

## 深層学習による画像認識を用いたトマトの成熟度分類 Estimation of tomato ripeness using deep learning image recognition

井上 多雲<sup>†</sup>      奥村 俊昭<sup>‡</sup>      矢島 邦昭<sup>‡</sup>  
Takumo Inoue      Toshiaki Okumura      Kuniaki Yajima

### 1. 研究背景と目的

著者らは収穫作業支援を目的として携帯可能なトマト収穫時期提示システムを研究開発している。トマトは成熟過程に応じて緑色から赤色まで 10 段階の着色度合いが農協で定められており、市場に出回る品質を安定させるために、その日に出荷する着色度合いを農協から指示され、それに合わせて農園で収穫している。経験不足の従業員は収穫のタイミングを逃すあるいは早すぎるということが多く、最適な収穫時期を見分けるノウハウを的確に教えることができるシステムが望まれている。

本研究では、撮影されたトマト画像を、CNN(畳み込みニューラルネットワーク)[1]によって学習することで、トマトの着色度を判断するシステムの開発を行っている。

### 2. 学習用データセットの作成

分類対象のトマトは収穫される前の状態であるため、トマトの向きや形状、萼の有無や撮影環境など、様々な条件を考慮する必要がある。そのため、本研究では農園や研究室内で撮影された、条件が一定ではないトマト画像を使用している。それらの画像を図 1 に示す宮城県産トマト標準出荷規格表を参考に、表 1 のように 5 段階で分類した。なお、用いる画像は背景に映り込んでいる枝葉や農園設備の影響を受けないようあらかじめマスク処理を行っている。総データ数は 3213 枚であり、そのうちの 2450 枚が学習用データ、残りの 763 枚が検証用データとして使用される。



図 1 宮城県産トマト出荷規格表(JA 全農みやぎ)

表 1 クラス分類とデータ数

標準出荷規格表	教師ラベル	データ数
① ②	Green	470
③ ④	Light Green	555
⑤ ⑥	Pink	1004
⑦ ⑧	Light Red	992
⑨ ⑩	Red	192
計		3213

<sup>†</sup> 仙台高等専門学校 情報電子システム工学専攻  
National Institute of Technology, Sendai College  
Division of Information and Electronic System Engineering

<sup>‡</sup> 仙台高等専門学校 総合工学科  
National Institute of Technology, Sendai College  
Department of Comprehensive Engineering

### 3. 学習用ネットワークの構築

表 1 のデータセットを用いて、CNN による学習を行った。構築したネットワーク図を図 2 に示す。畳み込み回数が増えると、学習時間が増加するとともに、学習が不安定になってしまう。そこで、学習時間短縮と安定化を図ることを目的に Batch Normalization[2]を行った。Batch Normalization はミニバッチごとの統計量を用いて、簡易的に正規化を行う手法である。これにより、パラメータのスケールや初期値の影響が少なくなり学習が安定し、高い学習率を設定することで、学習時間を短縮できる。

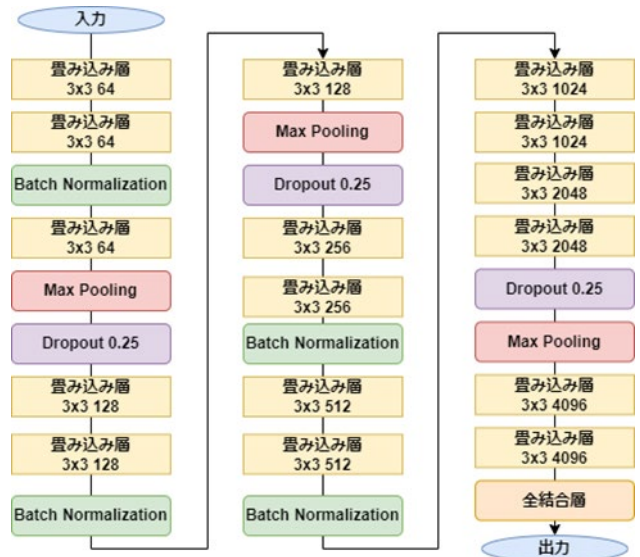


図 2 構築した学習用ネットワーク

### 4. 実験結果

表 1 のデータセットと、図 2 のネットワークを用いて、学習を行った。Epoch 数は 500、バッチ数は 32 に設定し、最適化アルゴリズムには Adadelta[3]を使用した。学習用 PC には Intel (R) Core (TM) i9-10900X CPU @ 3.70GHz が搭載されており、GPU には NVIDIA GeForce RTX 2080 Ti を 2 基使用している。学習にかかった時間は各 Epoch 2 秒、総学習時間約 1000 秒であった。学習結果における正解率の推移のグラフを図 3 に、損失の推移のグラフを図 4 にそれぞれ示す。学習率(Train accuracy)は 87%、正解率(Validation accuracy)は 83%という結果となった。検証用データを用いた際の損失は Epoch100 を超えたあたりから横ばいになっており、過学習は起きていないが、Epoch100 以降の学習が進んでいないことが確認できた。

学習結果の混同行列(Confusion matrix)を表 2 に示す。Green, Red のカテゴリでは正解率が 90%を超えているが、LightGreen, Pink のカテゴリでは正解率が 80%を下回っている。特に分類精度が低い LightGreen カテゴリの正解率は 70.82%であった。

