

コンシューマ・システム論文

画像局所特徴の類似度を用いた メロン等級判定システムの開発

小池 誠^{1,a)} 平原 健太郎^{2,b)} 渡辺 隆一^{3,c)} 小川 晋^{4,d)} 峰野 博史^{1,2,3,5,6,e)}

受付日 2022年7月22日, 採録日 2022年11月14日

概要: 高級果物として知られるメロンは、果実の見栄えなど外観の品質が重要視されており、出荷時には形状や色合い、ネットの模様、傷の有無などを加味した等級の選別が行われている。一般的に、このような外観検査は経験豊富な生産者の目視によって行われているが、等級判断基準の一貫性確保といった目的から等級判定のシステム化が望まれている。このような外観検査では近年、深層学習を用いた画像認識モデルを適用する研究が注目されている。しかし、深層学習を用いるためには大量のアノテーションされた画像データを収集する労力が課題となる。そこで、本研究では詳細画像認識 (Fine-grained Image Classification)、画像特徴ベクトルの類似度を用いた距離学習、Activation Map を用いたデータ拡張を組み合わせることにより、データ数が少ない場合であっても一般的な転移学習による画像認識モデルと比較し、高い精度で等級判定可能なモデルを提案する。加えて、この等級認識モデルを用いたメロン果実の等級判定システムのプロトタイプを構築し、実用性に関する基礎検証を実施した。その結果、約 82.1%の精度で熟練生産者と同等の等級判定が可能であることを示した。

キーワード: スマート農業, 詳細画像認識, 距離学習, データ拡張

Development of Melon Grading System Using Similarity Measure of Local Image Features

MAKOTO KOIKE^{1,a)} KENTARO HIRAHARA^{2,b)} RYUICHI WATANABE^{3,c)} SUSUMU OGAWA^{4,d)}
HIROSHI MINENO^{1,2,3,5,6,e)}

Received: July 22, 2022, Accepted: November 14, 2022

Abstract: Melons are known as expensive fruit, and the quality of their appearance is important. Therefore, when shipped, melons are graded by farmers based on shape, color, net pattern, presence of disease, etc. Generally, this inspection is done visually by experienced farmers. However, for consistency of grading criteria, systemization of visual inspection is desired. In recent years, there has been a lot of research on the application of image classification models using deep learning in the field of visual inspection. To use deep learning, it is necessary to collect a large amount of annotated image data. In this study, we propose a model that can determine the grade with higher accuracy than general image classification models based on transition learning (e.g., using imagenet) even when the number of data is small, by combining Fine-grained Image Classification, Metric Learning using similarity measure of an image feature vector, and Data Augmentation using an Activation Map. In addition, we developed a prototype of Melon Grading System using this model and conducted basic tests on its practicality. As a result, we showed that the developed grading system can grade melons with an accuracy of about 82.1% compared to an experienced farmer.

Keywords: smart agriculture, fine-grained image classification, metric learning, data augmentation

1. はじめに

ここ数年、国内の温室メロンの収穫量は、毎年約 1.8 万トン辺りを推移しており、主な産地は静岡県（約 7 千トン）、愛知県（約 4 千トン）、茨城県（約 3 千トン）となっている [1], [2]. 最も収穫量の多い静岡県では、主にアールスメロンの品種が栽培されており、この品種は成長過程で果実の肥大にともない果実表皮が裂け網目を形成することが特徴である。一般的に、温室メロンは贈答などに用いられる高級果物として知られている。そのため、温室メロン生産者は果実の品質を高めるため摘果を行うなど、多くの手間暇をかけて生産している。

果実の品質が重要視される温室メロンは、生産時の手間もさることながら、出荷時も時間をかけ厳しく検査されており、重量や糖度といった大きさや味の品質だけでなく果実の形状が整っていることや、色合い、網目パターンなど、果実の外観の品質も重要視されている。温室メロン生産者は、これら品質にかかわる要因を果実 1 つ 1 つに対して評価し、評価に基づく等級に選別する作業を行っている。現在、この選別作業は熟練生産者の目視評価によって実施されている。そのため、等級の判断を行う生産者の経験や主観に依るところが大きく、生産者の間で等級付けの一貫性が保たれないことが課題である。見た目の品位が評価観点となる網目パターンや形状は、定量的な判断基準を設けることが難しく、現状は生産者を一同に集め認識合わせを行う目揃え会や、個別生産者と出荷組合で 2 段階選別を行うなど、基準を合わせるための取り組みや若手就農者への教育に多くの時間と労力がかかっている。これら青果選別の課題については、秋元ら [3] がカキ果実を対象に分析しており、選別の基準点が選別人ごとにバラつくことや、目揃い会によって一時的に基準点を近づけることができたとしても、時間が経つにつれて元に戻ってしまう課題などが

報告されている。メロン果実の選別作業 (3.1 節) においても同様の課題があることを、静岡県の温室メロン生産者より確認しており、判定基準に一貫性を持った等級判定システムの実現が期待されている。

メロン外観の特徴の定量化については、川上ら [4], [5] が静岡県産のマスクメロンを対象に、ネット幅、ネット交点の分布、交点を結ぶパスの長さに着目し定量化を行い、等級上位の「山」と下位の「白」との比較実験にて、等級上位の方がネット幅はやや大きく、変動も小さく良く揃っていると報告している。また、等級と表面色について、色彩色差計 (ミノルタ CR-300) で L (明度), a (色度), b (色度) を測定し、等級上位の方が果実の果梗部と赤道部で Lab 色空間での色度 b の値が大きくなると報告している。しかし 3 年間の調査から、収穫年度によってこれらが大きくバラつくことも分かっており、網目パターンの構造と表面色に対する分析にとどまり、等級判定までは行われていない。

また近年、深層学習を用いた画像認識技術の発展により、農業分野においても深層学習を用いた研究がさかんに行われている。特に、農作業者の目視判断の代替を目的とした研究では、畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) [6] やそれに関連するモデルが用いられることが多い。ウメを対象とした研究では、デジタルカメラで撮影した果実画像から熟度を判定するシステムが提案されており、5 段階の熟度を 96% の精度で判別できたことが報告されている [7]。その他にも、キュウリやトマトの葉の病害判定 [8], [9] やリンゴ果実の生育状況の判定 [10] など、様々な作物の目視判定に CNN を用いたアプリケーションが研究されており良好な認識精度を示しているが、メロンの等級判定に適用したケースは報告されていない。

そこで本研究は、深層学習に基づく画像認識を用い、網目パターンといった外観の特徴を考慮した温室メロンの等級判定システム (以下、本システム) を提案し、プロトタイプを開発して基礎実験を行った。図 1 に本システムの構成を示す。本システムは、撮影部とサーバ部から構成される。撮影部は、メロンの全周映像をカメラで取得しサーバ部へ送信する。サーバ部は、撮影部から送信されてきた全周映像の前処理と等級判定モデルによる等級の推論を実行し結果を表示する。識別対象となるメロン果実は自然物であるため、外観、特に網目パターンは個体ごとに多様である。加えて、1 個体の全周においても同一パターンが一樣に分布している訳ではなく、見る場所によって網目が細かい、粗い、汚れているといった特徴がある。そのため、等級ごとの特徴を簡潔に定義することが難しい。そこで本システムでは、網目パターンの特徴を定義するのではなく、詳細画像分類 [11] とクラス別マージンを考慮した深層距離学習 [12] を用い、画像間の局所特徴量の「類似度」という

¹ 静岡大学大学院自然科学系教育部情報科学専攻
Graduate School of Science and Technology, Shizuoka University, Hamamatsu, Shizuoka 432-8011, Japan

² 静岡大学大学院総合科学技術研究科
Graduate School of Integrated Science and Technology, Shizuoka University, Hamamatsu, Shizuoka 432-8011, Japan

³ 静岡大学情報学部
Faculty of Informatics, Shizuoka University, Hamamatsu, Shizuoka 432-8011, Japan

⁴ 株式会社大和コンピューター
DAIWA COMPUTER CO., LTD., Takatsuki, Osaka 569-0054, Japan

⁵ 静岡大学学術院情報学領域
College of Informatics, Academic Institute, Shizuoka University, Hamamatsu, Shizuoka 432-8011, Japan

⁶ グリーン科学技術研究所
Research Institute of Green Science and Technology, Shizuoka 422-8529, Japan

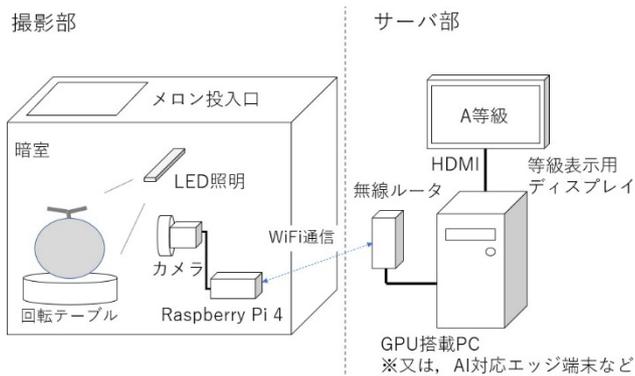
a) koike.makoto.18@shizuoka.ac.jp

b) hirahara.kentaro.18@shizuoka.ac.jp

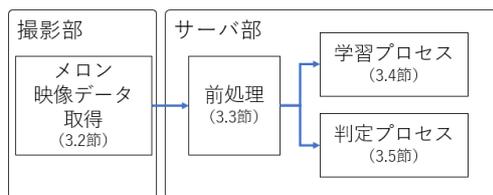
c) watanabe.ryuichi.18@shizuoka.ac.jp

d) susumu.ogawa@daiwa-computer.co.jp

e) mineno@inf.shizuoka.ac.jp



(a) システム構成図 (ハードウェア)



(b) 機能配置図 (ソフトウェア)

図 1 等級判定システムの構成

Fig. 1 Schematic diagram of melon grading system.

