

深層学習を用いた甲骨文字認識

紙徳 直生[†] 多田 晃己[‡] 伊藤大喜[‡] 孟 林[‡] 山崎 勝弘[‡]
立命館大学大学院 理工学研究科[†] 立命館大学 理工学部[‡]

1. はじめに

甲骨文字は、図1のような3000年以上前の中国殷代で使用された象形文字で2000種類以上存在し、亀の甲羅や獣の骨に刻まれている[1]。これらの解読と認識は、当時の歴史や文字の起源を知る上で非常に重要であるが、長い年月を経て劣化し、認識が困難である。我々は画像処理を用いたテンプレートマッチング[2]などによる甲骨文字認識を進めてきたが、ノイズなどが原因で誤認識するなどの問題がある。本研究では、深層学習の一種である畳み込みニューラルネットワーク(CNN)を用いて高精度な甲骨文字認識を目指す。学習に使用する甲骨文字を収集し、回転・切り取り等のデータ増強を行い、学習に使用するデータセットを大幅に増やす。また、ネットワーク構造やパラメータの最適化を行い、認識率を向上させることを目標とする。

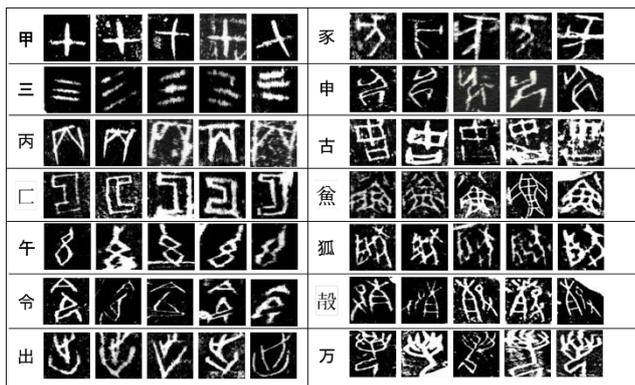


図1 甲骨文字原画像例

2. 甲骨文字データセット

CNNの学習には、大量のデータセットを作成しなければならない。データセットには、ネットワークを学習するための学習画像と教師画像、学習したネットワークを評価するテスト画像を用意する必要がある。

データセット構築で使用する原画像は、拓本集[1]をスキャンし、1358枚の画像を切り出したものである(図2)。そのうち、820枚を学習に使用する教師画像と訓練画像と定義し、残りの538枚をテスト画像として認識率の評価に使用する。各文字について、断裂やノイズが少ない最も綺麗な画像1枚を教師画像、残りを訓練画像とする。

3. データ増強

原画像の数には限りがあり、大量のデータを用意することは困難である。そこで、原画像1枚に対して、図3の

Recognition of Oracle Bone Inscriptions Using Deep Learning
Naoki Kamitoku[†], Koki Tada[‡], Ito daiki[‡],
LinMeng[‡], Katsuhiko Yamazaki[‡]
[†]Graduate School of Science and Engineering, Ritsumeikan University. [‡]College of Science and Engineering,

114種 1358枚



図2 実験で用いた甲骨文字

ように反転・回転・切り取り・輝度値変更・ノイズ付与の順で処理を行い、学習に使用するデータを増強する。これにより原画像一枚あたり、反転(2)×回転(13)×切り取り(17)×輝度値(3)×ノイズ付与(5)=6630枚の画像が得られる。

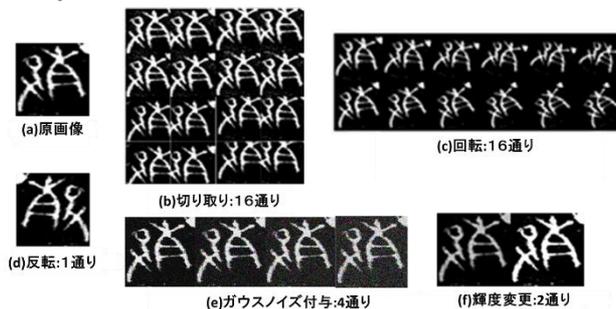


図3 データ増強

4. 深層学習による認識

CNNは深層学習の手法の一つであり、多数の層を交互に配置したものである。本研究で使用する Alexnet[3]を図4に示す。Alexnetは、5層の畳み込み、3層のプーリング、及び3層の全結合層で構成されるCNNである。畳み込み層では、入力に複数のフィルタをかけて複数の特徴マップを出力する。得られる特徴は、畳み込み層が進むにつれて段階的に抽出される。プーリング層では、特徴マップの次元を削減するほか、並進移動に対しての不変性を保つ。一層目と二層目の畳み込み層の出力にのみ、輝度の正規化を行う局所応答正規化(local response normalization, LRN)を適用する。そして、畳み込み層とプーリング層によって得られた特徴マップから、全結合層でクラス分類を行う。全結合層の出力にはSoftmax関数が使用され、クラス毎の類似度が確率的に算出される。Alexnetでは、6層目と7層目の全結合層で、パラメータ0.5のDropoutが使用されている。Dropoutは、学習時にランダムに選択されたユニット出力を0にすることで、パラメータの増加を抑えるとともに過学習を抑制する。

