

半教師あり学習を用いた果実画像からのメロン等級判定システムの検討

小池 誠¹ 平原健太郎² 渡辺隆一² 小川 晋³ 峰野博史⁴

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見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。このような外観検査は、経験豊富な生産者による目視確認が一般的であるが、近年の深層学習分野の発展により、外観検査に深層学習による画像認識モデルを適用する研究も注目されている。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問題となる。そこで、本研究では画像の類似度という尺度に着目し、深層距離学習とパッチ合成画像によるデータ拡張手法と、半教師あり学習を組み合わせることで、少ないラベルありデータから等級判定を行うシステムを検討する。回転テーブルを備えメロン果実の全周画像を取得する撮影部と、撮影した画像から等級を判定するサーバ部から構成される等級判定システムのプロトタイプを構築し、実現性についての基礎検証を実施した。その結果、画像の類似度という距離尺度を基にした等級判定モデルによって、約 60%の精度で等級判定が可能であることを確認した。また、等級判定システムが出力するデータ間の距離と、人間の等級判断との相関について分析することで、距離尺度を用いることの有用性を検討した。

キーワード: 深層学習, 半教師あり学習, メロン等級判定

1. はじめに

国内の温室メロンの収穫量は、ここ数年では毎年約 1.8 万トンで推移しており、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、温室メロンとして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表皮が裂けて網目を形成することが特徴である。一般的に、温室メロンは贈答用などに用いられる高級果物として知られており、温室メロン生産者は高級品を生産するため 1 株から 1 玉だけ成長させるよう摘果を行うなど、多くの手間をかけて生産されている。

高級果実として流通している温室メロンは、出荷時の検査も厳しく、重量や糖度といった大きさや味の品質だけではなく、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、果実の外見の品質も重要視されている。温室メロン生産者は出荷の際、これらの品質に関わる要因を果実 1 つ 1 つに対して評価し、品質ごとの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経験や主観に依るところが大きく、選別を行う者の間で等級付けの一貫性が保たれないことが課題となっている。特に、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のための目揃え会の実施や、個別生産

者と出荷組合で 2 段階の選別を行うなど、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。

メロンの外観の特徴の定量化については、川上ら [3,4] 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」との比較実験では、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、等級と表面色の関係について、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 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留まり、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盛んに行われている。特に、農作業者が作物に対し目視で行う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） [5] やそれに関連するモデルが用いられている。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、CNN を用いて 5823 枚の画像を学習し、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 [6]。一方、キュウリを対象とした研究では、キュウリの葉の画像から病害を判断するシステムが提案されており、健康な葉の画像と 7 種類の病害画像の計 8 クラスの分類を、5 層の CNN を用いて 7320 枚の画像で学習し、平均 82.3% の精度で判別できたことが報告されている [7]。また、トマトを対象とした研究でも、健康なトマトの葉の画像と 9 種類の病害画像の計 10 クラスの画像を 7000 枚含

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
2 静岡大学情報学部
Faculty of Informatics, Shizuoka University
3 (株) 大和コンピュータ
DAIWA COMPUTER CO., LTD.
4 静岡大学学術情報領域/グリーン科学技術研究所
College of Informatics, Academic Institute, Shizuoka University /
Research Institute of Green Science and Technology

むオープンアクセスなデータセット: Plant Village dataset[8]を使用し, 5層のCNNで学習を行い, 91.2%の精度で分類できたことが報告されている[9]. その他にも, リンゴ果実の生育状況の判定[10]やイネの病害[11]など, 様々な作物の状態判定に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まだない.

そこで本研究では, 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評価を行った. 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. サーバ部には画像認識モデルが配置されており, 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. 等級判定モデルの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様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・荒い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. そこで本研究では, 網目パターンの特徴を定義するのではなく, CNNに基づく距離学習[12]を用い, 画像間の「類似度」という指標で等級分類を試みる. 等級ごとに基準となる参照データを設定し, 最も類似度が高い参照データの等級をそのメロンの等級とする.

また, 深層学習を用いる場合に問題となるのが, 深層モデルを学習するための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, データ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]. 本システムの場合は, 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るため, 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必要不可欠であり, 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のは困難である. そのため, 本研究では少量のアノテーションされたデータと, アノテーションされていないデータの両方を使用して学習を行う半教師あり学習を採用する. アノテーションされていないデータも学習に用いることで, 学習用データ数を増やし, 過学習を低減させて等級判定精度の向上を図る.