指標を用いて等級分類を試みる。そして、本システムの有用性を示すために、プロトタイプを構築しメロン等級判定の基礎実験を行った。

また、深層学習を用いる場合に課題となるのが、質の良い大量のアノテーション済み学習用データを用意する労力である。深層モデルは学習可能なパラメータ数が多いため、学習用データセットに含まれるデータが少ない場合や、分類クラスを代表するようなデータが少ない、または、アノテーションのノイズが多すぎる場合において過学習の問題があり、モデルの汎化性能を著しく低下させることが知られている [13], [14]。本システムの場合、データの収集作業はメロン果実を回転テーブルに乗せるだけであり、誰でも比較的簡単に実施できる。しかし、収集したデータにアノテーションを行うには目利きのできる熟練生産者の協力が不可欠であり、現実的には大量かつ高品質なアノテーション済みデータを取得することは困難である。このような場合、一般的に転移学習によるモデルのファインチューニングや学習データ数を水増しするデータ拡張の適用が有効であることが知られている。そこで本システムでは、ファインチューニングの際に Activation Map [15] を利用したデータ拡張を適用することで、過学習を抑制し汎化性能の向上を図る。

以降、2章で関連研究について述べる。3章で本システムの主要部分であるサーバ部の等級判定モデルの構築方法について述べ、4章は構築した等級判定モデルの有用性に関する基礎実験を行う。5章で結果の考察を行い、最後に6章で本論文をまとめる。

2. 関連研究

2.1 詳細画像分類 (FGIC)

詳細画像分類 (Fine-grained Image Classification) は、一般画像分類で扱われるクラスカテゴリ (犬や車など) に対し、より細かいサブカテゴリ (犬種、車種など) を対象とする分類問題を指す。詳細画像分類では、クラス間のきめ細かい特徴の違いを見つけ出しクラス推定する必要があるため、一般画像分類よりもさらに難易度が高いタスクである。詳細画像分類では、画像空間内からクラス分類に有効な部位を抽出し、抽出部位の局所的な特徴量を分類に利用するアプローチをとることが多い [11]。初期の研究では、画像のラベルに加え、クラス分類に有効な部位をバウンディングボックス形式でアノテーションしたデータセットを作成し、Mask-CNN を用いてクラス分類に有効な部位を推定した後、クラス分類器を学習する手法が提案された [16]。しかし、細かな部位のアノテーションには大きな労力が必要になるという課題があり、部位のアノテーションを行わないデータセットを用いて、教師なしで細かい部位を予測する手法が注目されている。Liu ら [17] は、Region Proposal Network (RPN) による部位推定のサブネットワークと最終的なクラス推定を行うメインネットワークを、メインからサブへ知識蒸留 [18] を用いて学習することにより、ラベル情報のみでクラス推定に有効な部位を学習する手法を提案した。また、Sun ら [19] は、注目部位の抽出に Attention 機構を利用したネットワークを提案した。ネットワークが出力する Attention Map を使用し、画像内の注目すべき部位に絞った特徴抽出を複数回繰り返すことでクラス推定を行った。しかし、これらのアプローチはメインとサブの複数ネットワークが必要であったり、注目部位を絞るためにネットワークの順伝搬を複数回行う必要があったりなど、学習が煩雑になる傾向が見られる。一方 End-to-End な手法では、Chang ら [20] は、明示的な部位推定を学習するのではなく、クラス分類に有効な特徴量を特徴マップのチャンネル上で学習させることで、ネットワークに各クラスに対応する特徴量抽出を学習させるよう設計された損失関数 Mutual-Channel Loss (MC-Loss) を提案した。

本研究では、メロンの等級ごとのきめ細かい網目パターンの違いを学習するために、MC-Loss を使用する。ただし、クラス分類器には MC-Loss で使用される線形分類器ではなく、局所的な特徴の類似度を学習するため深層距離学習を用いる。

2.2 深層距離学習

深層距離学習 [12] は、入力データ間の類似性、または非類似性を確立するために深層学習のアーキテクチャを用いる手法である。入力データを任意の潜在空間上の特徴ベクトルに変換する深層ニューラルネットワークを用い、同

ジラベルのデータであれば変換後の特徴ベクトルがより近い距離に、異なるラベルのデータであればより遠い距離になるようネットワークの学習が設計されている。距離学習を用いたモデルはデータ数が少ない場合やテスト時に未知のクラスが含まれる場合であっても、通常のクラス分類モデルと比較し識別精度が高くなることが報告されており、画像を用いた顔認識・再認証タスクや異常検知タスクなどで深層距離学習を用いる研究が数多く行われている [21], [22], [23], [24]。この深層距離学習を用いたアプローチは、同じ等級であっても多様な網目パターンが存在するメロン画像の分類に対しても有効であると考えられる。

Deng ら [21] が提案した ArcFace は、顔画像を入力とし特徴抽出器で抽出した特徴ベクトルをクラスごとに角度距離による特徴空間に埋め込む。このとき、クラス間に角度マージンを設定することで、異なるクラス間の角度距離がより大きく、同じクラス内の角度距離がより小さくなるよう設計されている。なお、このマージンはハイパーパラメータであり、対象のデータ/タスクに合わせて適切に設定する必要がある。Zhang ら [24] はこのハイパーパラメータをクラス数、または、学習の進み具合から自動的に調整する機構を提案した。しかしながら、どちらもマージンの値はクラス間で共通の値が使用されている。

本研究では、距離学習として ArcFace を用いる。ただし、クラスごとに個別の適切なマージンを設定できるように変更を加えることで認識精度の向上を図る。

2.3 画像データ拡張

画像認識タスクにおいて、データ拡張とは学習データとなる画像に対し、平行移動、上下/左右反転、拡大縮小など、元画像に任意の変換処理を加えて新たな画像を生成することで学習データ数を水増しする技術である。データ拡張を行うことで、過学習を抑制し認識精度を向上させることが広く知られており、画像認識タスクにおいて標準的に使用されるテクニックである。今までにデータ拡張について様々な手法の研究が行われており、2つ以上の画像を混合させ新たな画像データを作り出す手法 [25], [26], [27] なども効果的であることが報告されている。特に、Summers ら [27] は2つの画像の混合手法について広く検証し、提案した14種類の混合方法のすべてにおいて認識精度向上の効果があることを報告した。また、Jiang ら [28] は、画像を用いた異常検知タスクにおいて、データ拡張に Attention Map を用いる手法を提案した。Attention Map を用いて異常検知に大きく寄与する部位をあえてマスクした画像を生成し、ネットワークが異常検知に有益な部位をより多く発見することを促す Attention-guided data augmentation (AGDA) を提案している。

本研究では、メロンのドメイン知識に基づき AGDA の機能を拡張することで、Attention Map 同様に注目部位を

示す Activation Map と2枚の画像を用いた画像混合によるデータ拡張手法を提案する。

3. メロン等級判定システム

本システムは、撮影部とサーバ部で構成される。図1(b)で示すように、撮影部ではカメラを用いてメロンの全周映像データの取得を行い、計算量の多い画像の前処理や等級判定モデルによる等級推論などはサーバ部で実施する。また、等級判定モデルの学習も事前にサーバ部で行うものとする。

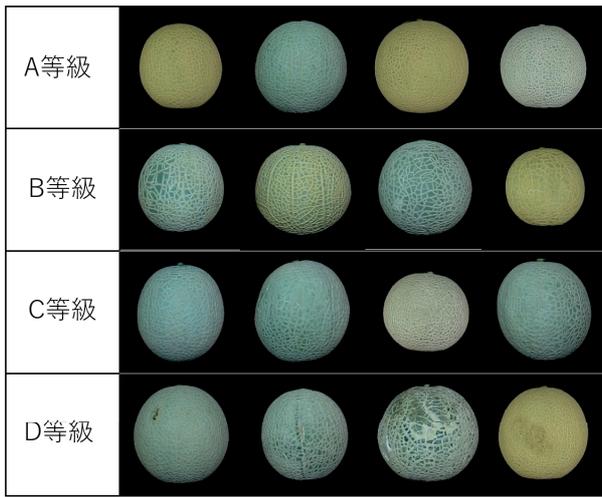
3.1 メロンの等級

本システムで対象としたメロンは、静岡県内で温室栽培されたアールスメロン系品種で、2月から11月にかけて収穫されたものである。出荷時の等級判定は、生産者による目視で行われており、品質が高い順に A, B, C, D の4等級に分類される。各等級の判定基準について、生産者へインタビュー調査を行った結果を下記に示す。

