

# 深層学習を用いたクラウド型画像分類システムの開発

渡辺 博己      岩田 靖三\*      生駒 晃大      棚橋 英樹

## Development of Cloud-Based Image Classification System using Deep Learning

Hiroki WATANABE    Seizo IWATA\*    Akihiro IKOMA    Hideki TANAHASHI

あらまし 深層学習 (Deep Learning) が注目され、クラウドサービスにより手軽に利用できるデータ分析環境も整いつつあるが、中小企業での活用には至っていない。本研究では、深層学習、特に畳み込みニューラルネットワーク (CNN, Convolutional Neural Network) を用いた画像認識技術に対する理解の促進と中小企業への導入支援を目的として、専門的な知識がなくても、容易にCNNを用いた画像分類技術を利用することが可能なシステムを構築した。本稿では、システムで利用可能なCNNについて述べるとともに、画像分類実験の結果を報告する。また、開発システムの概要について紹介する。

キーワード 人工知能, 深層学習, 畳み込みニューラルネットワーク, 画像分類, クラウド

### 1. はじめに

近年、人工知能 (AI, Artificial Intelligence) 分野の一つの技術である深層学習 (Deep Learning) が注目を集めている<sup>[1]</sup>。深層学習は、ニューラルネットワークの一種で、既に、画像認識の分野では、畳み込みニューラルネットワーク (CNN, Convolutional Neural Network) を用いることがデファクトスタンダードとなり<sup>[2]</sup>、その高い認識精度から外観検査や自動運転等、産業界での活用が期待されている。

一方、インターネット等のネットワークを通じて、コンピュータに関する様々なサービスを利用できるクラウドコンピューティングが多くの企業で導入されている。深層学習の利用が可能なSaaS (Software as a Service) やPaaS (Platform as a Service) を提供する企業が増加し<sup>[3]</sup>、今後、深層学習を活用する企業も増加していくと予想される。

これらの新技術を中小企業が活用するにあたっては、2017年4月に公表された中小企業白書<sup>[4]</sup>の第2-3-42図 (新技術の活用における課題) によると、「技術・ノウハウを持った人材が不足している (45.1%)」、「自社の事業への活用イメージがわからない (38.5%)」、「新技術について理解していない (30.2%)」等の課題があることが示されている。また、同第2-3-43図 (新技術の活用における課題別に見た、期待する公的支援) によると、これらの課題に対して、約4割の中小企業が「専門家等による助言、技術的支援の提供」を求めているとされている。

そこで、本研究では、深層学習、特にCNNを用いた画像

認識技術についての理解促進と、中小企業への導入支援を目的として、専門的な知識がなくても、独自に用意したラベル付きの画像データから学習モデルが構築でき、画像分類結果を簡単に取得することが可能なシステムを構築した。具体的には、クライアント端末からサーバにWebブラウザでアクセスし、認識性能を確認したい画像データをクラス別にアップロードすることにより、クラウド上のコンピュータリソースを用いてCNNによる学習・分類を実行し、分類結果をクライアント端末で確認できるシステムと、これらを実行するWebアプリケーションを開発した。以下、システムで利用可能なCNNの構造、認識精度について報告するとともに、開発したWebアプリケーションについて紹介する。

### 2. 畳み込みニューラルネットワーク<sup>[1]</sup>

畳み込みニューラルネットワーク (CNN) は、図1に示すとおり、入力層、畳み込み層、プーリング層、全結合層、及び出力層から構成されるのが一般的であり、畳み込み層とプーリング層を交互に接続した構造を持つことに特徴がある。

畳み込み層は、重みフィルタに応じた特徴抽出を行う層である。図1の例では、 $3 \times 3$ 画素分の領域を持つフィルタを1画素ずつ移動させながら、入力に対する重みフィルタの内積を計算し、チャンネルと呼ばれる単位の出力画像を新たに生成する。この時、 $N$ チャンネルから構成される畳み込み層は、 $N$ 個の重みフィルタを用いて生成された出力画像を持つ層となる。なお、畳み込み層で生成される出力画像

\* 岐阜県セラミックス研究所



図1 畳み込みニューラルネットワークの構成

のサイズは、元の画像よりフィルタサイズに応じて小さくなるが、出力画像の周囲にパディングと呼ばれる処理を行うことにより、同一サイズとすることが可能である。

プーリング層は、畳み込み層で生成された出力画像を縮小する層である。図1の例では、 $2 \times 2$ 画素分の注目領域に対して、領域が重ならないよう移動させながら、最大値、または平均値等を計算し、新たな出力画像を生成する。この時、出力画像のサイズは、元の画像サイズの縦横1/2倍となる。なお、チャンネルについては、チャンネル毎に独立して処理を行うのが一般的であるため、畳み込み層の出力チャンネル数とプーリング層の出力チャンネル数は一致する。