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効性につ

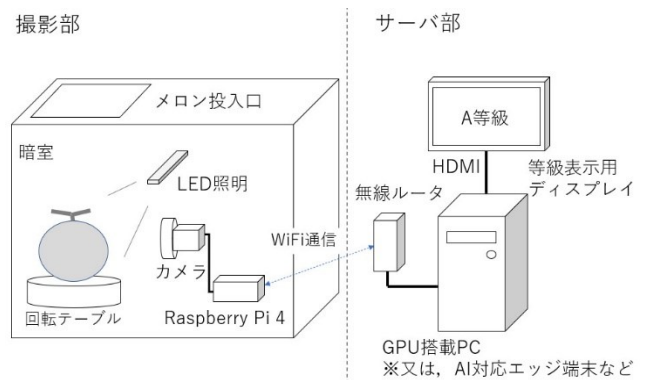


図1 メロン等級判定システムの構成

いて基礎実験を行う. 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

本論文では, メロンの外見の類似度という指標に基づく等級判定システムを提案する. その際, 深層距離学習の手法と半教師あり学習の手法を組み合わせ, 入力画像のバッチ合成によるデータ拡張を行うことで, 教師データ収集の労力の低減を図る. 本章では, 本システムに関連する研究について述べる.

2.1 深層距離学習

深層距離学習[16]は, 入力データ間の類似性, 又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, 似た入力であれば変換後の特徴ベクトルがより近い距離に, 似ていない入力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進むよう設計されている. 画像を入力とした深層距離学習が用いられる研究テーマの1つに, 顔認識・再認証タスクがあり, 顔画像から特定の人物かを判断するために深層距離学習を用いる研究が数多くなされている[12,17,18,19]. この深層距離学習を用いた距離指標によるアプローチは, メロンの網目パターンがどの等級のものに似ているかを定量化するために適用できると考える.

2.2 半教師あり学習

半教師あり学習[15]は, 教師あり学習と教師なし学習を組み合わせることを目的とした機械学習の一分野である. 特に分類タスクに焦点を合わせた研究が多く, 十分な量のアノテーションされた学習用データが入手できない状況において, データ分布における特定の仮定の下, アノテーションされていないデータも利用することで最終的なモデルの精度向上を図る. 例えば, 個々のデータをノードとしてデータ間の距離尺度に基づき近傍グラフを構築した際, 隣

接するノード(=データ)は同じラベルを持つという仮定の下、アノテーションされていないデータに擬似ラベルを付与する。そして、擬似ラベルありデータを学習データセットに加え、データ数を増強したのちに再度学習を行う手法[20]などが知られている。

2.3 パッチ合成によるデータ拡張

RICAP[21], CutMix[22]など、複数の入力画像からランダムに任意領域のパッチ画像を切り出し、それらを組み合わせたパッチ合成画像を新たな学習データとするデータ拡張手法が提案されている。この手法によって、CNNの過学習を防ぎ精度向上に有効であることが報告されている。

本システムで扱うメロン画像についても同様に、複数画像から切り出した画像を組み合わせたパッチ合成画像を入力画像とする。ただし、メロン画像の場合は特徴となる網目パターンが、等級ごとに一目見てわかるほどの明確な差異がなく、また、その差異も果実表面に一様に分布しているわけでもない。そのため、切り出した画像が切り出し元画像のラベルを必ずしも反映していないことを考慮する必要がある。そこで、本システムで用いるデータ拡張はデータの内挿のみに留めることとし、異なるクラスから切り出した部分画像との組み合わせは行わず、ラベルの不一致が比較的発生しない同一クラス内の画像の組み合わせのみを対象とする。

3. メロン等級判定システム

本システムは、メロンの全周画像を撮影する撮影部と取得した画像の前処理と等級判定を行うサーバ部で構成される。図2に本システムの機能構成図を示す。撮影部では、カメラを用いて映像データの取得のみを行い、プロトタイプでは、計算量が多い画像の前処理、深層モデルを使った等級の推論処理などはサーバ部で実施する。また、深層モデルの学習も事前にサーバ部で行うものとする。

3.1 メロンの等級判定

本システムが対象とするメロンは、静岡県内で温室栽培されたアールスメロン品種で、2月から7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、今回は簡単のため品質の順にA, B, Cの3等級に分類することとした。等級判定の観点について、等級判定者へインタビュー調査した結果を下記に示す。