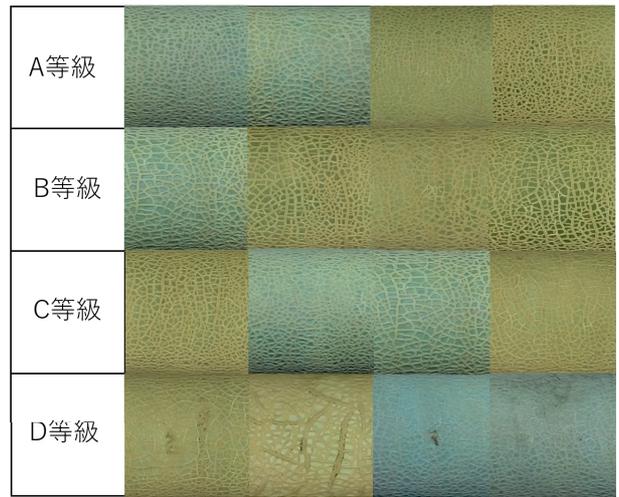
- A 等級：高品質
網目の密度が均一で果実の形状に歪みが少ない。
- B 等級：中品質
網目の乱れが目立つものの、果実の形状に歪みが少ない。
- C 等級：低品質
果実形状に扁平、面長、凸など歪みがある。
- D 等級：低品質
果実に傷、汚れ、黒カビ、著しい網目の乱れがある。

各等級の特徴のサンプルを図2に示す。図2(a)は輪郭の特徴であり、きれいな円形なのがA, B等級となり、形状が歪なのがC等級となる。なお、D等級は網目の特徴で判断されるため輪郭は不問である。図2(b)は網目の特徴であり、きめ細かいものがA等級、ばらつきが大きいものがB等級、傷や汚れが確認できるものがD等級となる。なお、C等級は形状で判断されるため網目は不問である。

等級付けの優先順位は、 $D > C > B > A$ 等級の順となり、品質が低い等級が優先される（例：CとDの特徴を持ったメロンはD等級となる）。また、A, B, D等級は網目の特徴によって判断され、C等級のみ果実の形状の特徴によって決定される。なお、網目の密度や形状の歪みについて定量的な判断基準はなく、生産者の主観によって判断される。当然、生産者は事前に目揃え会などの教育を受けているため、おおむね等級ごとの傾向は揃うものの、明確な基準がないことによるバラつきが含まれる。また、等級の判断根拠となる形状や網目パターンは、果実の局所的な部位にのみ現れる傾向にある。



(a) 輪郭の特徴



(b) 網目の特徴

図 2 各等級のサンプル

Fig. 2 Samples of each grade.

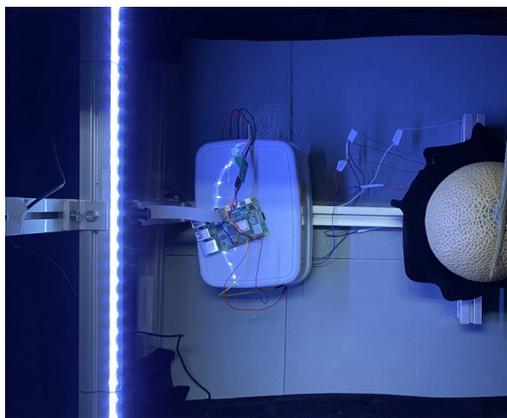


図 3 撮影部 (プロトタイプ)

Fig. 3 Camera part of melon grading system (prototype).

3.2 撮影部

撮影部を図 3 に示す。撮影部は、外部周辺の明るさの影響を軽減するため、暗幕で覆われたボックス内に設置する。ボックス内には、RGB カメラモジュール (OV5647) と LED 照明と重量センサ付きの回転テーブルがあり、それらを Raspberry Pi4 を用いて制御する。Raspberry Pi とサーバ部の PC は Wi-Fi 通信で接続されており、メロンを回転テーブル上に乗せると重量を感知し、撮影準備完了であることをサーバ部の PC へ通知し撮影スタンバイ状態となる。スタンバイ状態で、ユーザからの撮影開始の操作を受け付けると、Raspberry Pi がカメラの撮影を開始するとともに回転テーブルが回り始める。回転テーブルが 360° 回転すると撮影が停止し、撮影した映像データがサーバ部へ送信される。映像データは、解像度 1,000 × 1,000 ピクセル、25 FPS (全 599 フレーム) の MP4 フォーマットで保存される。

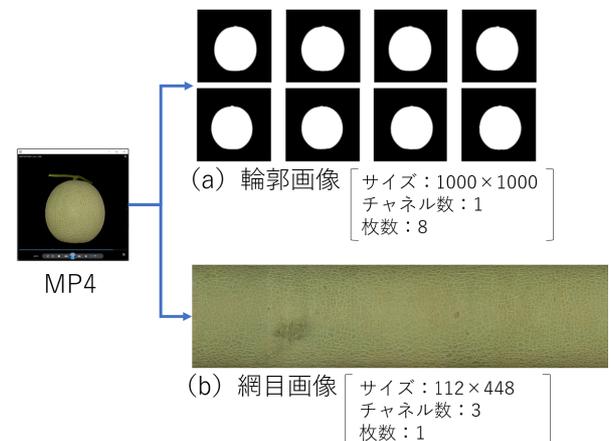


図 4 画像の前処理

Fig. 4 Image preprocessing.

3.3 画像の前処理

撮影部で撮影された映像データから画像へ変換する方法について述べる。変換方法は、輪郭画像と網目画像の 2 通りある。なお、これらの画像変換処理はすべて画像処理ライブラリである OpenCV [29] を用いて行った。

● 輪郭画像

輪郭画像の例を図 4(a) に示す。メロン果実の 360° 全周映像から、果実の向きが 45° ごとになるよう合計 8 フレームを画像として切り出す。切り出したフレーム画像に対し、背景領域と果実領域を区別できるような画像の 2 値化を行い、果実領域に対し収縮 (Erosion) 処理と膨張 (Dilation) 処理を各 20 回適用することで、果実領域画像から細い茎部分のみを消去した輪郭画像を生成する。

● 網目画像

網目画像の例を図 4(b) に示す。メロン果実の 360° 全周映像から、先頭の 1 フレームを画像として切り出す。切り出したフレーム画像に対し、背景領域と果実領域を区別で

きるよう画像の2値化を行う。次に、2値化画像に対し楕円フィッティングを行い、果実の中心座標、高さ、幅の情報を取得する。次に、中心座標を中心に幅1ピクセルで高さ分の領域を切り出す。これを全599フレームに対して行い、各フレームで切り出した画像を横方向に連結し、幅599ピクセルの画像を生成する。これは、簡易的なパノラマ画像を生成する方法と同じである。最後に、生成した画像を縦112×横448ピクセルにリサイズし網目画像とする。

3.4 メロン等級判定モデル

メロン等級判定モデルの学習アーキテクチャを図5に示す。メロン等級判定モデルは、大別すると輪郭画像から輪郭特徴ベクトルに変換する部分と網目画像から網目特徴ベクトルに変換する部分から構成され、最終的には両特徴ベクトルを結合し線形変換後にL2正規化したものをメロン1玉の特徴ベクトルとする。なお、メロン1玉のデータ x に対する正解ラベルは4等級を表す $y \in \{0, 1, 2, 3\}$ であるが、それとは別に輪郭画像と網目画像に対しても、それぞれラベルを付与する。輪郭画像 x^{Shape} のラベル $y^{Shape} \in \{0, 1\}$ は、画像がC等級であるかないかを表し、網目画像 x^{Net} のラベル $y^{Net} \in \{0, 1, 2\}$ は、画像がA, B, Dのどの等級であるかを表す。

3.4.1 輪郭ベクトルの埋め込み

輪郭画像から輪郭特徴ベクトルへの埋め込みは、輪郭画像から算出した輪郭情報の統計量を使用する。図6に示すとおり、輪郭画像から果実の横幅、高さ、アスペクト比、10°ごとの果実中心から輪郭までの長さを算出し、全8枚の平均値、最小値、最大値、標準偏差を求める。次に、算出したすべての統計量の中から、ランダムフォレストを用いた再帰的特徴量削減法[30]を用いて、果実形状に歪みのあるC等級 y^{Shape} の判定に有効な統計量を選択する。選択する統計量の数は、全統計量を使用した場合の正答率(約97%)を基準に、正答率を低下させずに最小となる数として実験的に求めた。表1に輪郭画像から算出した統計量とともに再帰的特徴量削減法により選ばれた7つの統計量(下線を引いたもの)を示す。果実形状の歪みを判断する際、熟練生産者は果実の縦横幅のバランスや一部が盛り上がったような歪みを確認している。本システムで選択された7つの特徴量は、それらを反映した統計量になっており、完全一致とはいえないまでも、妥当性のある特徴量選択ができていると考える。なお、この特徴量選択はArcFaceを用いた学習を行う前に実施しておき、以降、図5のEmbeddingモジュールでは入力された輪郭画像に対し、7つの輪郭特徴量の算出だけを行う。

