuracy[%]	Ave. Sensitivity[%]	Specificity[%]
BoVW + SVM	34.2	29.0	70.4
CNN	56.2	54.5	68.1

Sensitivity は各病変について計算可能である。以下、病変ごとの Sensitivity を Table. 4.5 示す。

Table. 4.5: 各病変に対する Specificity の値

手法	SE. MYSV [%]	SE. ZYMV [%]	SE. CCYV [%]	SE. CMV [%]	SE. PRSV [%]	SE. WMV [%]	SE. KGMMV [%]
BoVW + SVM	53.8	28.5	33.3	38.6	27.9	4.1	16.7
CNN	50.8	52.7	65.3	69.9	45.7	46.0	51.4

実験 B. 好条件下/悪条件下での画像における検証

実験 A を踏まえて、実験 B では好条件下、悪条件下で撮影された画像を CNN により分類する実験を行った。

使用した識別部検証用データセットは、好条件下で撮影された植物病害画像 8,000 枚（各クラス 1,000 枚）、悪条件下で撮影された植物病害画像 8,000 枚（各クラス 1,000 枚）である。

これらの画像は、学習時に 4-fold cross validation ポリシーにより training データ (6,000 枚) と validation データ (2,000 枚) に分離される。そこで、training データに対してはフロー (ii) に沿って前処理を施し、validation データに対してはフロー (i) に沿って前処理を施した。

以下に実験結果を Table. 4.6 として示す。

Table. 4.6: 好条件/悪条件下における識別結果

	Accuracy[%]	Ave. Sensitivity[%]	Specificity[%]
good w/o aug.	55.5	54.6	61.4
good w/ aug.	84.0	83.8	85.5
bad w/o aug.	50.2	48.8	59.5
bad w/ aug.	80.6	84.0	83.5

第5章 考察

5.1 検出部に関する考察

今回、葉検出部において葉領域の検出を実現することができなかった。この原因として考えられることとして、3つの事柄が考えられる。

1点目は、特徴量取得に用いた画像チャンネルの妥当性に関してである。今回用いたICFは、先行研究 [40] で実施されていた歩行者検出に即したものである。先行研究の論文内では、学習画像内で選択された特徴量に対しての議論がなされていたが、LUV色空間のUチャンネルにおける歩行者の顔部分において強い選択性が示されていた。

広範囲の葉領域検出にこの問題を適用すると、葉の表面の「緑色」に対して強い選択性がある特徴量を選定する必要がある。今回の「検出箇所なし」という結果から、今回選択したチャンネル（LUV色空間、15～165度の角度を持った直線フィルタ、勾配強度）の中に、強い選択性を持ったチャンネルが無く、結果AdaBoostの学習時に識別に有効な特徴量（特徴チャンネル）が選択されなかったことが考えられる。

2点目は、cascade検出器を学習させる際の次元数の不足である。今回、多くのチャンネル、矩形領域の位置から、50種類（=50次元）の矩形領域を抽出し特徴とした、しかしながら、10チャンネルを持つ64×64pixelのパッチから、16×16pixelの矩形領域を選択するよう今回のケースの場合、1枚の画像パッチから選ばれうる矩形領域の種類数は1万通りを超える¹。今回はそれと比較してかなり少ない次元数で実験を実施したため、望ましい性能が得られなかったとも考えられる。

3点目は、検出部検証用データセットに関してである。今回は1,000通りの正例画像を64×64pixelにリサイズ、負例画像には自然画像を用いたが、正例、負例の明確な区別がつかない画像が選択された可能性がある。例えば、負例画像の中には風景画像が含まれている。こういった画像は、形状にこそ類似性は無いが、色空間の選択の仕方によっては葉領域と定義した箇所と類似した特徴量を選択される危険性を孕んでいる。より正例、負例を分離するようなデータセットからICFを取り出すことによって、より有益な結果が得られた可能性もある。

以上3点を今後の課題としたい。

¹ $33 \times 33 \times 10 = 10,890$

5.2 葉識別部に関する考察

BoVW と SVM による識別実験 (実験 A) においては, 興味深い結果が獲得できた. Table. 4.2.3 および Table. 4.5 においてほとんどの指標において CNN の性能が BoVW+SVM の性能を上回っているが, Specificity および MYSV 病変葉における性能が, CNN の性能を僅かながら上回っている.

これは, BoVW の特徴設計に依る部分だと考えられる. 今回, BoVW の特徴点サンプリングには, SIFT 特徴のキーポイントを適用する形をとった. SIFT 特徴量はマルチスケールな画像中から, 物体の輝度変化が大きな箇所を特徴点として定める. 今回のケースの場合, 識別部検証用データセットに用いた画像群に, 病徴が比較的強く現れており, それによって病徴が持つ凹凸に特化した codebook が作成されたと考えられる. Table. 4.5 において, BoVW+SVM における WMV の識別率が非常に小さいのも, 同じ理由から来ているものと考えられる.