- ・ A 等級 (品質高)
網目の密度が均一で果実の形状に歪みが少ない。
- ・ B 等級 (品質中)
網目の密度に乱れがあるものの果実の形状に歪みが少ない。

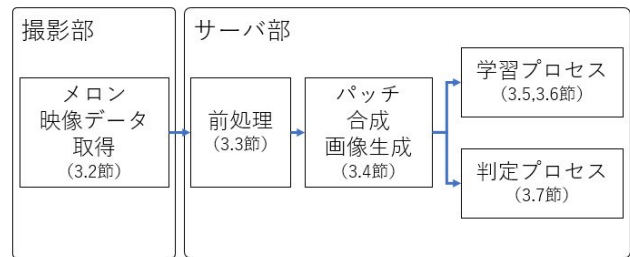


図2 等級判定システムの機能構成図

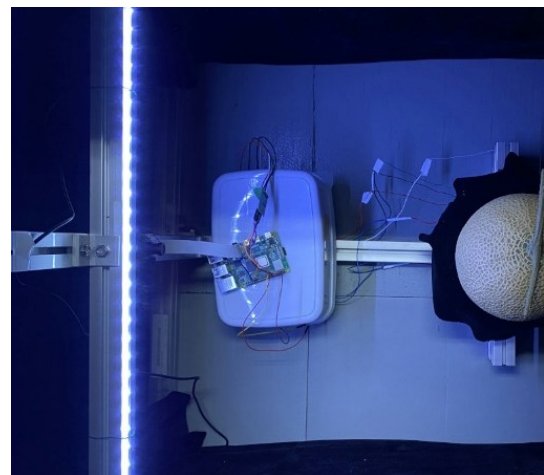


図3 メロン撮影部

- ・ C 等級 (品質低)
網目の密度に著しい乱れがある。又は、果実の形状が歪んでいる。

なお、網目の密度や形状の歪みについての定量的な基準はなく、等級判定者の主観によって判断されるのが一般的である。もちろん等級判定者は、事前に教育や目揃え会などにより認識合わせを行っているため、概ね等級ごとの傾向は揃うものの、明確な基準がないことや等級判定者ごとに基準がばらついてしまうことが課題である。

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される。ボックス内には、RGBカメラ(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により制御する。Raspberry Piとサーバ部PCはWiFi通信で接続されており、メロンを回転テーブルの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PCへ通知し撮影スタンバイ状態となる。撮影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がカメラによる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をサーバ部へ送信する。映像データは、解像度1000×1000pix, 599フレーム(25fps)

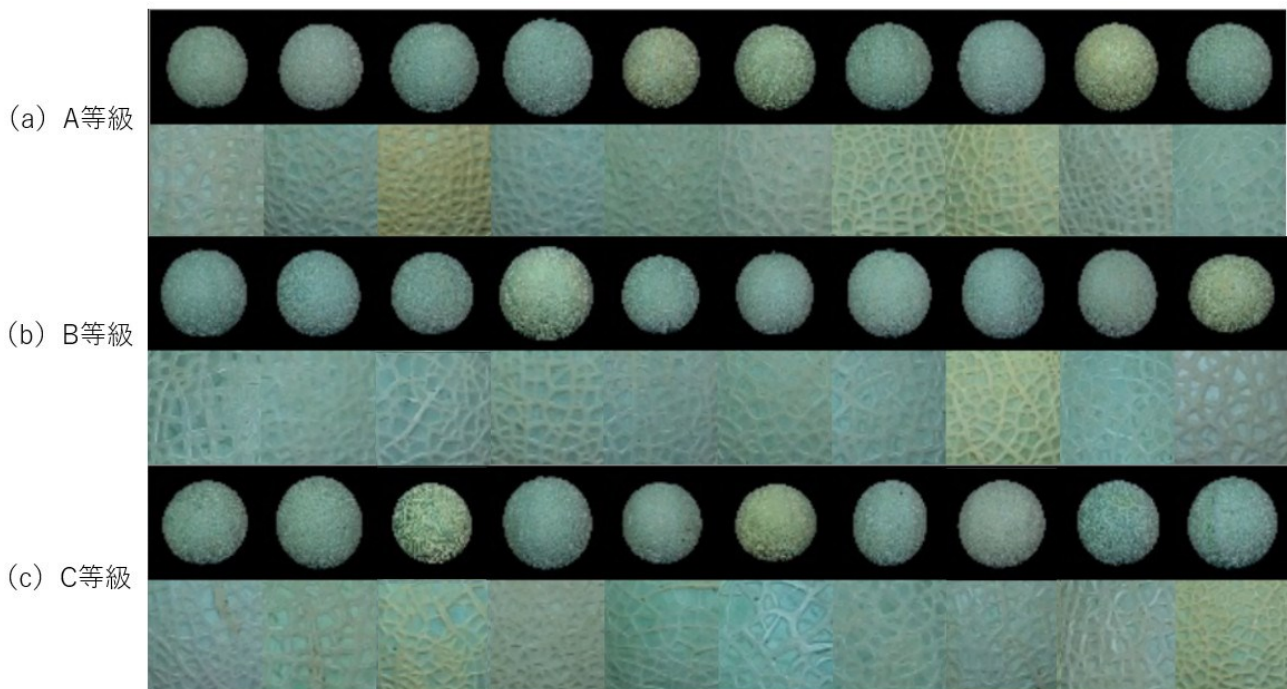


図4 等級別の切り出し画像例（上段：輪郭切り出し画像 下段：網目パターン切り出し画像）

であり、MP4 フォーマットで保存される。

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処理方法について述べる。まず、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。次に、切り出したフレーム画像から、背景領域と果実領域が区別できるよう 2 値画像に変換し、果実領域の収縮 (Erosion) 処理と膨張 (Dilation) 処理を各 20 回適用することで、果実領域から茎部分を消去する。その後、果実領域に輪郭の楕円フィッティングを行い、果実領域の中心座標 (x, y) を取得する。取得した中心座標を基準に元フレーム画像から、下記に述べる 2 種類の画像を切り出し、サイズを $56 \times 56 \text{pix}$ として保存する。