3.4.2 網目特徴ベクトルの抽出

網目特徴ベクトルの抽出は、網目画像から画像全体の特徴を抽出するのではなく、画像の等級を表す数~数十ピクセルの局所的なクラス固有の特徴を検出することが目的で

ある。したがって、MC-Lossを使い畳み込みニューラルネットワーク(CNN)により構成されたエンコーダを学習する。まず、入力となる網目画像はエンコーダによって特徴マップに変換される。このとき、CNNの畳み込み層の最終層が出力する抽象度の高い特徴マップを使用するのではなく、局所的な特徴を用いるため抽象度の低い中間の畳み込み層の出力を使用する。具体的には、エンコーダにResNet50[31]を使用する場合、Conv5ブロックの7×7の特徴マップを用いるのではなく、1つ手前のConv4ブロックの14×14の特徴マップを用いる。

変換された特徴マップを、高さ H 、幅 W 、チャンネル数 N として $F \in \mathbb{R}^{W \times H \times N}$ とする。このとき、MC-Lossを用いる場合は $N = C \times \varepsilon$ となるよう設定する必要がある。ここで C は分類クラスの数で、 ε はクラスごとの特徴マップのチャンネル数である。その結果、クラス i に対応する特徴マップは、 $F_i \in \mathbb{R}^{W \times H \times \varepsilon}$ 、 $i = 0, 1, \dots, C-1$ と表すことができる。MC-Lossではこの特徴マップに対し損失を次のように定義する。

$$L_{MC}(F) = L_{dis}(F) - \lambda \times L_{div}(F) \quad (1)$$

$L_{dis}(F)$ はエンコーダが抽出した特徴の識別性に関する損失で、クロスエントロピー損失関数 L_{CE} を用いて次のように表す。

$$L_{dis}(F) = L_{CE} \left(y^{Net}, \frac{[e^{g(F_0)}, e^{g(F_1)}, \dots, e^{g(F_{C-1})}]^T}{\sum_{i=0}^{C-1} e^{g(F_i)}} \right) \quad (2)$$

ここで、 $L_{CE}(y, \hat{y})$ は正解ラベル y とモデルの推定ラベル \hat{y} のクロスエントロピー損失である。また、関数 $g(\cdot)$ は次のように定義する。

$$g(F_i) = \frac{1}{WH} \sum_{k=1}^{WH} \max_{j \in \{0, 1, \dots, \varepsilon-1\}} [M_i \cdot F_{i,j,k}] \quad (3)$$

ここで、 $M_i = \text{diag}(Mask_i)$ 、 $Mask_i \in \mathbb{R}^\varepsilon$ であり、関数 $\text{diag}(\cdot)$ は引数のベクトルを対角行列の対角成分に配置する。このとき、 $Mask_i$ の半分(ランダムに選択された $\varepsilon/2$ 分)を0とし、もう半分を1とする。いい換えると、式(3)の $[M_i \cdot F_i]$ はクラス別の特徴マップ F_i のチャンネルの半分を0に置き換えることであり、チャンネル単位のドロップアウトと考えることができる。

$L_{div}(F)$ は抽出した特徴の多様性に関する損失で、次式で表す。

$$L_{div}(F) = \frac{1}{C} \sum_{i=0}^{C-1} h(F_i) \quad (4)$$

ここで、関数 $h(\cdot)$ は次のように定義する。

$$h(F_i) = \sum_{k=1}^{WH} \max_{j \in \{0, 1, \dots, \varepsilon-1\}} \left[\frac{e^{F_{i,j,k}}}{\sum_{k'=1}^{WH} e^{F_{i,j,k'}}} \right] \quad (5)$$

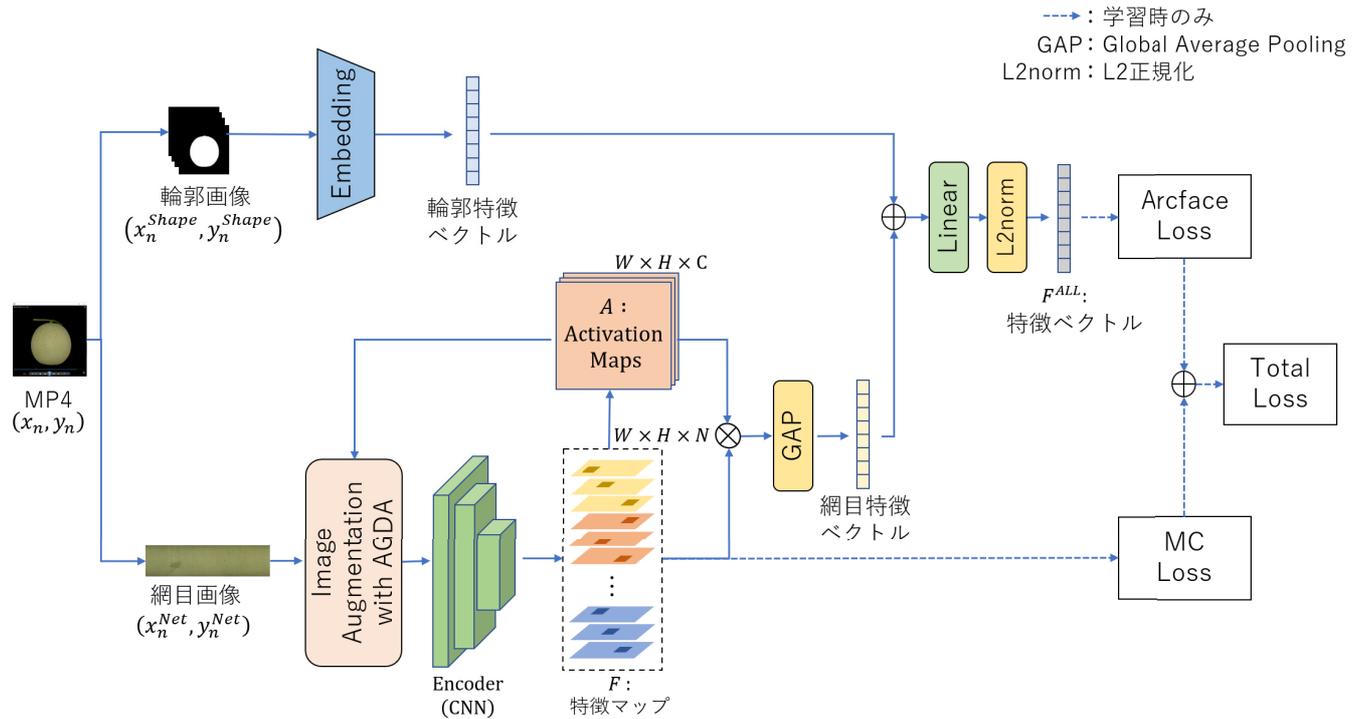


図 5 等級判定モデルのアーキテクチャ

Fig. 5 Architecture of melon grading model.

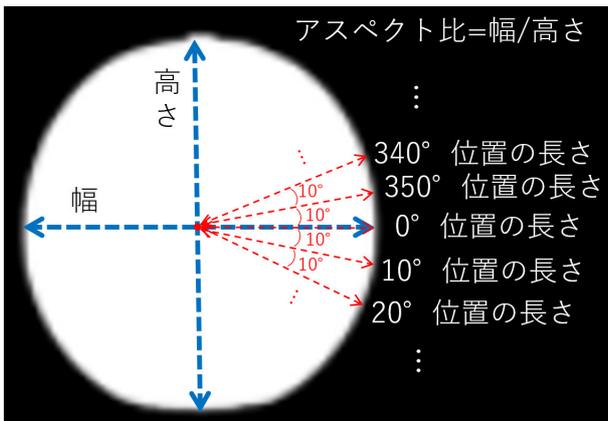


図 6 輪郭特徴量の算出

Fig. 6 Computation of contour features.

