

複数ラベルを用いた機械学習による画像認識の研究

Research on image recognition by machine learning using multiple labels

茨城大 ○澤田篤彦 ◎尾島裕隆, 周立波, 清水 淳, 小貫哲平, 金子和暉

要旨

現在、機械学習を用いた農業における野菜出荷時の良否判別を考えており、この判別を行うには、野菜の種類と良否という複数ラベルを持った物体検出が必要となる。そこで、今回は、その第一段階として、機械学習で使われる VGG16 を応用し、複数ラベルの画像認識のモデルを考案した。この作成したモデルに複数のラベルを持った手書き数字を用いて画像認識を行った。

1. 研究背景

現在の日本の農業従事者の平均年齢は、図 1(a)を見るとわかるように平成 31 年時点で約 68 歳、そして平均年齢及び高齢化率がともに伸びていることが分かる。今まで、農業従事者が、図 1(b)のように野菜出荷時に手作業で行っていた仕分け作業に深層学習を用いることで AI が人間の眼の役割を担い、農業従事者への負担を少しでも減らせると思い本研究に至った。

2. 深層学習による物体検出と良否判別

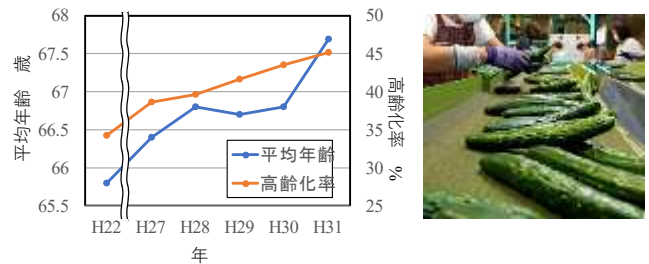
画像認識とは、何らかの画像からその画像または、画像の中にある単一の物体が何であるかを識別する手法である。つまり、図 2(a)のような画像中の一本のキュウリを識別することができる。使われるモデルの代表例として、VGG16¹⁾などがある。次に、物体検出とは、ある画像の中の単一の物体だけでなく、複数の物体を同時に識別し、かつ物体の位置までも検出できる手法である。使われるモデルの代表例として、SSD がある。物体検出は、図 2(b)のように、画像中の二本のキュウリを識別することができ、かつ赤い枠線でそれぞれのキュウリの位置も検出できる。そして、私が考案するモデルは、この物体検出を用いた良否判別であり、リアルタイムでの検出が可能である SSD に良否判別のアルゴリズムを付け加えたものである。図 2(b)のように複数のキュウリの識別と位置の検出に加えて、図 2(c)のような曲がったキュウリを出荷不可として判定できる。

私が提案するモデルは、図 2 のようにキュウリを検出する場合には、対象物の種類（キュウリ）を示すラベルに加え、対象物の出荷可否のラベルを与えることで、出荷可能なキュウリを識別する。今回は、その第一段階として、複数のラベルを同時に与え、VGG16 を応用した物体識別により、その有用性を検証した。

3. 複数ラベルをもつ学習

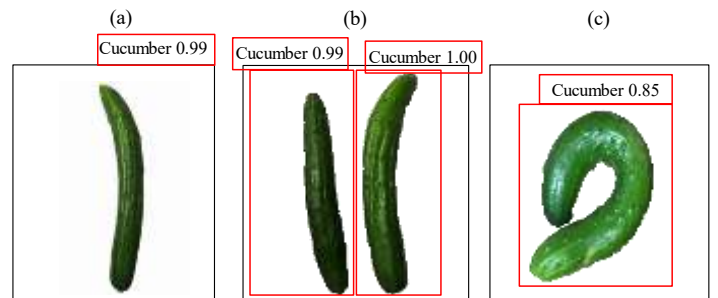
本研究では、0 から 9 の手書き数字の画像データセット MNIST を 0° から 179° でランダムに回転させて、数字と角度の二つのラベルを与えて、画像認識を行った。モデルは、画像認識でよく使われる VGG16 を選んだ。VGG16 とは、CNN の代表的なモデルであり、図 3 のように畳み込み層 13 層＋全結合層 3 層の計 16 層で構

成されており、この 16 層が名前の由来となっている。今回ラベルを数字と角度で二つ与えているため、モデルの出力も 2 つに分ける必要がある。そこで、図 3 に示すモデル右端のように全結合層の最後の 2 層を数字用と角度用で分けた。