- 輪郭切り出し画像

果実全体の形状の特徴を含む画像を切り出す。元フレーム画像に対し、左上座標 $(x - 390, y - 390)$ 、右下座標 $(x + 390, y + 390)$ となる矩形領域を切り出す。

- 網目パターン切り出し画像

網目パターンの特徴が反映された画像を切り出す。 -100 以上 $+100$ 以下の様に分布する整数の中からランダムに取得した $s_i \in [-100, 100] (i = 0, 1)$ を中心座標からのシフト量とし、左上座標 $(x + s_0 - 100, y + s_1 - 100)$ 、右下座標 $(x + s_0 + 100, y + s_1 + 100)$ となる矩形領域を切り出す。

1 つのメロン全周映像から切り出す枚数は、輪郭切り出し画像を 8 枚、網目パターン切り出し画像を 80 枚とする。なお、これらの処理はすべて画像処理ライブラリである OpenCV[23] を用いて行った。図 4 に等級ごとの切り出し画像例を示す。

3.4 等級判定に使用するパッチ合成画像

等級判定に使用する画像は、 $56 \times 56 \text{pix}$ の切り出し画像を 4×4 の格子状に結合することで生成する。その結果、生成後の画像サイズは $224 \times 224 \text{pix}$ となる。結合対象の切り出し画像は、輪郭切り出し画像から 4 枚、網目パターン切り出し画像から 12 枚をランダムに選択し、 4×4 の格子の上段横一列に輪郭切り出し画像を配置し、それ以外に網目パターン切り出し画像を配置することとした。図 5 最左部に生成画像の例を示す。この方法により、1 つのメロンから最大約 1.2×10^{27} 枚の画像を生成できるため、データ拡張によって過学習を防ぐ効果も期待できると考える。

3.5 深層距離学習モデル

画像の類似度の学習に用いる深層距離学習アーキテクチャを図 5 に示す。入力画像からの特徴量抽出には EfficientNet[24] の畳み込み層の出力を用いる。EfficientNet は、CNN の効率的なスケリング方法と精度に関する研究を基に提案されたアーキテクチャであり、B0 から B7 へとスケールアップ可能である。EfficientNet を本システムに採用することで、要求精度に応じて最適な構成を選択可能になると考える。EfficientNet の後段に 1 層の畳み込み層を追加し 4×4 の格子ごとの特徴量を抽出した後、1 次元配列へ

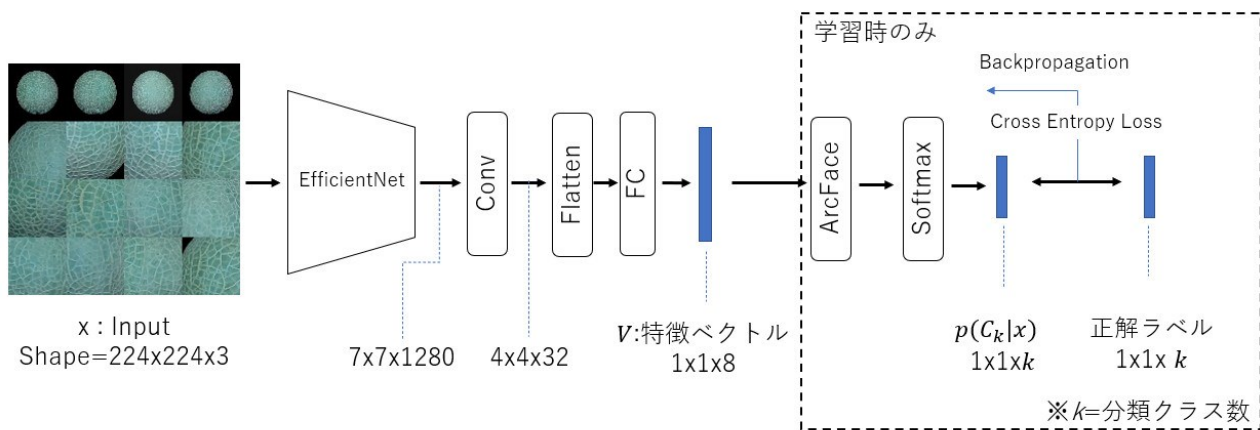


図5 学習モデルのアーキテクチャ

変形し、最終的には全結合層を通し8次元の特徴量ベクトルを出力するネットワーク構成とした。なお、EfficientNetの学習対象パラメータの初期値は、学習効率の向上と過学習防止を目的として、大規模画像データセットであるImagenetで事前学習された値を用いてFine-tuningを行う。