CNNの認識精度は、年々向上しているが、層が深くなる傾向にある。そのため、本研究で事前に用意したCNNについては、層の深さの比較ができるよう、単純な4層CNNと、AlexNet<sup>[5]</sup>を参考にした8層CNN、及びVGGNet<sup>[6]</sup>を参考にした12層CNNの3種類を構築した。図2に各CNNの構成図を示す。各CNNにおいては、入力層で $32 \times 32$ にリサイズ処理したグレースケール画像を生成した後、畳み込み層とプーリング層を数回繰返しながら特徴量を抽出することにより、全結合層で特徴量に基づき分類し、出力層で各クラスに分類される確率に変換した結果を生成する。なお、本研究においては、畳み込み層と全結合層の数をCNNの層数として数え、出力層も全結合層に含めている。

### 3. 画像分類実験

前章で構築した各CNNの画像分類性能を評価するため、

実験用に撮影した画像データを用いて、種類（クラス）別に分類する実験を行った。また、CNNとの比較のため、他手法による画像分類実験も行った。

#### 3. 1 実験概要

実験に用いる画像データは、陶器の裏面にある焼印を対象として、画像ファイルを連続して記録することができるAndroidアプリケーション（岐阜県セラミックス研究所が開発）により取得した。なお、分類対象とする焼印は5種類とした。

画像データの取得にあたっては、撮影対象を等速回転させ、JPEG形式によりファイルに記録した。撮影では、各種類200枚の画像データを1セットとして、照明や、カメラと対象との距離、及び角度を変更し、8000枚（各種類1600枚）の画像データを取得した。また、別途取得した画像データ500枚（各種類100枚）について、画像位置を水平方向に移動させたり、サイズを変更したりして、画像データを加工し、2000枚（各種類400枚）の画像データを用意した。

実験では、これらの10000枚の画像データについて、交互にサンプリングし、一方を学習用、他方を評価用として利用した。さらに、別途評価用として、撮影条件の異なる2000枚（各種類400枚）の画像データを取得し、実験を行った。実験では、先に取得した評価用の画像データをTest-A、別途取得した評価データをTest-Bとして、画像分類精度の差を評価した。実験に用いた画像データにおける各画像のラベル（クラス名）と学習データの例を図3に示す。

画像分類性能の評価にあたっては、前章で述べたCNNの構成に基づき、5クラスに分類する出力層を持つCNNを構築した。また、CNNの分類精度との比較のために、昨年度開発したシステム<sup>[7]</sup>で利用したBag-of-Features (BoF)により特徴量を抽出し、Support Vector Machine (SVM)により分類処理を行う方法<sup>[8]</sup> (BoF+SVM)についても分類精度を求めることにした。