この $L_{div}(F)$ は、エンコーダによって抽出されたクラスごとの特徴マップが、チャンネル方向で重複していない場合にボーナスを与える。具体的には、式 (1) で表すとおり、パラメータ λ を掛けた値が損失から減算される。この損失関数は、エンコーダがなるべく重複のない多様な特徴抽出を学習するように作用する。

3.4.3 Activation Map の生成

Activation Map は、特徴マップを集約することで画像空間上のどの部位の特徴がモデルの推定結果に寄与しているかを可視化する。これは説明可能な AI (XAI) 技術の 1 つとして知られており、Zhou ら [15] の Class Activation Map (CAM) では、エンコーダが出力する特徴マップと全結合層の重みを用いて Activation Map : $A \in \mathbb{R}^{W \times H \times C}$ を算出し

表 1 輪郭画像から算出した特徴量 (下線: 特徴量選択で選ばれた特徴量)

Table 1 Features computed from contour image (underlined: selected features).

| 項目 | 統計量 |
|------------------------|--|
| 幅 | 最小値, 最大値, 平均値, 標準偏差 |
| 高さ | 最小値, 最大値, 平均値, 標準偏差 |
| アスペクト比 | 最小値, 最大値, <u>平均値</u> , 標準偏差 |
| 果実中心から輪郭までの長さ (10° ごと) | 最小値, 最大値, 平均値, 標準偏差 <u>(20° 位置の平均値, 20° 位置の最大値, 170° 位置の平均値, 170° 位置の最大値, 180° 位置の最大値, 240° 位置の最大値)</u> |
| ※図 6 参照 | |

ている。本システムにおける Activation Map は、基本的には CAM と同じ考え方で生成するが、全結合層の重みを用いない点が異なる。本システムの生成手順を図 7 に示す。エンコーダが出力する各クラスの特徴マップ F_i にカーネルサイズ 1×1 のチャンネル方向の最大プーリング関数 (Cross-Channel Max-Pooling: CCMP) を適用した後、チャンネル方向のソフトマックス関数 (Channel-wise Softmax: CWSM) を適用することで、各特徴マップ F_i に対応したクラス別の Activation Map A_i , $i = \{0, 1, \dots, C - 1\}$ を求める。そして、各クラスの特徴マップをクラス数分 (= C) まとめて本システムの Activation Map : $A \in \mathbb{R}^{W \times H \times C}$ とする。言い換えると、Activation Map は特徴空間 $W \times H$ 上

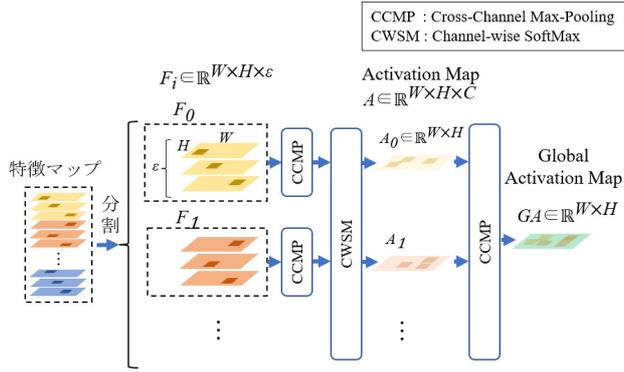


図 7 Activation Map の生成手順

Fig. 7 Generation method of Activation Map.

で, CWSM を適用することにより, 各クラスに対する寄与度を 0.0~1.0 の範囲, かつ, 合計が 1.0 になる形式で表現した 3 次元配列データとなる. さらに, A に対し CCMP を再度適用し, 各クラスの Activation Map を集約した Global Activation Map: $GA \in \mathbb{R}^{W \times H}$ を求める. つまり, GA はあるクラスの特徴が表れている画像部位の値が大きく, どのクラスの特徴も表していない部位の値が小さくなるように表現した 2 次元配列データといえる. Activation Map A と Global Activation Map GA は以下の式で定義する.

$$A = \text{softmax} \left(\left[\max_{j \in \{0, 1, \dots, C-1\}} [F_{i,j}] \right], i = 0, 1, \dots, C-1 \right) \quad (6)$$

$$GA = \max_{i \in \{0, 1, \dots, C-1\}} [A_i] \quad (7)$$

本システムでは, エンコーダが出力する特徴マップに対し GA を掛け合わせた後, 全体平均プーリング (GAP) でチャンネルごとに平均値を取り出しベクトル化したものを網目の特徴ベクトルとする. また, A を後述する画像データ拡張の手法として用いる.

3.4.4 クラス別マージンを用いた特徴ベクトルの埋め込み

輪郭特徴ベクトルと網目特徴ベクトルを連結し, 線形変換後に L2 正規化したベクトルをメロンの特徴ベクトル $F^{ALL} \in \mathbb{R}^{N+7}$ とする. この特徴ベクトルを, ArcFace を用いてベクトルの類似度を指標とした角度距離空間上に埋め込む. ArcFace の損失関数は次の式で表すことができる.

$$L_{AC}(F^{ALL}) = L_{CE} \left(y, \frac{[e^{s(\cos \theta_0)}, \dots, e^{s(\cos(\theta_y + m_y))}, \dots, e^{s(\cos(\theta_{C-1}))}]^T}{e^{s(\cos(\theta_y + m_y))} + \sum_{j=0, j \neq y}^{C-1} e^{s(\cos \theta_j)}} \right) \quad (8)$$

$$\cos \theta_y = \frac{W_y^T \cdot F^{ALL}}{\|W_y\| \|F^{ALL}\|} \quad (9)$$

ここで, $\cos \theta_y$ は特徴ベクトル F^{ALL} とその正解ラベル y に対応する重みベクトル W_y とのなす角度を表す. また, s と m はハイパーパラメータである. ただし, 本システム

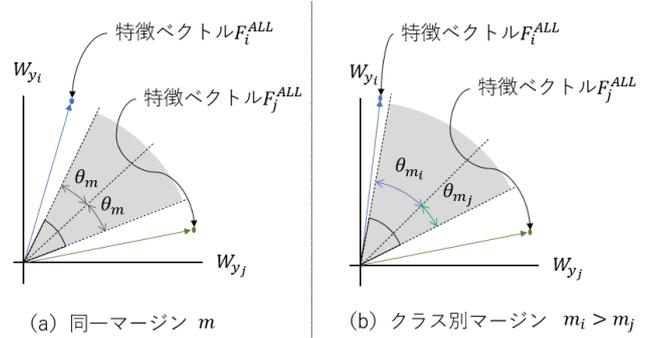


図 8 クラス別マージンを用いた埋め込み

Fig. 8 Embedded using margin by classes.

ではマージンに少し変更を加え, クラスごとに値を設定可能とした. これにより, 図 8 に示すように灰色の領域がクラス間のマージン領域となり, この領域を避けて特徴ベクトルが埋め込まれる. クラスごとに異なるマージン領域を設定することで, 特徴ベクトルを埋め込む領域を調整することが可能となる. 今回扱うメロンは, A 等級は比較的均一な品質でありクラス内のばらつきが少なく, 逆に C や D 等級は傷など様々な不良パターンが存在するため, クラス内のばらつきが大きいことが経験上分かっている. そこで, データの性質に合わせ, クラス内のばらつきが少ないデータは埋め込む領域を小さく, クラス内のばらつきが大きいデータは埋め込む領域を大きく取って過学習の軽減を図る.

最後に, MC-Loss の損失と ArcFace の損失にハイパーパラメータ $\alpha = \{0.0 \sim 1.0\}$ を掛けて合計した値を, モデル全体の損失 L として学習を行う.

$$L(F^{ALL}, F) = L_{AC}(F^{ALL}) + \alpha(L_{MC}(F)) \quad (10)$$

3.4.5 Activation Map によるデータ拡張 (AGDA)

データ拡張は過学習を抑制するテクニックとして, 画像認識モデルの学習に広く適用されている. 本システムにおいても, 学習に使用できるデータ数が少ない中, 過学習を抑制するためにデータ拡張が有効であると考えられる. 本システムでは, 入力される網目画像に対し基本的なデータ拡張手法であるカラージッタ, ランダム反転 (上下/左右), 画像ローリング (横軸方向) を適用する. なお, ランダムスケールなど, 網目の見た目が大きく変わってしまう変換は対象外とした. さらに, 2つの画像を組み合わせたデータ拡張についても検討する. 網目画像の場合, ラベル情報を反映した網目の特徴が局所的にしか存在しないため, CutMix [25] のようにランダムに画像を切り出す手法では, 元ラベルを正しく反映した画像を生成することができない. そこで, AGDA [28] と同様に, Activation Map を用いて等級判定に大きく寄与した特徴が存在する部位のみを切り出す方法を採用する. ただし, AGDA では元画像の対象部位をマスクするのみであったが, 本システムでは 2 枚の画像

を組み合わせ新しい画像を生成する．具体的には，拡張対象の網目画像とそのラベルを (x_n^{Net}, y_n^{Net}) とし，同じラベルの別の網目画像を (x_m^{Net}, y_m^{Net}) とする．網目画像 x_m^{Net} をエンコーダへの入力として 3.4.3 項で述べた手順で生成した Activation Map を A^m とし，生成した A^m の中で拡張対象のラベル y_n^{Net} に対応するクラス別の Activation Map を $A_{i=y_n^{Net}}^m$ とする．新たな網目画像 \hat{x}_n^{Net} は， $A_{i=y_n^{Net}}^m$ の値を用いて網目画像 x_n^{Net} と網目画像 x_m^{Net} をアルファブレンドすることで生成し，学習データに加える．

$$\hat{x}_n^{Net} = (1 - A_{i=y_n^{Net}}^m) \times x_n^{Net} + A_{i=y_n^{Net}}^m \times x_m^{Net} \quad (11)$$

このように Activation Map を用いて等級判定に大きく寄与した画像部位のみを画像合成対象とすることで，画像の特徴とラベルの整合性を担保したデータ拡張が可能となる．