モデルの学習時には、EfficientNetで抽出した8次元特徴ベクトルは、深層距離学習の手法であるArcFace[12]を用い、各クラスの基準を表す学習パラメータとのマージンを加味した角度(=類似度)が算出される。最終的な出力にはSoftmaxを適用し、入力を x 、分類クラス数を K 、分類クラスを $C_k(k=1, \dots, K)$ としたとき、各クラスの事後確率 $p(C_k|x)$ となる。以降は、一般的なCNNの学習と同様に誤差逆伝搬法によりネットワークの学習を行う。

3.6 学習プロセス

モデルの学習は、半教師あり学習を用いる。最初の100epochは、ラベルありデータのみを用いた教師あり学習を行う。次に、学習中のモデルで出力した特徴ベクトル V を用い、似た特徴ベクトルを持つラベルなしデータに対し、ラベル拡散法[20]を用いて擬似ラベルを付与する。ラベル拡散法では、特徴ベクトルをL2正規化した後、 k 近傍法を使って5つの最近傍ベクトルのラベルから擬似ラベルを推定する。そして、擬似ラベルを含めたラベルありデータを使って再度100epochの学習を行う。この擬似ラベルの付与と100epochの学習を1ステップとし、任意のステップ数だけ学習を繰り返す。

3.7 等級判定

学習済みモデルを用いて等級を判定する方法について、2種類の方法を述べる。等級判定時には学習済みモデルの入力層から特徴ベクトルの出力までの層を使用する。また、図2に示すメロンの全周映像からパッチ合成画像生成までのプロセスは、学習時と同様に行うこととする。まず事前に学習済みモデルにラベルありデータを入力し、ラベルあ

りデータの特徴ベクトルを取得し、等級ごとに平均を取った等級ごとの平均特徴ベクトル $REF_k(k=1, \dots, K)$ を求め、各等級を表す参照ベクトルとする。等級判定を行う際は、対象となるメロンの全周映像からパッチ合成画像を生成し、学習済みモデルを用いて特徴ベクトル V に変換する。これを1つのメロンに対し8回行い、取得した8個の特徴ベクトルの平均を、そのメロンを表す平均特徴ベクトル \bar{V} とする。

$$\bar{V} = \frac{1}{8} \sum_{i=1}^8 V_i \quad (1)$$

次に、等級ごとの参照ベクトルと対象メロンの平均特徴ベクトルとの類似度をコサイン距離により求め、最も距離が近い参照ベクトルが示す等級をそのメロンの等級 y とする。

$$y = \underset{k}{\operatorname{argmin}} \left(1 - \frac{REF_k \cdot \bar{V}}{\|REF_k\| \|\bar{V}\|} \right) \quad (2)$$

もう1つの方法は、ある1つの参照ベクトルを用いて、1つの距離尺度で等級を判定する方法である。例えば、A等級の参照ベクトルを基準にコサイン距離 d を求め、等級間のしきい値 $T_k(k=1, \dots, K-1)$ をもって等級 y を決定する。分類する等級数 $K=3$ の場合、以下となる。

$$y = \begin{cases} 1, & d < T_1 \\ 2, & T_1 \leq d < T_2 \\ 3, & d \geq T_2 \end{cases} \quad (3)$$

$$d = 1 - \frac{REF_1 \cdot \bar{V}}{\|REF_1\| \|\bar{V}\|} \quad (4)$$

4. 基礎実験

本システムの有用性を検証するため、3つの基礎実験を行った。まず、パッチ合成画像による等級判定の有効性を確認し、次に学習に用いるラベルありデータ数を変えての等級判定精度の評価を行った。最後に、画像の類似度の尺

度を用いた等級判定が、等級判定者の尺度と一致しているかについて検証を行った。

4.1 データセット

基礎実験で使用したデータは、3.2 節で述べた撮影部で撮影した A 等級 100 玉、B 等級 100 玉、C 等級 100 玉と、アノテーションのされていない 375 玉の映像データを使用した。データの取得は、2021 年 2 月から 2021 年 7 月にわたり 12 回実施され、等級のラベル付けは 3 名の等級判定熟練者によって実施された。

4.2 パッチ合成画像の有効性の確認

3.4 節で述べたパッチ合成画像を用いた等級判定の有効性を確認するための実験を行った。入力画像にパッチ合成画像を用いる場合と、パッチ化されていない元フレーム画像を用いる場合とで、テストデータに対する正答率の比較を行った。なお、元フレーム画像には左右反転、明度変更といった画像データ拡張を適用した。

