

曇み込みニューラルネットワークを用いたトマトの向き推定

Tomato Orientation Estimation using Convolutional Neural Network

堀野 泰輔[†]奥村 俊昭[‡]矢島 邦昭[‡]

Taisuke Horino Toshiaki Okumura Kuniaki Yajima

1. 背景・目的

近年、日本では急速な少子高齢化が進む中、人口は減少に転じており、近い将来における労働力不足が懸念されている。特に農業の分野では、他分野よりもさらに急激な高齢化と人手不足が進み、すでに深刻な状況となりつつある。そのため農林水産省では、ロボット技術や人工知能等、革新的な技術を農業現場に導入することで、安定かつ高品質生産を実現する新たな農業を確立させることを目指す方向性が示されている。そこで、農業の出荷工程の 1 つである箱詰め作業に着目した。箱詰めには各農産物において様々なルールが存在し、例えばトマトでは箱詰めの際には必ず蒂部分を下にして出荷する必要がある。この作業をロボットに行わせることを想定すると、ロボットは掴んだトマトの向きを正確に判断し、蒂部分が下になるように調整をする必要がある。これに対し著者は、正確なトマトの向きの判断を、深層学習によって作成される分類器によって行わせることができると考えた。本研究では Convolutional Neural Network (CNN) [1] を構築し、角度別に撮影したトマトの画像を学習させ、トマトの向き分類器を作成する。そして分類器の学習率、検証用データの正解率を検証する。

2. データセット

本研究で用いるデータセットは、X 軸、Y 軸、Z 軸を 45 度刻みで向きを変化させたトマトを第 1 象限のみ撮影し、第 2、第 3、第 4 象限の向きのデータセットは撮影した画像を 90 度ずつ回転させたものをデータとしている (図 1)。したがって 26 方向分のデータセットを学習し、それぞれの方向ごとの確率で分類を行う。また、箱詰め作業を行う環境を想定すると、認識したトマトの背景は一律であると考えられるため、データとして扱う画像の背景も統一している。各方向のデータ数は 100 枚の一律としており、訓練データと検証用データの比率は 8:2 としている。よって画像データ計 2600 枚のデータセットとなり、このデータセットを用いて分類の学習を進める。



(a) X=0, Y=90, Z=0 (b) X=45, Y=45, Z=45

図 1 使用した角度別のトマト画像

[†] 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

[‡] 仙台高等専門学校 総合工学科
National Institute of Technology, Sendai College
Department of Comprehensive Engineering

3. CNN 構築

CNN は空間フィルタ畳み込み演算層や Pooling 層を何層も重ねて構築されるニューラルネットワークであり、主に画像などの 2 次元データの処理に適している。本研究では、この CNN の構築を TensorFlow 上のニューラルネットワークライブラリである Keras を用いて行っている。構築した CNN のネットワークモデルを図 2 に示す。

今回作成したデータセットにおいて、各ラベルにおけるデータの特徴の違いは、トマトの蒂部分の位置や、花落ち部分の位置、また蒂部分から花落ち部分にかけて伸びる薄い筋等だが、特に花落ち部分の認識と薄い筋の認識が重要であると考えられる。トマトの蒂部分が写らないデータセット間の分類において、ラベルごとにはっきりとした特徴の違いが存在しないため、誤分類が発生しやすい問題がある。そのため、花落ち部分等の微小な特徴を抽出させ、それらの有無や位置によって分類を行わせる必要がある。

CNN にデータセットを学習させる際、原画像サイズを小さくリサイズし、計算量を少なくすることで大量の画像データを短期間で学習させる前処理を組み込む。このリサイズによる縮小の度合いを小さくすることで、花落ち部分等微小な特徴を保持したまま学習させることができると考えられる [2][3]。今回は細かな特徴が認識でき、かつその状況で一番小さい入力画像サイズである 300×300 画素に設定した。さらに、畳み込み演算層では入力画像の細かな特徴をより正確に検出させるようカーネルサイズを 3×3 で処理させている。一方 Pooling 層は、入力データに対し情報圧縮処理を施しサイズを縮小する層となっているため、データごとの微小な位置変化等に対応させることなどに期待されるが、ダウンサンプリングすることにより特徴がぼやけてしまうため、微小な特徴を抽出することを目的とする今回の学習には向かない処理層となっていると考えられる [4]。そのため Pooling 層の数は畳み込み演算層と比べ、比較的少なく設置させている。