各CNNでは、入力画像のサイズを $32 \times 32$ としていること

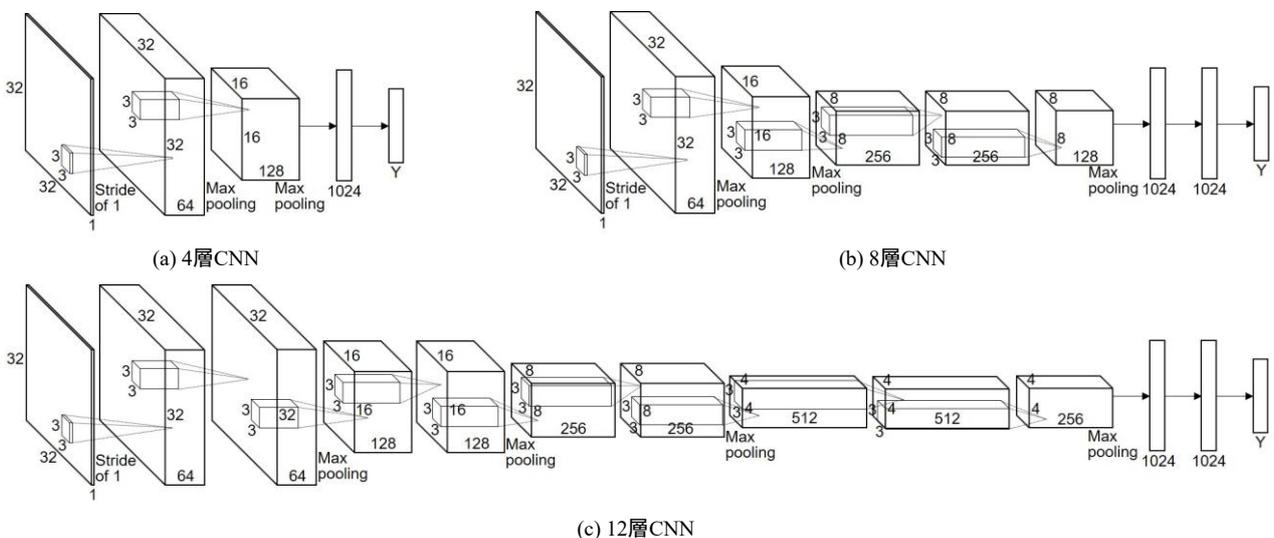


図2 構築した各畳み込みニューラルネットワークの構成図



(a) 分類対象画像のラベル (クラス名)



(b) 学習に使用した画像例 (Class-3)

図3 実験に使用した画像データ

から、BoF+SVMにおいても入力画像のサイズをCNNと同一とした。しかし、このサイズで特徴量抽出を実行すると、昨年度の実験で使用した局所特徴量の中には、十分な数のキーポイント(特徴量が記述された画像上の特徴点)が得られない局所特徴量もあったため、最も多くのキーポイントが得られたSIFT<sup>9)</sup>を特徴量抽出に使用した。また、Visual Word (クラスタリングされた特徴量の代表) 数は1000として、コードブック (Visual Wordから構成される辞書) を生成した。

### 3. 2 実験結果

実験結果を表1, 2に示す。なお、表1の分類率については、キーポイントが得られなかった画像データを不正解として全画像数で除した場合の値であり、括弧内は、キーポイントが得られた画像数で正しく分類された画像数を除した場合の値である。

表1, 2を比較することより、CNNの分類性能の高さがよく分かる。本実験では、入力画像サイズを32×32としたため、BoF+SVMにとっては、厳しい条件ではあったが、それにもかかわらずCNNは高い分類率を達成することができた。このことから、CNNを画像検査等に利用することにより、検査精度の向上が期待できると考えられる。

また、表2のTest-Aに対する結果より、同一条件で取得した画像データであれば、CNNを利用することにより、未学習データであっても高い精度で分類可能であることが分かる。しかし、取得条件が異なるTest-Bの分類結果については、高い精度で分類することが可能ではあるものの、12層CNNについては、他のCNNと比較して低い分類率となった。一般的に、層が深くなれば分類性能が向上すると思われるが、どのような画像データを何クラスに分類するかにより、適した層の深さがあると考えられ、設定したパラメータや、学習に利用した画像データ数等についても検討する必要があると考えられる。

これらの結果から、本実験で対象とした画像分類問題に対しては、4層から8層のCNNを構成すれば、十分な精度で分類可能であることが分かる。勿論、層が深くなるにつれて処理時間を要するため、どのような条件でCNNを利用するかも踏まえてCNNを構築し、画像検査装置等のシステムに実装する必要がある。

表1 BOF+SVMによる画像分類率 (単位: %)

クラス名	Test-A	Test-B
Class-1	71.50 (96.75)	77.50 (95.98)
Class-2	89.80 (99.89)	94.00 (94.00)
Class-3	80.10 (96.27)	83.75 (83.75)
Class-4	89.20 (94.99)	82.25 (82.66)
Class-5	89.40 (99.55)	99.25 (99.50)
計	84.00 (97.52)	87.35 (90.99)

表2 CNNによる画像分類率 (単位: %)

クラス名	4層CNN		8層CNN		12層CNN	
	Test-A	Test-B	Test-A	Test-B	Test-A	Test-B
Class-1	100.00	100.00	100.00	100.00	100.00	100.00
Class-2	100.00	100.00	100.00	100.00	100.00	99.75
Class-3	100.00	99.50	100.00	99.50	100.00	100.00
Class-4	100.00	96.25	100.00	99.50	99.90	96.00
Class-5	100.00	98.50	99.90	100.00	100.00	91.00
計	100.00	98.85	99.98	99.80	99.98	97.35