実験 B の結果より, Data Augmentation が好条件/悪条件下の識別において有効に機能することが示された. しかし, あらゆる環境下に対応可能か否かにおいては, 未だ議論の余地が残されている.

説明のため以下に画像を示す.

これらは, 悪条件下のデータ (w/ data augmentation) において CNN を学習し, 評価を実施した際にある fold において誤識別をした画像群である.

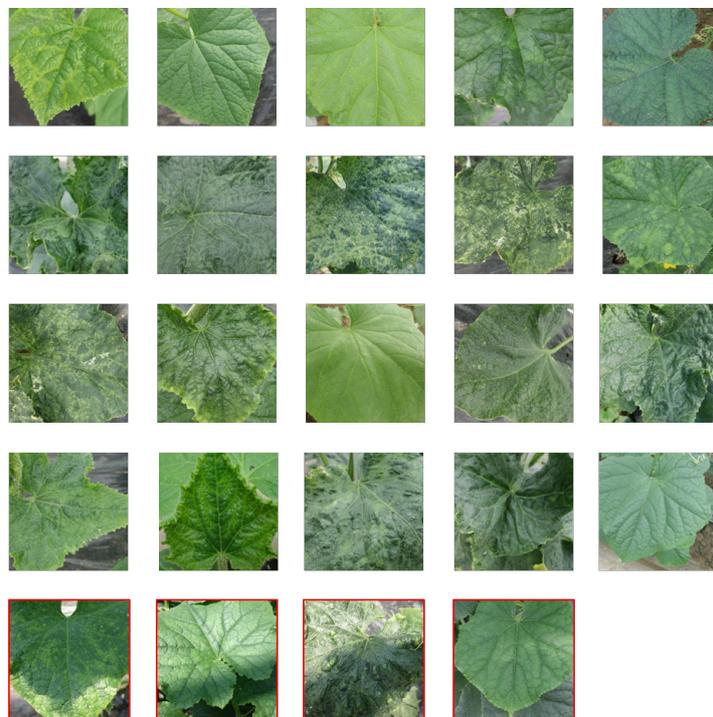


Fig. 5.1: 誤識別画像群. 赤枠で囲まれている画像群は, 明暗の影響を強く受けている.

これらを眺めると, 光や影が写り込んだ画像群と比較して, むしろ好条件な画像の方が多く誤識別をしている. これは, 悪条件の画像が入ったことにより, 白もしくは黒の画素に強

く反応するユニットが生成されてしまったためと考えられる。また、CNNは僅かな画素値の変化で識別結果が変動することが知られており [49]、汎用的なシステムの構築にあたっては、より精緻な学習が求められると考えられる。

第6章 まとめ

本研究では、画像工学、機械学習技術を用いた、実際の農場で運用可能な状況下における植物病害自動診断システムの検討を実施した。検出部においては、integral channel featuresを用いて葉の画像から特徴量を取得し、attentional cascadeによる葉領域の検出を試みた。今回検出までは至らなかったが、より厳密なチャンネルの生成、特徴量ベクトルの次元数の調整等によって今回の手法を用いて実現する可能性は残されている。一方、葉識別部においては、convolutional neural networkを用いることによって、7種類のキュウリ葉と健全葉のクラス分類問題において80%を超える識別能を達成した。今後の課題として、現在未解決となっている葉領域検出の実現、及びより多種多様な背景、病変、植物に対して適用できる診断装置の実現が挙げられる。

謝辞

本論文は、多くの皆様方の多大なご尽力のおかげで執筆することができました。

まず、研究室配属から4年間指導していただいた彌富仁准教授に心より御礼申し上げます。研究に関するアドバイスのみならず、英論文執筆、就職活動等様々なご指導を賜りました。最後の1年が日本と Baltimore を繋いだ遠隔でのディスカッションのみになってしまったことは残念ですが、先生の Johns Hopkins 大学での経験は、必ずや研究室の後輩たちに良い影響をもたらすと確信しております。

続いて、弊研究に必要不可欠な植物病害写真のご提供、および病害写真の撮影にご協力頂きました埼玉県農林技術研究センターの宇賀博之様に感謝申し上げます。宇賀様の快いご協力がなくては、本研究は成立しませんでした。葉領域検出を実現できなかったことは誠に心苦しいですが、引き継いだ後輩が実現させてくれると願っております。

また、植物病害に関する様々な知識をご教示頂いた他、当研究を多方面にPRして下さった法政大学生命科学部の鍵和田聡専任講師にお礼申し上げます。アグリビジネスフェア、イノベーション・ジャパンの際は大変にお世話になりました。