3.5 判定プロセス

学習済みモデルを用いて等級判定する方法について，2通りの方法を考えた．等級の判定には，学習済みモデルが出力する特徴ベクトル F^{ALL} を使用する．また，画像の前処理などのプロセスは学習時と同様に行うこととする．

● 最大類似度による判定

事前に学習データを用いて，各クラスの平均特徴ベクトル $REF_{i=0,1,\dots,C-1}$ を求めておく．判定対象のメロンの特徴ベクトル F_n^{ALL} と平均特徴ベクトル REF_i のコサイン類似度を求め，最も似ている平均特徴ベクトルのクラスをそのメロンの等級 y'_n とする．以下にその式を示す．

$$y'_n = \operatorname{argmax}_{i \in \{0,1,\dots,C-1\}} \left(\frac{F_n^{ALL} \cdot REF_i}{\|F_n^{ALL}\| \|REF_i\|} \right) \quad (12)$$

● 距離しきい値による判定

最大類似度による判定と同様に，事前に平均特徴ベクトルを求めておき，判定対象のメロンの特徴ベクトルと平均特徴ベクトルのコサイン距離を求め，クラスごとのしきい値 $T_{i=0,1,\dots,C-1}$ との比較によって等級を判定する．クラス数 $C = 4$ の場合の判定方法を以下に示す．ここで，どの等級を優先的に判断するかは，判断対象によって調整可能である．メロンの場合は， $D > C > B > A$ ($3 > 2 > 1 > 0$) 等級の順に判断する．

$$y'_n = \begin{cases} 3, & d_3 < T_3 \\ 2, & d_2 < T_2 \\ 1, & d_1 < T_1 \\ 0, & \text{それ以外} \end{cases} \quad (13)$$

$$d_i = 1 - \frac{F_n^{ALL} \cdot REF_i}{\|F_n^{ALL}\| \|REF_i\|} \quad (14)$$

しきい値を用いた判定のメリットは，運用時に D 等級の判定を少し厳しくするなど，運用に合わせた事後調整が可能になることである．

4. 基礎実験

4.1 データセット

基礎実験で使用したデータセットは，3.2 節で述べた撮影部を用いて撮影した各等級 122 玉とアノテーションされていない 80 玉のメロン全周映像データである．データの取得は，2021 年 2 月から 2021 年 12 月にわたり 18 回実施され，等級ラベルのアノテーションは同一の熟練生産者によって実施されている．

4.2 実験条件

本実験で使用する等級判定モデルは，輪郭特徴として表 1 における下線で示した 7 つの特徴量を使用し，網目特徴を抽出するエンコーダは，ImageNet で事前学習された ResNet50 を採用した．MC-Loss にかかわる設定として，エンコーダへの入力画像サイズは $114 \times 448 \times 3$ のカラー画像とし，ハイパーパラメータ $\varepsilon = 5$ ， $\lambda = 10$ とした．エンコーダは ResNet50 の畳み込み層 Conv4 ブロックの後に畳み込み層を 1 層追加し， $7 \times 28 \times 15$ の特徴マップを出力する．ArcFace にかかわる設定として，ハイパーパラメータ $s = 10$ ， $m = [0.8, 0.4, 0.4, 0.4]$ ， $\alpha = 2$ とした．ここでマージンの配列 m は，先頭からそれぞれ A, B, C, D 等級に対するマージンを表す．学習は，モーメンタムを 0.9 に設定した確率的勾配降下法を使用し，バッチサイズは 32，学習率は 0.001 としエポックごとに $e^{-0.01}$ 減衰させ 250 エポック学習させた．テストデータに対する等級判定のプロセスは，基礎実験では運用時のしきい値調整を考慮しないため，3.5 節に述べた最大類似度による判定で行った．なお，等級判定モデル (図 5) の実装は，Python と機械学習ライブラリである TensorFlow [32]，scikit-learn [33] を用いて行った．ImageNet で事前学習された ResNet50 モデルは，TensorFlow が提供するモデルを利用し，輪郭ベクトル埋め込み処理は sklearn.feature_selection.RFE モジュールを利用し，それ以外は新規実装した．また，学習時の損失関数である ArcFace と MC-Loss は，それぞれ参考文献を元に TensorFlow で実装した．

4.3 等級判定精度の比較

提案する等級判定モデルを用いた等級判定精度に関する実験を行った．この実験では，各等級 122 玉のデータを用い，学習用データとして各等級 $N \in \{32, 52, 72, 92\}$ 玉，テスト用データとして各等級 30 玉を用いた．まず，学習用データを用いて等級判定モデルの学習を行い，テストデータを用いて等級判定を行った．実験は，AGDA ありの場合と AGDA なしの場合とで，交差検証法を用いて 3 回行い，学習に使用したデータ数別の正答率を調べた．また，比較対象として，通常の ResNet50 と ViT-B/16 [34] を用いた場合と，人間の熟練生産者が判断した場合について実

験を行った。熟練生産者の実験では、メロン全周映像を見て等級を判断してもらった。ここで、熟練生産者は映像での等級判定をふだん行っていないため、映像からの等級判定に慣れてもらうためのトレーニングを事前に実施し、十分判定できる状態になったのちに本実験を実施した。また、ResNet50 と ViT-B/16 を用いた実験では、どちらも ImageNet の事前学習モデルを用い、網目画像を入力とし輪郭特徴ベクトルは GAP 後に連結して、一般的な分類問題としてファインチューニングを行った。

学習の結果、すべてのモデルで損失の収束を確認できた。学習データ数別の正答率を表 2 に示す。最も正答率が高かったのは 82.1% で、提案したモデル ($N = 92$) を AGDA ありで学習した場合であった。人間の熟練生産者が判定した場合でも 84.6% であったことから、人間に近い精度で等級を判定できていることが確認できる。また、少ないデータ数 ($N = 32$) であっても比較対象としたモデルより、提案したモデルの方が高い精度で等級判定できることが確認できた。

次に、熟練生産者と提案モデル ($N = 92$, AGDA あり) の混同行列を表 3 に示す。熟練生産者も提案モデルも比較的 A, B 等級間の判定に誤りが多いことが確認できた。また、果実形状の歪みの判定に着目すると、熟練生産者が A, B 等級を C 等級と誤判定したのは 2 玉であったが、提案モデルは 13 玉と誤判定が多く C 等級における適合率低下の要因となっていることが確認できた。一方、D 等級においては、提案モデルの再現率、適合率はどちらも 90% 以上であり、高精度で等級判定可能であることが確認できた。

4.4 クラス別マージンに関する検証

クラス別マージンを用いた特徴ベクトルの埋め込み (3.4.4 項) と、等級判定精度の関係について実験を行った。学習用データ数 $N \in \{32, 92\}$ として、マージンなし $m = [0.0, 0.0, 0.0, 0.0]$ から、全クラスマージンあり $m = [0.8, 0.8, 0.8, 0.8]$ の間でマージンの付け方を変え、AGDA なしで学習を行った。これを、交差検証法を用いて 3 回行い、マージン別の正答率を求めた。

クラス別マージンの正答率を表 4 に示す。学習データ数によらず、A 等級のみマージンを大きく、その他の等級のマージンを半分を設定した場合が最も正答率が高い結果となった。また、特徴ベクトルの主成分分析による散布図を図 9 に示す。マージンがない場合 (図 9 (a)) は、クラス間の識別性が低く正答率が低くなり、逆に、マージンが大きすぎる場合 (図 9 (c)) は、学習データに対するクラス間の識別性は高くなるが、テストデータが学習データから離れた位置に分布する傾向が見られ等級判定精度も低下したと考えられる。A 等級のみマージンを大きく、それ以外を半分とした場合 (図 9 (b)) は、比較的学習データとテストデータが同じ領域に分布していることが確認できる。

表 2 学習データ数別の正答率

Table 2 Accuracy by number of training data.

| Method | Accuracy (%) | | | |
|----------------|--------------|--------|--------|-------------|
| | $N=32$ | $N=52$ | $N=72$ | $N=92$ |
| 熟練生産者 | 84.6 | | | |
| ResNet50 | 55.8 | 61.1 | 63.5 | 72.1 |
| ViT-B/16 | 62.1 | 65.0 | 67.7 | 72.9 |
| 提案モデル (AGDAなし) | 64.3 | 69.7 | 71.7 | 81.7 |
| 提案モデル (AGDAあり) | 65.2 | 70.8 | 72.9 | 82.1 |

表 3 熟練生産者と提案モデルの混同行列

Table 3 Confusion matrix between expert and proposed model.

| | 熟練生産者の判定 | | | | 再現率 [%] |
|---------|----------|------|------|------|---------|
| | A | B | C | D | |
| A | 55 | 3 | 1 | 1 | 91.7 |
| B | 21 | 37 | 1 | 1 | 61.7 |
| C | 7 | 0 | 52 | 1 | 86.7 |
| D | 0 | 0 | 1 | 59 | 98.3 |
| 適合率 [%] | 66.3 | 92.5 | 94.5 | 95.2 | |

| | 提案モデルの判定 ($N=92$, AGDAあり) | | | | 再現率 [%] |
|---------|-----------------------------|------|------|------|---------|
| | A | B | C | D | |
| A | 38 | 14 | 7 | 1 | 63.3 |
| B | 6 | 46 | 6 | 2 | 76.7 |
| C | 1 | 0 | 59 | 0 | 98.3 |
| D | 0 | 5 | 1 | 54 | 90.0 |
| 適合率 [%] | 84.4 | 70.8 | 80.8 | 94.7 | |