## 4. クラウド型画像分類システム

一般的に、CNNによる画像分類プログラムの開発においては、ネットワークの構築やパラメータの設定にノウハウを必要とされるが、中小企業においては、どのようなデータを、どのような手順で入力したらいいかといった前処理の段階で、戸惑うケースが少なくない。書籍やWeb等に掲載されているCNNに関する情報は、ネットワークやサンプルプログラムに関する情報が殆どであり、サンプルプログラムにおいては、既にパッケージ化された入力データが使われている。そのため、利用者自身が用意した画像データについて、それらの入力方法に関する情報にたどり着けず、CNNの利用に至らないことがある。

そこで、本研究では、クライアント端末からサーバ上のコンピュータリソースを用いて、CNNによる画像データの学習・分類が可能なクラウド型の画像分類システムを構築するとともに、サーバ上で画像分類処理を実行するために必要な操作を、クライアント端末から簡単に行えるWebアプリケーションを開発した。開発したアプリケーション画面を図4に示す。本アプリケーションは、4つの機能から構成されており、データセット/クラスの作成、画像ファイルのアップロード、CNNによる学習モデルの作成、CNNによる画像ファイルの分類が可能となっている。

入力データを準備する際には、画像データとそれらの保存先のファイルパス、ラベルが記述されたCSV形式のドキュメントファイルを用意することが一般的であるが、多くの入力データを用いる場合は、手間のかかる作業となる。そこで、本アプリケーションでは、この作業を省略することができるよう、データセット/クラス作成機能により画像データの保存先を作成し、画像ファイルアップロード機能により保存先毎にデータを管理することで、データ構成



図4 開発したWebアプリケーション画面

を読み取り、画像データとクラスとの対応関係を自動生成する仕様とした。なお、データセットは課題別に、クラスは各課題における対象別に画像データを管理することを目的としている。

CNNにより画像进行分类するためには、予め学習モデルを生成しておく必要がある。CNN学習モデル作成機能では、データセット、CNNの種類を選択した後、学習を繰り返す回数（エポック数）を入力し、学習ボタンをクリックすることで学習モデルを生成する仕様とした。学習モデルの生成に要する時間は、入力する画像サイズ、CNNの種類、エポック数等により異なるが、3章の実験における4層CNNの学習モデル作成時間は1分弱であった。なお、エポック数に到達する前に、識別精度が1（学習データの分類率が100%）となるエポックが5回に達すると、過学習と呼ばれる学習データに特化したモデルが生成されることを防止するため、自動的に学習を終了する。

CNN画像分類機能では、選択した学習モデルを使って、送信した画像ファイルをクラス毎に分類する。分類結果については、クラス毎に集計して出力する仕様としているため、複数ファイルを同時に送信しても、CNNによる分類性能の把握が容易となっている。

## 5. まとめ

本研究では、CNNを用いた画像認識技術に対する理解の促進と中小企業への導入の支援を目的として、簡単にCNNを用いた画像分類技術を利用することが可能なシステムを構築した。システムの構築にあたっては、層の深さの異なる3種類のCNNを構成し、分類性能の差を評価した。本システムを使用することで、CNNを利用した画像分類処理を容易に試行できるため、深層学習の導入の検討に有用である。

今後は、CNNだけでなく、他用途を対象としたDNN（Deep Neural Network）についても検討し、システムへの実装を目指す。

## 文献

- [1] 麻生秀樹, 安田宗樹, 前田新一, 岡野原大輔, 岡谷貴之, 久保陽太郎, ボレガラ・ダヌシカ, “深層学習—Deep Learning—”, 近代科学社, 2015.
- [2] 内田祐介, 山下隆義, “畳み込みニューラルネットワークの研究動向”, 信学技報, Vol.117, No.362, PRMU2017-105, pp.25-38, 2017.
- [3] 内山育海, “ソフト基盤でAIが身近に, 開発支援サービスはこう選ぶ”, 日経エレクトロニクス, 2017年6月号, pp.61-69, 2017.
- [4] 中小企業庁, “2017年版中小企業白書”, <http://www.chusho.meti.go.jp/pamphlet/hakusyo/index.html>, 2017.
- [5] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, In Proc. of NIPS, pp.1097-1105, 2012.
- [6] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, In Proc. of ICLR, 2015.
- [7] 渡辺博己, 曾賀野健一, 松原早苗, 棚橋英樹, “機械学習を用いた作業時間推定システムの開発”, 岐阜県情報技術研究所研究報告, No.18, pp.15-21, 2017.
- [8] G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray, “Visual Categorization with Bags of Keypoints”, In Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp.59-74, 2004.
- [9] D.G. Lowe, “Object Recognition from Local Scale-Invariant Features”, In Proc. of ICCV, pp.1150-1157, 1999.