G-26

ディープラーニングによる画像認識
Image recognition by deep learning

○青山美寿々[1]，木村剛志[1]，山内耀介[1]
*Misuzu Aoyama, Takeshi Kimura, Yousuke Yamauchi

Abstract: Recently, deep learning (DL) has played a central role in the field of artificial intelligence. In particular, DL has achieved rapid and dramatic advancements in the image and speech recognition areas of study. Modern devices, such as smart-phones or tablet PCs, also benefit from such advancements. In this paper, we focus upon a convolutive neural network (CNN) and DL applied to image recognition tasks and report some simulation results.

１．はじめに

近年，Deep Learning は画像認識や音声認識分野において多くの成果を収め，例えばスマートフォンやタブレット端末の開発においても重要な技術の一つとなりつつある．Deep Learning はここ数年で非常に速いスピードで発展を遂げている．本論文では Deep Learning による画像認識に焦点をあて，代表的なモデルの一つである CNN を用いたシミュレーション実験結果について報告する．

２．画像認識

画像認識とは，画像データの内容を分析し，その形状を認識する技術のことである．その技術の一つとして，画像データから対象物となる輪郭を抽出し，背景から分離した上で，その対象物が何であるかを分析する技術が挙げられる．

３．Deep Learning[1][2]

Deep Learning とは文字通り深層学習を意味し，従来ニューラルネットワークを用いた人工知能の機械学習の一つであり，人間の脳のおける知識獲得に近い仕組みを実現できると考えられている学習アルゴリズムである．通常，深く大規模なネットワーク構造の学習を伴うことが特徴であり，それゆえに深層学習と呼ばれている．

Deep Learning では，これまでは人手で行う必要のあった特徴量の抽出が自動で行われる．画像から何らかの特徴量を抽出する場合，SIFT（Scale-Invariant Feature Transform）や SURF（Speed-Upped Robust Feature）などと呼ばれる，いわゆる“設計済み”の特徴抽出法を用いることが多い．また教師信号を用いずにデータから自動的に特徴量を抽出する教師なし学習が使われている．ディープラーニングは教師のあるなしに関わらず大量のデータを用いて，種々様々な色やエッジを組み合わせ，与えられたデータのクラス分類や回帰を行うネットワークの学習を行う[3].

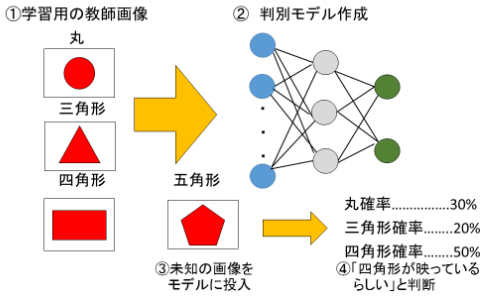

Figure 1. 丸，三角，四角の判別モデル

ここで，一つ具体例を示す：まず，大量の画像と，画像に映っている物体を訓練データとして用意する．これらを学習データとして用い，学習モデルを作成する．Figure 1 は，「丸」，「三角形」，「四角形」が含まれる大量の画像を準備し，それぞれ画像と物体名とを組にした学習データを基に，「識別器」を作成する，といった一例を示している．学習後，Figure 1 における識別器は,未知の画像を入力すると「丸」，「三角形」，「四角形」のどれが映っているのかという結果が出力される．また画像認識の精度は，通常，用いられる学習アルゴリズムに大きく依存する．

4. CNN[4]

パターン認識とは，実際に与えられた情報から特徴を読み取り，どのような種類に属するかを判断することである．一方，そのような判別を行うニューラルネットワーク（NN）は生物の脳の神経の動きを数学的手

１：日大理工・学部・数学

法により情報処理メカニズムに転換したアルゴリズムの一つである．ニューラルネットワークに「畳み込み」の意味を持つ Convoltionual を加えた CNN は通常のニューラルネットワークに加えて画像認識で上下左右に位置関係を考慮し，より抽象的な情報に転換，それらを通常のニューラルネットワークでパターン認識するという二段階の構成になっている. すなわち, 入力層, フィルタをスライドさせて画像に畳み込む畳み込み層, 畳み込み層から出力された特徴マップを縮小するプーリング層, 認識するクラス数の確率を求める全結合層, 出力層から構成される．畳み込み層とプーリング層が複数連なることで「深い層」が形成され，その後の全結合層も同様に何層か続く構造から成る．従来の機械学習ではあらかじめが明示的に特徴を定義する必要があり, 複雑な特徴になると表現することが困難となる．一方ディープラーニングでは人工知能が学習データから特徴を抽出することによって複雑な特徴も抽出することが可能となる．

5. Caffe[5]

Caffe は, Berkeley Vision and Learning Center においてオープンソースとして開発されている Deep Learning のフレームワークの一つである．また Caffe では，ネットワークのモデルや最適化の方法をパラメータファイルを用いて設定することができるため，簡単なモデルがあれば，コーディングすることなく動かすことができ．さらに，CUDA，GPU 環境での動作に対応しており，GPU での学習の実行時間が他のツールよりも高速であることも評価されており,他のフレームワーク（Theano や Pylearn2 など）と比較しても,学習，評価を行う際の処理速度も高速である．また，Web アプリケーションの開発によく利用されている言語の一つである Python と Caffe を併用すると，Web アプリケーションを組み込むことが比較的容易に行える．

6. 予備実験と本実験

Caffe は事前に 1000 種類のカテゴリに分類する学習済みのリファレンスモデルが内蔵されており，本予備実験では実際に未知の画像を読み込ませどの画像に分類されるかを検証する．

実験では，Caltech101 と呼ばれる画像データセットの中に入っているカップの画像を読み込ませ，それらの分類を試みた．

番号	認識対象物	類似度
#1	cup	33.1%
#2	caldron, cauldron	20.2%
#3	candle, taper, wax light	11.6%

Figure 2. 予備実験の結果

実験では Figure 2 に示されるように, カップに 33.1%, 大鍋に 20.2%. ロウソクに 11.6%類似したという結果が得られた．

なお本実験では，引き続き Caffé を用いて，学習器を生成し，元画像 2 種類とその他誤り画像をそれぞれ 1000 枚ずつ用意し，それぞれをフォルダ分け，正規化処理を施し，画像分類を実行する．以上述べた結果については，学術講演会当日に報告する予定である．

7. まとめ

本研究では，教師あり・なし学習で画像をより的確に識別できることがわかった．今後は，ディープラーニング・フレームワークの一つである Caffe を引き続き用い，認識率の向上についてさらに検証を進めて行く予定である．

8. 参考文献

[1] Yee-Whye Ten，Geffrey E.Hinton，Simon Osindero :「A Fast Learning Algorithm for Deep Belief Nets」，2005 年 6 月．

[2] 中山　英樹 :「Deep learning と画像認識　～歴史・理論・実践～」，http://www.slideshare.net/mobile/nlab_utokyou/deep-learning-40959442，　2014 年 10 月．

[3] 山下隆義 :「イラストで学ぶディープラーニング」，講談社，　pp．1~3，pp．36~47，　2016 年 3 月．

[4] blog.li.nu :「Caffe を使った CNN による手書き文字分類[1]　- CNN の概要，データセットの準備」，http://li.nu/blog/2016/02/handwritten-number-classification-with-caffe.html，　2016 年 2 月．

[5] 武井宏将 :「初めてのディープラーニング」，リックテレコム，pp．12~14，pp54~57，pp68~95，2016 年 1 月．