表 4 クラス別マージン設定ごとの正答率

Table 4 Accuracy by each class margin setting.

| m | Accuracy (%) | |
|----------------------|--------------|-------------|
| | $N=32$ | $N=92$ |
| [0.0, 0.0, 0.0, 0.0] | 60.5 | 75.8 |
| [0.8, 0.4, 0.4, 0.4] | 64.3 | 81.7 |
| [0.8, 0.8, 0.4, 0.4] | 62.2 | 75.8 |
| [0.8, 0.8, 0.8, 0.4] | 63.3 | 77.9 |
| [0.8, 0.8, 0.8, 0.8] | 62.8 | 77.9 |

4.5 AGDA の有効性と生成画像の妥当性の確認

AGDA の有効性とデータ拡張された画像の妥当性について確認する。250 エポック学習を行った時点のデータ拡張で生成された画像例を図 10 に示す。図中の白点線は Activation Map を用いて合成された部位である。

4.3 節の実験結果 (表 2) から、AGDA ありの提案モデルとなしの提案モデルの結果を比較することで、データ拡張の適用により正答率が平均 0.9% 向上することが確認できた。また図 10 から、生成される画像が元ラベルの特徴表現を反映した自然な網目画像となっていることが確認で

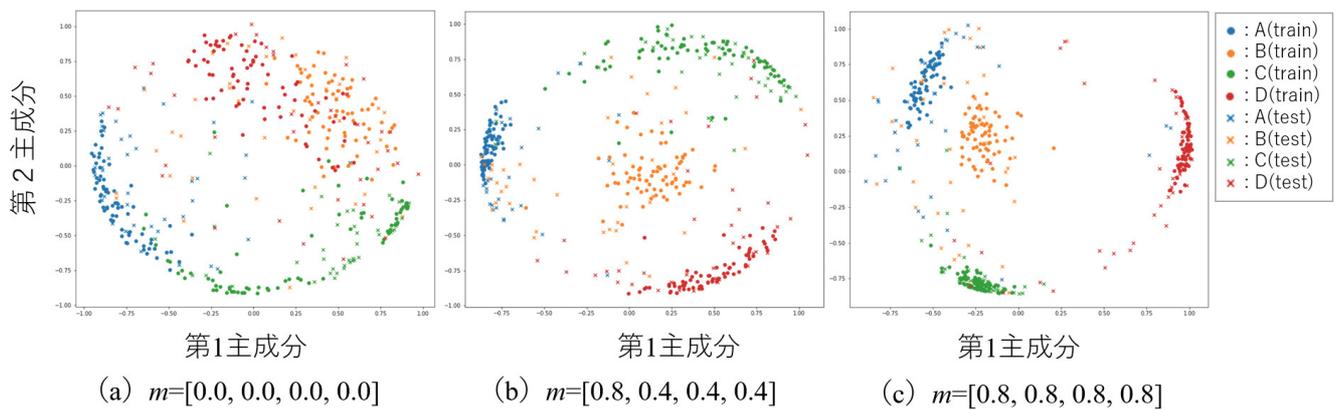


図 9 クラス別マージン設定ごとの特徴ベクトルの主成分分析結果 (N = 92)

Fig. 9 PCA of feature vectors by each class margin setting.

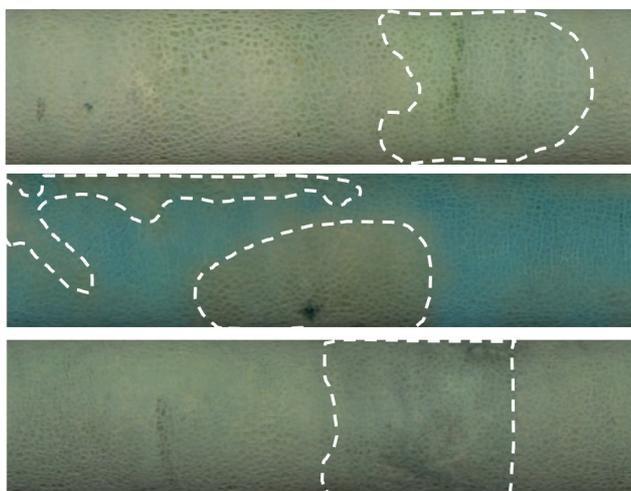


図 10 AGDA によるデータ拡張の画像例 (D 等級)

Fig. 10 Example images of data augmentation with AGDA (rank D).

きた。

4.6 等級判定モデルの定性評価

等級判定モデルを用いた判定結果が、熟練者の感覚とどの程度一致しているかを検証するため、モデル判定結果に対し定性評価を行った。4.3 節の実験の結果から、アノテーションを行った熟練者であっても、後日改めて等級判定を行うと 15%程誤ってしまうことが分かった。メロンの中には、熟練者であってもどの等級に分類するか迷ってしまう個体も含まれている。そこで、データセットを使った正答率の評価に加え、定性的な評価も実施することとした。判定対象となるメロン映像データと等級判定モデルの判定結果を熟練者に提示し、モデルの判断の妥当性を 4 段階で評価してもらった。

- I. 判断が一致している
- II. 一致していないが許容できる
- III. 一致していないし許容し難い
- IV. 明らかに間違っている (見間違いである)

表 5 等級判定モデルの定性評価結果

Table 5 Qualitative evaluation results of melon grading model.

| システムの 判定結果 [等級] | 熟練者の評価 [回答数] | | | | 一致率 (%) | 許容率 (%) |
|-----------------------|-----------------|-----|------|-----|------------|------------|
| | I. | II. | III. | IV. | | |
| A | 24 | 7 | 7 | 2 | 60.0 | 77.5 |
| B | 21 | 9 | 6 | 4 | 52.5 | 75.0 |
| C | 29 | 3 | 5 | 2 | 74.4 | 82.1 |
| D | 40 | 1 | 0 | 0 | 97.6 | 100.0 |
| Total | 114 | 20 | 18 | 8 | 71.3 | 83.8 |

2 名の熟練者にそれぞれアノテーションしていない 80 玉に対するモデルの判定結果の妥当性を評価してもらった。熟練者の評価結果を表 5 に示す。システムが判断した等級別に、熟練者の 4 段階評価の回答数を集計した。等級別に熟練者の評価が「I.」であった割合を一致率、「I. または II.」であった割合を許容率として実験結果の評価を行ったところ、一致率は 71.3%であり許容率は 83.8%であった。また、D 等級では許容率が 100%であり、熟練者がおおむね妥当であると判断する等級判定ができていることが確認できた。

5. 考察

等級判定精度の比較結果 (表 2) から、各等級 92 玉のメロンを学習することで正答率 82.1%の等級判定モデルを構築できることが確認できた。これは人間の熟練者の 84.6%より若干劣るものの、人間とほぼ同程度の等級判定が実現できたといえる。それを裏付ける結果として、定性評価においても 83.8%と高い確率で、熟練生産者の目から見ても妥当であると判断された。特に、D 等級に関しては許容率 100%であることから、高精度な不良判定の自動化などの活用も期待できる。学習データ数を 32 玉から 92 玉に増加させるにつれて判定精度も向上することが確認できた。さらにデータを集めることで、熟練生産者以上の一貫性をもって等級判断が可能になることも期待できる。



(a) システムはA等級と判定したが、熟練者は網目が若干不均一であるためB等級と判定
(b) システムはC等級と判定したが、熟練者は強い網目を傷と判断してD等級と判定

図 11 熟練者が許容できないと判断したメロンの例

Fig. 11 Examples judged by expert to be unacceptable.