また、多くの助言をして下さった彌富研究室の皆様、特に共同研究者の藤田恵梨香さん、同じ研究チームの諏訪勝元さん、榊原友助さん、Cap Huu Quan さんにお礼申し上げます。特に藤田さんには多くの実験をサポートしていただき、本当に感謝しております。

そして、同期の小原崇司さん、小谷龍ノ介さん、島田大樹さん、吉田拓也さんに深い感謝の意を示します。様々な方面から研究のディスカッションをしたり、共に研究できたことは私の誇りです。

最後になりましたが、6年間の間、八丈島から暖かく私の研究生生活を見守り、支援をしてくれた両親、妹、祖父母に感謝申し上げ、謝辞とさせていただきます。

参考文献

- [1] 農林水産省. 農林水産基本データ集. <http://www.maff.go.jp/j/tokei/sihyo/>, 2016.
- [2] 農林水産省. 農業生産工程管理 (gap) の共通基盤に関するガイドライン. <http://www.maff.go.jp/j/seisan/gizyutu/gap/guideline/pdf/>, 2012.
- [3] Kazusato Oshima. Plant potyvirus evolution the survey of the genetic structure of populations. *Virus*, Vol. 2, No. 1, pp. 151–160, 2012.
- [4] 宇賀博之, 彌富仁, 川崎雄介, 藤田恵梨香, 鍵和田聡. キュウリウイルス病の画像による診断システムの開発. *植物防疫*, Vol. 70, No. 1, pp. 315–318, 2016.
- [5] 独立行政法人農畜産業振興機構. 野菜の品目別産出額の推移. <http://vegetan.alic.go.jp/toukeiouran.html>, 2016.
- [6] 柳井啓司. 一般物体認識の現状と今後. *情報処理学会論文誌：コンピュータビジョン・イメージメディア*, Vol. 48, No. Cvim 19, pp. 1–24, 2007.
- [7] Yusuke Uchida. Local Feature Detectors, Descriptors, and Image Representations: A Survey. *Computing Research Repository(CoRR)*, *arXiv:1607.08368*, 2016.
- [8] David G Lowe. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision (IJCV)*, Vol. 61, No. 2, pp. 91–110, 2004.
- [9] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. *ECCV International Workshop on Statistical Learning in Computer Vision*, pp. 59–74, 2004.
- [10] F Perronnin and C Dance. Fisher Kenrels on Visual Vocabularies for Image Categorization. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2007.
- [11] Florent Perronnin, Jorge Sánchez, and Thomas Mensink. Improving the Fisher Kernel for Large-Scale Image Classification. *In Proceedings of the 11th European Conference on Computer Vision(ECCV)*, pp. 143–156, 2010.

- [12] Hervé Jégou, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid. Aggregating Local Image Descriptors into Compact Codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 34, No. 9, pp. 1704–1716, 2012.
- [13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, Vol. 115, No. 3, pp. 211–252.
- [14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems (NIPS)*, pp. 1–9, 2012.
- [15] Matthew D Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. *In Proceedings of the 13th European Conference on Computer Vision (ECCV)*, pp. 818–833, 2014.
- [16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Chapel Hill, and Ann Arbor. Going Deeper with Convolutions. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 771–780, 2016.
- [18] Alexey Dosovitskiy and Thomas Brox. Inverting Visual Representations with Convolutional Networks. *Computing Research Repository(CoRR)*, *arXiv:1506.02753*, pp. 1–15, 2015.
- [19] Paul Viola and Michael J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision (IJCV)*, Vol. 57, No. 2, pp. 137–154, 2004.
- [20] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, 2005.
- [21] Pedro F. Felzenszwalb, Ross Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 32, No. 9, pp. 1627–1645, 2010.

- [22] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision (IJCV)*, Vol. 59, No. 2, pp. 167–181, 2004.
- [23] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587, 2014.
- [24] J. R R Uijlings, K. E a Van De Sande, T. Gevers, and a. W M Smeulders. Selective search for object recognition. *International Journal of Computer Vision (IJCV)*, Vol. 104, No. 2, pp. 154–171, 2013.
- [25] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann LeCun. OverFeat: Integrated recognition, localization and detection using convolutional networks. *Computing Research Repository(CoRR)*, *arXiv:1312.6229*, 2013.
- [26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 37, No. 9, pp. 1904–1916, 2015.
- [27] Ross Girshick. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.
- [28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances In Neural Information Processing Systems (NIPS)*, pp. 1–10, 2015.
- [29] Jayme Garcia Arnal Barbedo. Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus*, Vol. 2, No. 1, pp. 660–671, 2013.
- [30] Ms. Kiran R. Gavhale and Prof. Ujwalla Gawande. An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. 16, No. 1, pp. 10–16, 2014.
- [31] Kuo-Yi Huang. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in Agriculture*, Vol. 57, No. 1, pp. 3–11, 2007.
- [32] Song Kai, Zhikun Liu, Hang Su, and Chunhong Guo. A research of maize disease image recognition of corn based on BP networks. *In Proceedings of the International*

- Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 246–249, 2011.
- [33] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad. A framework for detection and classification of plant leaf and stem diseases. *In Proceedings of the IEEE International Conference on Signal and Image Processing (ICSIP)*, pp. 113–118, 2010.
- [34] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, and S. Vishnu Varthini. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal*, Vol. 15, No. 1, pp. 211–217, 2013.
- [35] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino. Deep-plant: Plant identification with convolutional neural networks. *In Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 452–456.
- [36] David Hall, Chris McCool, Feras Dayoub, Niko Sunderhauf, and Ben Upcroft. Evaluation of Features for Leaf Classification in Challenging Conditions. *In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 797–804, 2015.
- [37] Erika Fujita, Yusuke Kawasaki, Hiroyuki Uga, Satoshi Kagiwada, and Hitoshi Iyatomi. Basic investigation on a robust and practical plant diagnostic system. *In Proceedings of the IEEE International Conference on Machine Learning and its Applications (ICMLA)*, pp. 989–992, 2016.
- [38] Sharada Prasanna Mohanty, David P. Hughes, and Marcel Salathé. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, Vol. 7, pp. 1–7, 2016.
- [39] David Hughes and Marcel Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *Computing Research Repository (CoRR)*, *arXiv:1511.08060*, 2015.
- [40] Piotr Dollár, Zhuowen Tu, Pietro Perona, and Serge Belongie. Integral Channel Features. *In Proceedings of the British Machine Vision Conference (BMVC)*, pp. 1–11, 2009.
- [41] Zhuowen Tu. Probabilistic Boosting-Tree : Learning Discriminative Models for Classification , Recognition , and Clustering. *In Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1589–1596, 2005.

- [42] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Face detection without bells and whistles. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 720–735, 2014.
- [43] Adriana Olmos and Frederick A A Kingdom. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, Vol. 33, No. 12, pp. 1463–1473, 2004.
- [44] Yi-qing Wang. An Analysis of the Viola-Jones Face Detection Algorithm. *Image Processing On Line (IPOL)*, Vol. 4, pp. 128–148, 2014.
- [45] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research (JMLR)*, Vol. 10, No. 9, pp. 1755–1758, 2009.
- [46] Alfredo Canziani, Eugenio Culurciello, and Adam Paszke. An Analysis of Deep Neural Network Models for Practical Applications. *Computing Research Repository (CoRR)*, *arXiv:1605.07678v1*, 2016.
- [47] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *Computing Research Repository (CoRR)*, *arXiv:1408.5093*, 2014.
- [48] Alexey Dosovitskiy, Philipp Fischer, Jost Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 38, No. 9, pp. 1734–1747, 2016.
- [49] Ilya Sutskever Joan Bruna Dumitru Erhan Ian J. Goodfellow Rob Fergus Christian Szegedy, Wojciech Zaremba. Intriguing properties of neural networks. *Computing Research Repository (CoRR)*, *arXiv:1312.6199*, 2013.

業績

研究論文

- “キュウリウイルス秒の画像による診断システムの開発,” 宇賀博之, 彌富仁, 川崎雄介, 藤田恵梨香, 鍵和田聡, 植物防疫, Vol.70, No.5, pp.315-318, 2016.

国際会議発表 (主著)

- ”Basic Study of Automated Diagnosis of Viral Plant Diseases using Convolutional Neural Networks,” Yusuke Kawasaki, Hiroyuki Uga, Satoshi Kagiwada and Hitoshi Iyatomi, Lecture Notes in Computer Science, International Symposium on Visual Computing 2015 (ISVC 2015).

国内会議発表 (主著)

- “Convolutional Neural Networks を用いた植物病自動診断支援システムの開発,” 川崎雄介, 宇賀博之, 鍵和田聡, 彌富仁, 2015年電子情報通信学会総合大会 (滋賀), 2015年3月
- “Convolutional Neural Networks を用いたウイルス性植物病害自動診断支援に関する基礎検討,” 川崎雄介, 宇賀博之, 鍵和田聡, 彌富仁, 第31回ファジイシステムシンポジウム (東京), 2015年9月

その他の会議 (共著)

- ”Basic investigation on a robust and practical plant diagnosis system,” Erika Fujita, Yusuke Kawasaki, Hiroyuki Uga, Satoshi Kagiwada and Hitoshi Iyatomi, IEEE International Conference on Machine Learning and Applications 2016 (ICMLA 2016).
- “Convolutional Neural Networks を用いた多種類ウイルス性植物病害自動診断システム,” 藤田恵梨香, 川崎雄介, 鍵和田聡, 宇賀博之, 彌富仁, 2016年電子情報通信学会総合大会 (福岡), 2016年3月