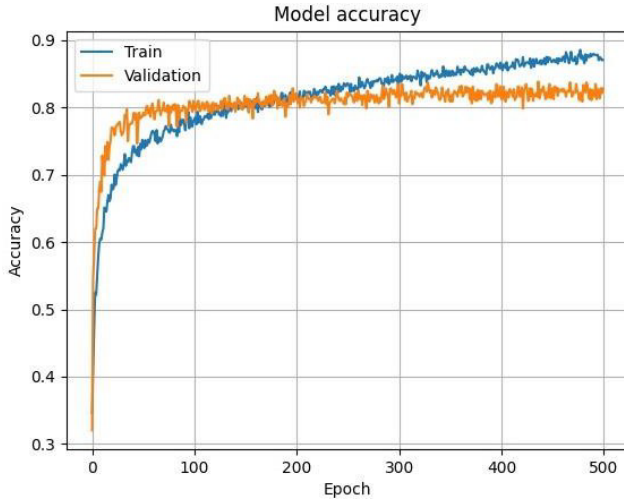


図 2 正解率のグラフ

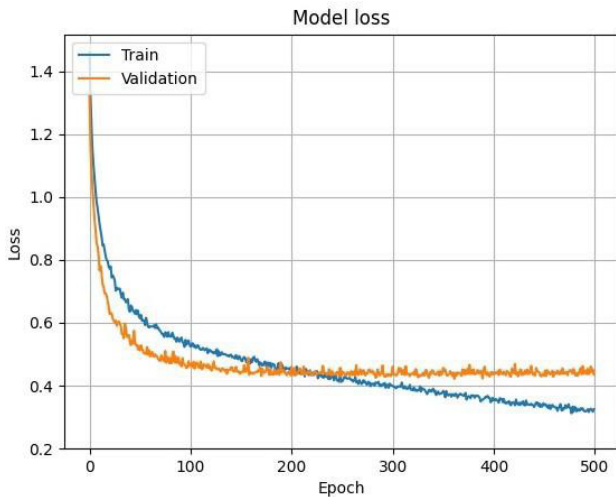


図 4 損失のグラフ

表 2 学習結果の混同グラフ

教師ラベル	Green	114	12	0	0	0
	Light Green	9	102	31	2	0
	Pink	1	20	171	25	0
	Light Red	0	0	33	241	5
	Red	0	0	0	4	44
		Green	Light Green	Pink	Light Red	Red
分類されたカテゴリ						

## 5. 考察

実際に本システムを収穫現場で使用することを想定した場合、正解率は低くとも 90%以上は必要であり、現段階では精度が不十分であると考えられる。また、農園で収穫してから消費者に届くまでに時間がかかるため、収穫現場では図 1 の規格表における着色度 3~6 のトマトを収穫することが多い。これらのトマトは、特に分類精度が低い Light Green や Pink のカテゴリに含まれており、これらのカテゴリの分類精度向上が主な課題である。



図 5 誤分類画像の例

図 5 に示す誤分類した画像を観察すると、蒂が映っている画像が多くみられた。図 1 にあるように、着色度はトマトの下部を見て判断される。CNN が蒂を特徴量として認識してしまい、着色度に着目した分類がうまく行っていないと考えられる。また、影が重なって明度が落ちている画像にも誤分類が見られた。本システムは撮影条件が一定ではないことを想定しているため、Data Augmentation[4]で明度を変化させた画像を生成し、汎化性能を高めることが必要となる。

学習用ネットワークは徐々に畳み込み回数を増やしながらか実験を行っていったが、これ以上の畳み込み回数の増加は正解率に影響しなかった。分類精度向上には畳み込み回数やプーリング層の位置だけではなく、ネットワークの構造を改善することが必要だと考える。ResNet[5]などで用いられている、Shortcut Connection(恒等写像)と残差ブロックによって、さらに層を深くしたネットワークでの学習が可能になる。また、Shortcut Connection には浅い層で抽出した特徴量を深い層で用いることができるという特徴がある。これらの層を用いて、ネットワークを再構築、もしくは転移学習をすることによって精度が向上すると考える。収穫現場では、10 段階の着色度合いのうち、指定された 1 つの着色度を収穫する。そのため、最終的に 10 段階にラベル付けたデータセットを作成し、学習を行う必要がある。

## 6. おわりに

本研究では、CNN により着色度を自動判別する分類器を作成し、84%の正解率を得た。5 段階の分類は本システムにおける前段階であり、実際の収穫現場で使用するには 10 段階の分類システムを開発する必要がある。そのため、データセットやネットワークの改善を行い、分類精度の向上を図る。5 段階での分類精度 90%以上を目標にし、10 段階での分類システム開発が今後の展望として挙げられる。

### 謝辞

本研究は JSPS 科研費 JP20K06330 の助成を受けたものである。トマト画像撮影には(株)サンフレッシュ小泉農園と(有)地域資源活用研究所にご協力頂いた。

### 参考文献

- [1] A. Krizhevsky, I. Sutskever, and G. Hinton. "ImageNet classification with deep convolutional neural networks.", Advances in Neural Information Processing Systems 25 (2012).
- [2] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International conference on machine learning. PMLR, (2015).
- [3] Zeiler, Matthew. D. "Adadelat: an adaptive learning rate method." arXiv preprint arXiv:1212.5701, (2012)
- [4] Xiaodong Cui, et. al. : "Data augmentation for deep neural network" IEEE/ACM Trans. Audio, Speech, and Lang. Proc, 23(9):1469–1477.(2022)
- [5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. (2015).