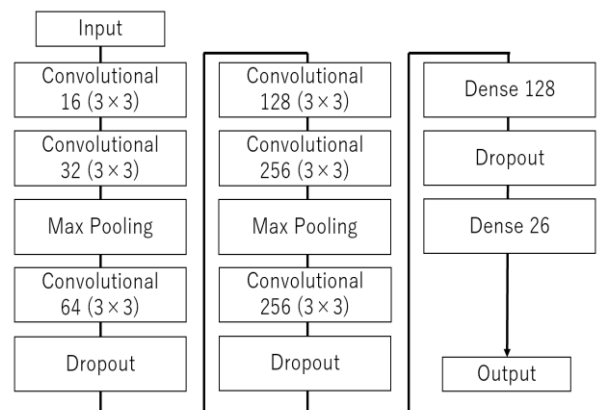


図 2 構築したネットワークモデル

4. 過学習抑制

本研究の学習において、過学習の対策は非常に重要な項目となっている。トマトの蒂部分が写らない角度のラベル同士の分類を正確に行わせるため、入力画像中の微小な特徴を抽出する CNN を構築しているが、細かな特徴を学習すると同時に過学習の発生リスクが高まる問題点がある。過学習は訓練データ 1 つ 1 つをより忠実に学習しすぎてしまい、検証用データやテストデータ等の未知データに対して対応できず正解率が低くなる現象である。今回構築した CNN では Dropout 処理と、畳み込み演算層に対して L2 ノルムを設定し重みにペナルティを設けることによって過学習の抑制を行っている [5]。Dropout 処理は任意の確率でランダムにニューロンを無視して学習を進める正則化処理であり、学習の反映率を下げる効果がある。この Dropout 処理をネットワーク内に 3 つ設置し、30%の確率でニューロンを透過するよう設定した。L2 ノルムではニューロンの重みパラメータにペナルティを設け、重み値が大きくなりすぎないようにする正則化処理である。これら 2 つの正則化により、微小な特徴を抽出させつつ過学習を抑制することができると考えられる。

5. 学習率と検証用データの正解率

構築した CNN を用いてデータセットを学習させ、訓練データの学習率と検証用データの正解率を検証した。Epoch 数（データセットの学習回数）は 150 回に設定し、最適化アルゴリズムを Adam[6] とし、学習率を 0.001 としている。Epoch ごとの訓練データの学習率と検証用データの正解率を表すグラフを図 3、訓練データと検証用データの各損失を表すグラフを図 4 に示す。訓練データの最終学習率と検証用データの正解率は、どちらも 85% となった。また訓練データの損失は 0.25、検証用データの損失は 0.4 と、値の差が生じているが、図 4 を見ると検証用データにおける損失の値は上昇しておらず、0.4 で一定となっているため、過学習は抑制できていると考えられる。

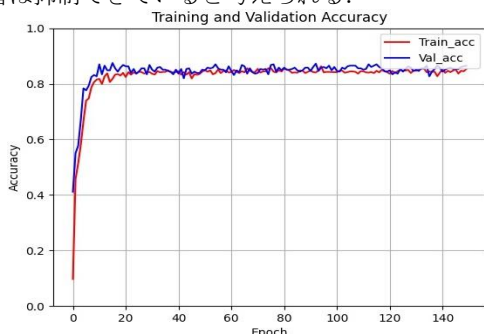


図 3 訓練データの学習率と検証用データの正解率

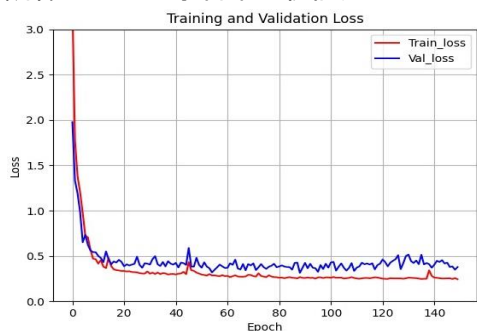


図 4 訓練データと検証用データの損失

6. 考察

訓練データの学習率と検証用データの正解率の 85% という結果は、深層学習による分類において十分であると結論しがい数值である。本研究の背景に存在するトマトの箱詰め作業では、1 回の作業に数百から数千個のトマトを扱っている。そのため 15%の確率で誤分類を行ってしまうと、数として膨大な量のトマトが蒂部分を下に向けられていない状態で納品されてしまうことになる。したがってさらなる学習率と正解率の向上に向け、データセットや CNN の改善が必要であると考えられる。

現在データセットの総数が 2600 枚となっており、各方向におけるデータ数は 100 枚となっているが、データセットの拡張により訓練データの学習率、検証用データの正解率の向上が期待できる。また本研究ではトマトの色の度合いが赤のもので統一しているが、着色度合いが異なるものもデータセットに組み込むことで、よりさまざまな状況下で分類を行わせることができるようになる。かつラベル内の画像間に特徴の違いが生じるため、これまで以上に過学習の発生リスクを抑えながら学習を行わせることができると考えられる。そして現在のトマトの撮影状況は周囲の明かりが一定となっているが、現地で分類器を使用する際は周囲の環境によって色の写り方が変化すると予測される。そのため現在のデータセットの輝度値を変化させたものも組み込むことで、さらに実用性が増すと考えられる。

7. おわりに

本研究では深層学習によるトマトの方向判別を行う分類器を作成し、過学習を抑制しつつ 85%の訓練データの学習率と、85%の検証用データの正解率を得た。しかし実用化に至るには、さらなる学習率と正解率の上昇が必要であると考えられる。学習率、正解率の向上の手法として、データセットの拡張が挙げられる。データ数を増加させることに加え、トマトの着色度合いが異なるものを追加することや画像の輝度値を変化させたものを追加することによって、分類器を周囲の状況に柔軟に対応させることができ、汎用性・実用性が高まると考察される。そしてデータセットの拡張・調整を行った上での CNN の最適ネットワークの構成を追求することや、特徴抽出過程の可視化を行い推論の正確性を確認することなど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す。

参考文献

- [1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2022).
- [2] Babenko, Artem and Lempitsky, Victor. Aggregating deep convolutional features for image retrieval. In *ICCV* (2022).
- [3] 玉木 徹, 画像識別のための特徴量, *Medical Imaging Technology*, 2017, 35 巻, 1 号, p. 23- 28 (2022).
- [4] Cui, Yin, et al. "Kernel pooling for convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition* (2022).
- [5] Li, Hao, et al. "Pruning filters for efficient convnets." *arXiv preprint arXiv:1608.08710* (2022).
- [6] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), *Conference Track Proceedings, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9* (2022).