一方で、表 3 の結果より、提案モデルは果実形状の特徴を見る C 等級の判定で誤りが多いことから、輪郭特徴ベクトル埋め込み手法に改善の余地が確認できた。改善案としては、今回は 7 つの統計量を用いたが、統計量に加え網目特徴ベクトルのように画像特徴量を用いるなどが考えられる。また、表 5 の結果より、熟練生産者が許容できないと判断したメロン (図 11) を確認すると、網目の大きさや均一度が A と B 等級の中間に位置するようなパターンが多いことが確認できる。特に、最高品質である A 等級に対する判断は厳しくなる傾向にあり、許容できないと判断される割合が大きくなっていった。また、B, C 等級では、傷と強めに出ている縦方向の網目のきわどいパターン判断で不一致となっている場合と、学習データにほとんど含まれないレアケースの傷パターンの判断が不一致となる場合が多く確認できた。よって、輪郭特徴ベクトル埋め込み手法の改善に加え、クラス間の境界に位置するようなデータを積極的に集めることによって、等級判定精度の向上が期待できる。また、レアケースへの対応として、3.5 節で述べた距離しきい値による判定を適用することで、どの等級にもあてはまらない場合は人間に判断を委ねるといったシステム設計も考えられる。

表 2 の結果より、ResNet50 を一般的な画像分類モデルとして使用する場合と比較し、本システムは平均 5.8% の精度向上が確認できた。これにより、メロンの等級判定では、画像全体ではなく局所的な特徴に絞った等級判定が有効であることを示すことができた。また、図 10 に示すように、AGDA によるデータ拡張を用いることで、メロン画像として違和感ない形で新たなデータを生成でき、判定精度が平均 0.9% 向上することが確認できた。従来、このようなデータ拡張を行うためには、事前に特徴部位をアノテーションしておく必要があったが、AGDA によるデータ拡張を用いることで部位のアノテーションを行う労力を大幅に削減できる可能性を示すことができた。

6. まとめ

本研究では、メロン外観の局所的な特徴の類似度に基づく等級判定システムを提案し、プロトタイプを用いて有用性について基礎的な検証を行った。ImageNet で学習済みの ResNet50 をベースに MC-Loss, クラス別マージンを加味した ArcFace, AGDA を組み合わせたシステムを構築することで、約 82% の精度で等級判定が可能なシステムを構築し、熟練生産者が約 84% で許容できる等級判定を行えることを確認した。

今後は、さらなる判定精度向上や季節性を考慮した精度評価を行うため、運用しながらデータ蓄積が可能なシステムへの改良を検討する。また、Activation Map により等級判断の根拠となる部位を提示できることを活用し、本システムを新規就農者に対する等級判定技能の習得支援システムとして展開することも検討していく。

謝辞 本研究の一部は、JST 創発的研究支援事業 (JP-MJFR201B) ならびに東北大学電気通信研究所における共同プロジェクト研究の支援を受けたものである。また、データセットならびに有意義なディスカッションの機会をいただいた (株) 大和コンピュータの皆様に深い感謝の意を表す。

参考文献

- [1] 品目別 (季節区分別) 作付面積, 収穫量及び出荷量累計統計: 農林水産省ホームページ統計情報, 入手先 (<https://www.maff.go.jp/j/tokei/kouhyou/kensaku/hin5.html>) (参照 2021-07-23).
- [2] 静岡県公式ホームページ「My しずおか日本一」, 入手先 (<http://www.pref.shizuoka.jp/j-no1/m.melon.html>) (参照 2021-07-23).
- [3] 秋元浩一, 黒田佐俊: カキ果実の等級選別について (第 4 報) —選別基準に及ぼす目揃会の効果, 農業機械学会誌, Vol.43, No.1, pp.121–126 (1981).
- [4] 川上昭太郎, 坂口栄一郎, 梅田重夫: メロンのネットパターン, 農業機械学会誌, Vol.58, No.1, pp.17–32 (1996).
- [5] 川上昭太郎, 早川千吉郎, 梅田重夫, 坂口栄一郎: メロンの表面色と品質, 農業機械学会誌, Vol.56, No.3, pp.117–122 (1994).
- [6] Krizhevsky, A., Sutskever, I. and Hinton, G.E.: Imagenet Classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, pp.1097–1105 (2012).
- [7] 建本 聡, 原田陽子, 今井健司: 深層学習を利用したウメ「露茜」の画像による熟度分類, 農業情報研究, Vol.28, No.3, pp.108–114 (2019).
- [8] Fujita, E., Kawasaki, Y., Uga, H., Kagiwada, S. and Iyatomi, H.: Basic investigation on a robust and practical plant diagnostic system, *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp.989–992 (2016).
- [9] Agarwal, M., Singh, A., et al.: ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network, *Procedia Computer Science*, Vol.167, pp.293–301 (2020).
- [10] 源野広和, 小林一樹: 深層学習を用いたデータクレンジングとリンゴ果実画像への応用, 農業情報研究, Vol.29,

No.3, pp.47–61 (2020).

[11] Zhao, B., Feng, J., Wu, X. and Yan, S.: A Survey on Deep Learning-based Fine-grained Object Classification and Semantic Segmentation, *International Journal of Automation and Computing*, Vol.14, No.2, pp.119–135 (2017).

[12] Kaya, M. and Bilge, H.: Deep Metric Learning: A Survey, *Symmetry*, Vol.11, No.9, 1066 (2019).

[13] Ying, X.: An Overview of Overfitting and its Solutions, *Journal of Physics: Conference Series (JPCS)*, Vol.1168 (2019).

[14] Ghosh, A., Kumar, H. and Sastry, P.S.: Robust Loss Function under Label Noise for Deep Neural Networks, *Proc. 31st AAAI Conference on Artificial Intelligence*, pp.1919–1925 (2017).

[15] Zhou, B., Khosla, A., Lapedriza, A. and Torralba, A.: Learning deep features for discriminative localization, *Proc. IEEE Conference on Computer Vision of Pattern Recognition*, pp.2921–2929 (2016).

[16] Wei, X., Xie, C., Wu, J. and Shen, C.: Mask-CNN: Localization parts and selecting descriptors for fine-grained bird species categorization, *Pattern Recognition*, Vol.76, pp.704–714 (2018).

[17] Liu, C., Xie, H., Zha, Z., Ma, L., Yu, L. and Zhang, Y.: Filtration and Distillation: Enhancing Region Attention for Fine-Grained Visual Categorization, *Proc. AAAI Conference on Artificial Intelligence*, Vol.34, No.7, pp.11555–11562 (2020).

[18] Hinton, G., Vinyals, O. and Dean, J.: Distilling the knowledge in a neural network, arXiv preprint arXiv:1503.02531 (2015).

[19] Sun, M., Yuan, Y., Zhou, F. and Ding, E.: Multi-attention Multi-Class Constraint for Fine-grained Image Recognition, *Proc. European Conference on Computer Vision*, pp.805–821 (2018).

[20] Chang, D., Ding, Y., Xie, J., et al.: The Devil is in the Channels: Mutual-Channel Loss for Fine-Grained Image Classification, *IEEE Trans. Image Processing*, Vol.29, pp.4683–4695 (2020).

[21] Deng, J., Guo, J., Xue, N. and Zafeiriou, S.: ArcFace: Additive Angular Margin Loss for Deep Face Recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4690–4699 (2019).

[22] Liu, W., Wen, Y., et al.: SpheroFace: Deep hypersphere embedding for face recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.6738–6746 (2017).

[23] Wang, H., Wang, Y., et al.: Cosface: Large margin cosine loss for deep face recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5265–5274 (2018).

[24] Zhang, X., Zhao, R., et al.: AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.10823–10832 (2019).

[25] Yun, S., Han, D., et al.: CutMix: Regularization strategy to train strong classifiers with localizable features, *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp.6023–6032 (2019).

[26] Takahashi, R., Matsubara, T. and Uehara, K.: Ri-cap: Random image cropping and patching data augmentation for deep cnns, *Asian Conference on Machine Learning*, pp.786–798 (2013).

[27] Summers, C. and Dinneen, M.J.: Improved mixed-

example data augmentation, *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp.1262–1270 (2019).

[28] Jiang, L., Wang, Y., Tang, Z., Miao, Y. and Chen, S.: Casting defect detection in X-ray images using convolutional neural networks and attention-guided data augmentation, *Measurement*, Vol.170, 108736 (2021).

[29] OpenCV (ver.3.8), available from (<https://opencv.org/>) (accessed 2021-07-23).

[30] Sklearn.feature_selection.RFE, available from (https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html) (accessed 2022-07-11).

[31] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[32] TensorFlow (ver2.7.0), available from (<https://www.tensorflow.org/>) (accessed 2022-09-13).

[33] scikit-learn (ver.1.0.2), available from (<https://scikit-learn.org/stable/index.html>) (accessed 2022-09-13).

[34] Dosovitskiy, A., Beyer, L., et al.: An image is worth 16x16 words: Transformers for image recognition at scale, arXiv2010.11929 (2020).



小池 誠 (学生会員)

2005年静岡大学大学院情報学研究科情報学専攻修了。2018年静岡大学大学院自然科学系教育部情報科学専攻入学。農業に関連したデータ分析やシステム開発の研究に従事。



平原 健太郎 (学生会員)

2022年静岡大学情報学部卒業。2022年同大学院総合科学技術研究科入学。画像処理や機械学習の研究に従事。



渡辺 隆一

2018年静岡大学情報学部入学。2021年より機械学習を用いた植物の品質評価, および, 自走車を用いた栽培データ収集の研究に従事。



小川 晋

2012年近畿大学大学院水産学専攻博士前期課程修了。2015年(株)大和コンピューター農業開発部に所属。自社農場にて農業生産を行いつつ、農業生産の経験を生かし栽培システムの基盤研究に従事。



峰野 博史 (正会員)

1999年静岡大学大学院理工学研究科計算機工学専攻修了。同年日本電信電話(株)入社。NTTサービスインテグレーション基盤研究所を経て、2002年10月より静岡大学情報学部助手。2018年4月より同大教授。博

士(工学)。Heterogeneous Network Convergence, Agri-Cyber Physical Human System等に関する研究に従事。電子情報通信学会, 人工知能学会, IEEE, ACM各会員。本会シニア会員。