まず、ラベル付きデータから各等級 50 玉分をテストデータとして抽出し、それ以外の各等級 30 玉分とラベルなしデータ 375 玉分を学習用データとする。等級判定モデルは、画像特徴量の抽出に EfficientNet-B0 を用い、分類クラス数は $K=3$ とする。学習には、損失関数にクロスエントロピー、最適化関数に Adam[25]を用い学習率は 0.0001 とした。3.6 節の学習プロセスに従って 5 ステップ (=500epoch) の学習を行い、ステップ毎に 3.7 節式(2)の方法でテストデータに対する等級判定で正答率を確認した。これを、交差検証法を用いて 3 回行い平均正答率をもって評価を行った。

パッチ合成画像を用いた場合と非合成画像を用いた場合の正答率を図 6 に示す。学習用データに対する損失(実線)を見ると、どちらも 0 付近に収束しており十分学習が進んだことが確認できる。また、パッチ合成画像と非合成画像を比較すると、非合成画像の方が損失の収束速度が速いことが確認できる。テストデータに対する正答率(破線)を見ると、2 ステップ (=200epoch) までは、パッチ合成画像の正答率は約 60%であり、非合成画像よりも平均 5.7%高かった。

4.3 ラベルありデータ数による正答率の比較

半教師あり学習に用いるラベルありデータ数と正答率との関係を調べるための実験を行った。ラベルありデータから等級ごとに 50 玉分をテスト用データとして抽出し、それ以外の $N = \{10, 30, 50\}$ 玉分を学習用データとする。それ以外は 4.2 節の実験と同じ条件で行い、ステップ毎の正答率を確認した。

ラベルありデータ数別の正答率を図 7 に、等級ごとの再現率と適合率を表 1 に示す。図 7 より、最も高い正答率を示したのは $N = 30$ で 60.0%であった、次に $N = 50$ で 58.0%、

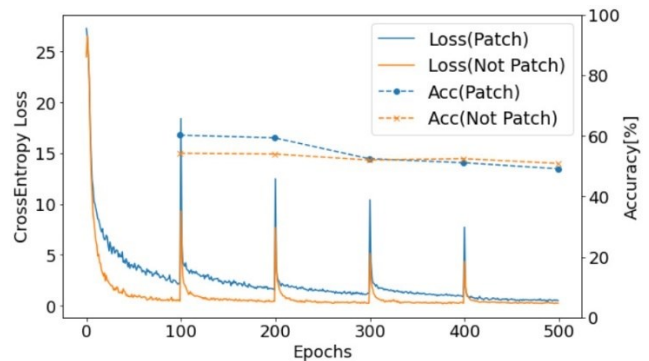


図 6 パッチ合成画像と非合成画像の正答率

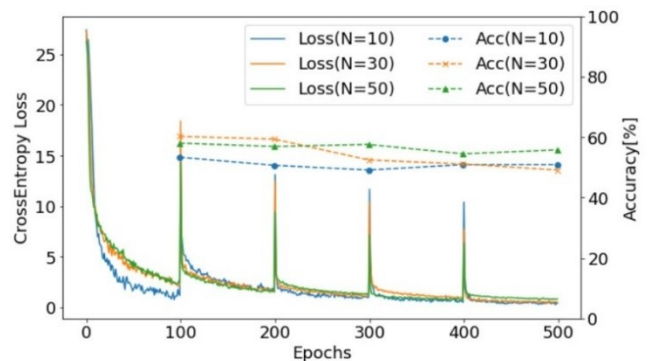


図 7 ラベルありデータ数別の正答率

表 1 等級別の再現率と適合率

	N=10		N=30		N=50	
	再現率	適合率	再現率	適合率	再現率	適合率
A等級	0.73	0.50	0.58	0.59	0.67	0.54
B等級	0.34	0.61	0.60	0.68	0.64	0.57
C等級	0.53	0.55	0.63	0.55	0.43	0.66

$N = 10$ で 53.3%であった。また、 $N = 30$ では他と比較し、学習ステップが進むにつれて過学習の傾向が見られた。また、表 1 より、 $N = 10$ では A 等級に判定の偏る傾向が見取れるが、 $N = 30, 50$ では著しい偏りは見取れなかった。

4.4 類似度による等級判定の有用性評価

画像の類似度を指標とした等級判定が、実際の等級判定者の行う判定と一致しているかを調べるための実験を行った。3.1 節で述べた通り、A 等級から C 等級になるにつれ形状や網目の歪みが増えていく。これは、類似度の観点から言えば、等級が下がるにつれて A 等級との類似度が低下する (= コサイン距離が広がる) と考えることができる。

そこで、まずテストデータを対象に、等級判定モデルで出力した特徴ベクトルと A 等級の参照ベクトル REF_1 とのコサイン距離を求め、しきい値を用いて A 等級と B|C 等級、A|B 等級と C 等級の 2 クラス分類を行った際の ROC 曲線を求め、3.7 節式(3)の方法で等級判定が可能か評価を行