(a) 平均年齢と高齢化 (b) 出荷判別作業例

図 1 農業従事者の現状



(a) 1本のキュウリ (b) 複数のキュウリ (c) 出荷不可キュウリ

図 2 それぞれのモデルに対するキュウリの検出結果

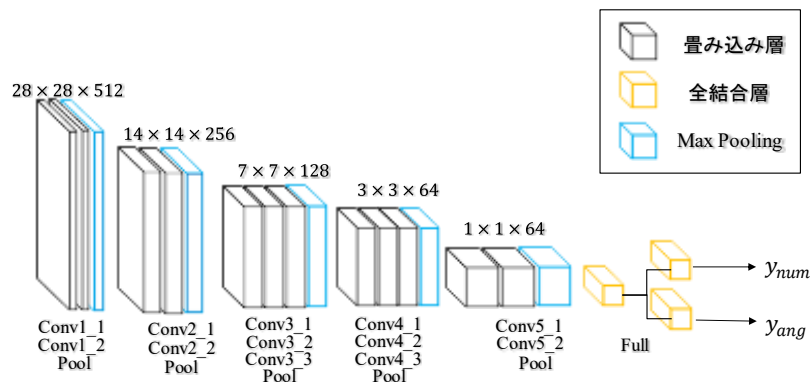


図 8 VGG16 の構造

4. 学習条件

学習に使用するデータセットについて MNIST は、訓練データ 6 万枚、検証データ 1 万枚の計 7 万枚の手書き数字画像である。本研究の角度の与え方については、 0° から 179° を 0° から 9° をクラス 1, 10° から 19° をクラス 2 というように 10° 刻みで、18 クラスに分けてラベル付けすることを考え、18 クラスをランダムに 20 個生成し、MNIST の画像を回転させた画像データを生成した。そのため、訓練データ 120 万枚、検証データ 20 万枚の計 140 万枚となった。詳細を表 1 に示す。画像に与えるラベルは、数字のラベルについては MNIST による 0 から 9 の 10 個のクラスを使用し、角度のラベルは、上で述べたように、計 18 個のクラスを用いた。さらに、学習時に用いた角度は整数のみとし、テスト時に用いた画像では実数の角度まで与えた。また、損失関数は数字、角度ともに交差エントロピーを用いた。交差エントロピーの式を式(1)に示す。 x は入力であり、 $p_1(x)$ 、 $p_2(x)$ はそれぞれ数字と角度のラベルであり、 $q_1(x)$ 、 $q_2(x)$ はそれぞれ数字と角度におけるモデルの推論結果である。

$$H(p_1, p_2, q_1, q_2) = -\sum_x p_1(x) \log(q_1(x)) + \left(-\sum_x p_2(x) \log(q_2(x)) \right) \quad (1)$$

5. 学習結果と考察

図 4 に学習と検証結果の正解率を示す。実線が学習結果、点線が検証結果を表している。また、青線が数字、赤線が角度の結果を示している。縦軸を正解率、横軸をエポック数としておりエポック数がわずか 5 であるが、数字の正解率は 98.6%、角度の正解率は 95.1% にまで達した。図 5 は、学習と検証結果のロスである。それぞれの線は、図 4 と同様である。縦軸をロス、横軸をエポック数としており、ロスもしっかり収束できていることが分かる。正解率、ロスともにエポック 1 を境に急激な変化が見られる。続いて図 6 は、テストデータに対する、数字と角度の正解率を示したものである。数字に関しては、どれも特にばらつきはなく精度も 97% を上回っており、全体の正解率は 98.7% であり、非常に良い結果が得られた。一方、角度はランダムに与えられているため、各クラスにおいて与えられる数にばらつきが生じる影響で、生起率に多少のばらつきが見られるが、どれも 86% を上回っている。加えて、全体の正解率は 91.9% であり、良い結果が得られた。 0° から 9° 正解率が非常に高くなっているが、数字本来の形に近いため精度が良いと考えられる。

6. 結言

今回は、複数ラベルを与えての画像認識を行うことに成功した。テスト画像では、実数まで与えたにも関わらず、数字の正解率が 98.7%、角度の正解率が 91.9% であった。また、キュウリの良否判別でもキュウリの曲がり具合などで判別するため角度の設定が必要になると感じるので、今回の結果をもとにアルゴリズムを構築していきたい。

表 1 学習条件

Params	value
Training : validation	1200000 : 200000
Batch Size	1000
Epochs	5
Activation Function	Softmax

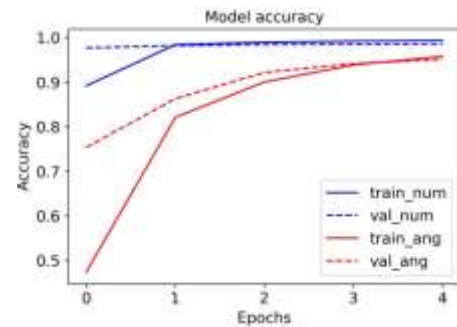


図 4 学習と検証結果の正解率

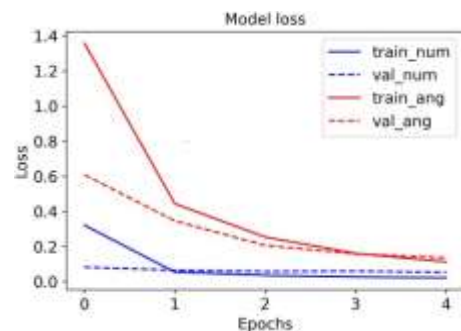
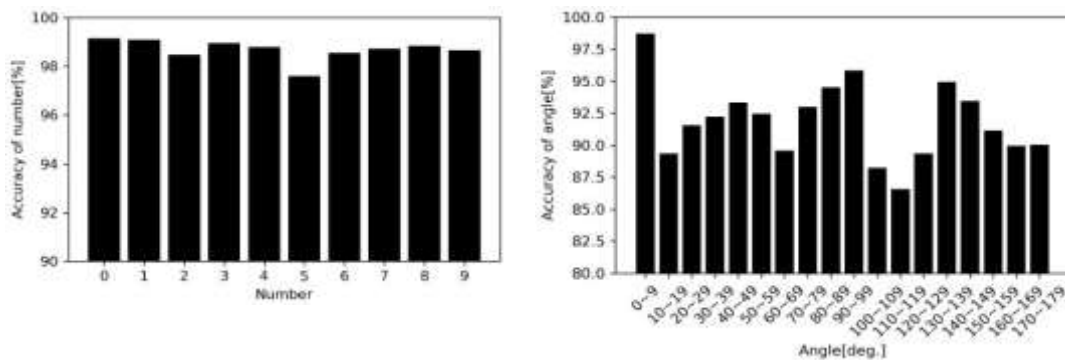


図 5 学習と検証結果のロス

- 1) Keren Simonyan, Andrew Zisserman : Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014



(a) 数字

(b) 角度

図 6 テスト結果における数字と角度のそれぞれの正解率