った。

次に、ラベルなしデータに対し等級判定を行い、判定結果の等級ごとに10玉のデータを抽出し、それらの特徴ベクトルとA等級の参照ベクトル REF_1 とのコサイン距離を求めた。そして、抽出した30玉のデータに対し、等級判定者に品質の良し悪しをA等級の中でも品質が良いものを1、それ以外を2、B等級を3、C等級の中でも品質が良いものを4、品質が悪いものを5として5段階で評価してもらい、等級判定モデルで出力した距離と等級判定者の品質の指標の比較評価を行った。

図8にコサイン距離による等級判定のROC曲線を示す。コサイン距離による画像の類似度を基にA等級とB/C等級を分類した場合のAUCは0.71であった。同じく、A/B等級とC等級のAUCは0.72であった。

等級判定者が行った評価と対象となったメロンのコサイン距離を図9に示す。等級判定者が品質：良(2)と答えたメロンの平均コサイン距離は0.0025であった。逆に、品質：最低(5)と答えたメロンの平均コサイン距離は0.35であった。また、等級判定者の評価が低下するにつれ等級判定モデルが出力した特徴ベクトルもA等級の参照ベクトルから離れていくことが確認できた。なお、今回の実験では等級判定者が最も品質が良い1と判断したメロンはなかった。次に、コサイン距離と等級判定者の判定が大きく異なった例を図10に示す。局所的な傷や網目の乱れがある場合に見逃してしまう場合や、局所的な果実の盛り上がりなど人間であれば画像から3次元形状を予測して判断できるものを見逃してしまう傾向が確認できた。

5. 考察

パッチ合成画像によるデータ拡張と半教師あり学習の組み合わせにより、アノテーションの労力を抑えつつ判定精度向上を目指した。図6に示されるようなパッチ合成画像を用いたデータ拡張手法を適用することで、過学習を抑えて約6%の精度向上の効果を確認できた。しかし、半教師あり学習の効果については、図7で示す通り擬似ラベルを用いた学習を進めても、汎化性能の向上を確認できたとは言い難い。半教師あり学習の効果が表れなかった要因として、メロンの画像は、等級の違いによる画像の差異が小さく、画像の特徴空間上で密な分布になっていることが考えられる。また、ラベルありデータ数を30から50に増やしても、テストデータに対する正答率が増加しなかったことから、学習に用いたラベルありデータのラベルの一貫性が低いことも要因として考えられる。半教師あり学習が効果を発揮する前提として、分類クラス間の分布が疎である必要[15]があり。そのためにも、画像上で等級間の違いがより明確になるようデータ取得方法の見直しや、アノテーションを行う等級判定者を一人に絞ってラベルの一貫性を高めるな

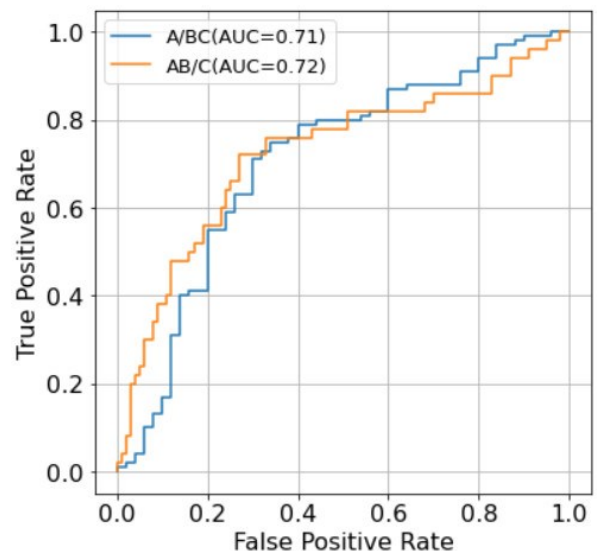


図8 コサイン距離による等級判定のROC曲線

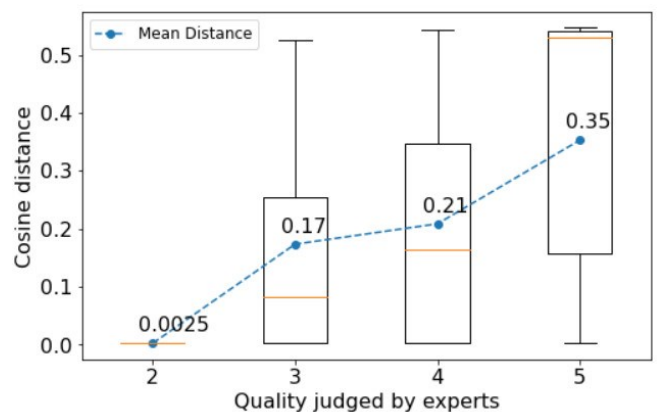


図9 等級判定者による評価別のコサイン距離

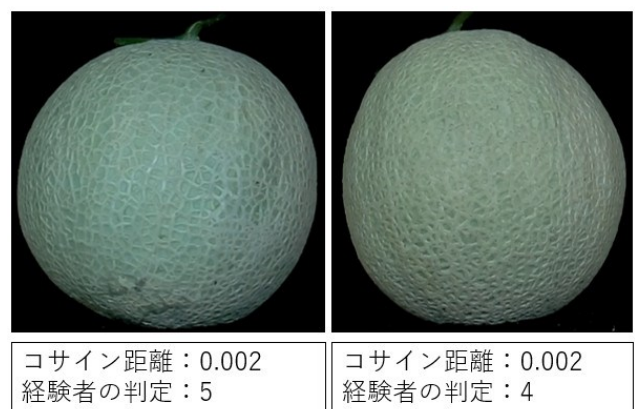


図10 コサイン距離と判定で乖離がある例

ど、データセットの質の向上が重要と考える。

メロンの外見の類似性に基づく等級判定の有用性に関して、図8,9で示す通り、等級が低くなるにつれA等級との画像特徴量の類似性が低くなるという直感的な仮説に則した結果を得ることができた。つまり、人間が行う等級判定

の再現に、類似度を用いることの有用性を確認できた。また、従来の CNN を用いたクラス分類手法では、あらかじめ定まったクラス数のみの判定しかできなかったが、類似度を用いた等級判定を適用することで、出荷先ごとに分類する等級数を分けることも容易となる。さらに、どの距離までを A 等級として判定するかといった、しきい値による等級判定の厳しさを調整できるようになるため、季節や出荷先、生産状況に合わせてより柔軟な運用のできる等級判定システムを実現できると考える。

6. まとめ

本研究では、メロンの外見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の基礎的な評価を行った。深層距離学習の一手法である ArcFace を用いて埋め込んだ特徴ベクトル空間上でのコサイン距離を等級判定に用いることで、正答率 60% の等級判定システムが実現可能であることを示した。また、本システムにより算出される A 等級を基準とした距離尺度と、等級判定者が主観で行う品質の判定に相関があること確認し、類似度を用いた等級判断の有用性を示した。

今後、さらなる判定精度向上のため、熟練等級判定者の知見を取り入れつつ、より質の高いデータを取得するための撮影部の改良が挙げられる。また、季節変動や収穫年度ごとの果実の成長のばらつきに対する評価も期待されている。さらに、見た目が重要視される他の果実への展開も考慮して本システムの改良を進めていく。

謝辞 本研究の一部は、JST 創発的研究支援事業 (JPMJFR201B) 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) 大和コンピューターの皆様に深い感謝の意を表す。

参考文献

- [1] 品目別 (季節区別) 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、
<<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」:
<http://www.pref.shizuoka.jp/j-no1/m_melon.html> (参照 2021-07-23).
- [3] 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.
- [4] 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.
- [5] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet Classification with deep convolutional neural networks. In Advances in neural information processing systems, pp.1097-1105, 2012.
- [6] 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.
- [7] Fujita, E., Kawasaki, Y., Uga, H., Kagiwada, S., 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.
- [8] Hughes, D., Salathe, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. <https://plantvillage.psu.edu/>, 2015.
- [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] Liang, W., Zhang, H., Zhang, G., Cao, H.: Rice blast disease recognition using a deep convolutional neural network. Scientific Reports, Vol.9, Art.no.2869, 2019.
- [12] Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.4690-4699, 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., Sastry, P.S.: Robust Loss Function under Label Noise for Deep Neural Networks. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp.1919-1925, 2017.
- [15] Engelen, J., Hoos, H.: A survey on semi-supervised learning. Mach. Learn., Vol.109 No.2, pp.373-440, 2020.
- [16] Kaya, M., Bilge, H.: Deep Metric Learning: A Survey. Symmetry, Vol.11 No.9:1066, 2019.
- [17] Liu, W., Wen, Y., et al.: Sphreface: Deep hypersphere embedding for face recognition. In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.6738-6746, 2017.
- [18] Wang, H., Wang, Y., et al.: Cosface: Large margin cosine loss for deep face recognition. In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.5265-5274, 2018.
- [19] Zhang, X., Zhao, R., et al.: AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations. In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.10823-10832, 2019.
- [20] Zhou, D., Bousquet, O., Lal, T.: Learning with Local and Global Consistency. In Advances in Neural Information Processing Systems, pp.1233-1240, 2004.
- [21] Takahashi, R., Matsubara, T., Uehara, K.: Ri-cap: Random image cropping and patching data augmentation for deep cnns. In Asian Conference on Machine Learning, pp.786-798, 2013.
- [22] Yun, S., Han, D., et al.: CutMix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.6023-6032, 2019.
- [23] OpenCV, Ver.3.8.8, <<https://opencv.org/>> (参照 2021-7-23)
- [24] Tan, M., Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In International Conference on Machine Learning (ICML), 2019.
- [25] Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv:1412.6980.2014.