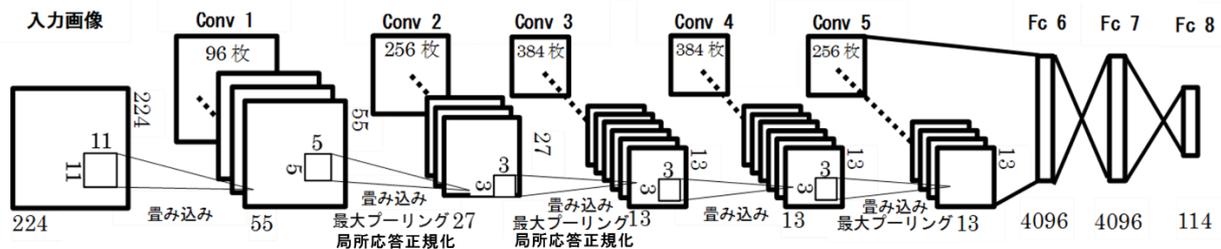


図4 Alexnet

本研究では、パラメータの最適化に確率的勾配降下法 (SGD) を利用し、パラメータは 0.01 とする。Alexnet を採用した理由は、分類するクラス数が多い点と、一般的な手書き文字よりもノイズや断裂が多く、Lenet のような層が浅いネットワークでは、特徴が上手く抽出できないからと考えたからである。

5. 実験

5.1. 実験内容

データ増強の手法による認識率の変化を調査する。また、6,7 層目の Dropout の値をそれぞれ 0.3~0.7 の 25 通りに変更して学習し、テスト画像 538 枚で認識を行った。

CNN の実装には、caffe を使用した。入力画像サイズは 224×224、教師画像の認識率が収束するまで学習を行う。実験に使用した GPU マシンは CPU: Xeon E5-1620v3, GPU: TITAN X(Pascal) ×2, メモリ: 64GB である。

5.2. 実験結果

表1にデータ増強と Dropout による認識率の比較を示す。データ増強では、切り取り・ノイズ付与・回転の効果が高いことが分かった。また、輝度単体では 40% の認識率しかないが、回転・切り取り・反転の3種で 87%、輝度を加えた4種で 90% に向上したことから、組み合わせると有効であることが分かった。さらに、ガウスノイズを付与することで認識率は 92.9% になった。Dropout を変更した認識実験では、6,7 層目の Dropout をそれぞれ 0.3~0.7 の 25 通りに変更した結果、6 層目 0.7、7 層目 0.5 にすることで認識率 94.4% を達成することができた。

認識できた画像を図5に示す。"今","貞"のようなノイズの激しい画像や、"不","万"のような一部が欠損している画像が認識できていることから、深層学習を用いた甲骨文字認識は、ノイズやある程度の断裂に有効であることがわかった。

5.3. 考察

学習したモデル自体の認識率とテスト画像の認識率はほとんど変わっておらず、甲骨文字認識において、過学習はあまり起こっていないと考えられる。そのため、過学習を抑制するための Dropout の値を変更することで認識率を向上させることができた。

また、データ増強の手法では切り取りが最も有効であったが、デメリットも存在する。例えば、"三"という文字に対して 80% 切り取りを行うと、線が一本無くなり"二"という文字になってしまう。学習時に"三"と"二"を混合してしまうので誤認識の原因となる。この問題は回転でも起こるケースがある。そこで、すべての文字に対してすべてのデータ増強を行うのではなく、文字ごとに最適

なデータ増強を行うことによってデータセットの質を向上させることが重要である。

現在、甲骨文字を 114 種 820 枚から 184 種 2000 枚に文字種を増やして実験を行っている。認識率が 90% に低下したので、最適なデータ増強の手法やネットワーク構造・パラメータの最適化を行っている。

また、深層学習は認識率が高くて、推論過程がわからない問題がある。そこで、なぜその結果になったかを考察するために、特徴マップなど様々なものを可視化し、推論過程の説明機能の実現を目標に研究を行っている。

表1 データ増強と Dropout による認識率の比較

データ増強	学習画像数	認識率 (%)	Dropout	認識率 (%)
なし	820	39.2	0.5,0.5	92.9
輝度	2460	40.1	0.7,0.5	94.4
反転	1640	51.1	0.6,0.5	94.1
回転	10660	68.9	0.6,0.7	93.5
ノイズ付与	4100	76.1	0.8,0.5	92.7
切り取り	13940	87.5	0.5,0.7	92.2
すべて	5436600	92.9	0.5,0.6	92



図5 認識できた画像

6. おわりに

本研究では、深層学習を用いた甲骨文字認識において、94.4% の認識率を達成した。また、甲骨文字のデータ増強やパラメータの調整によって認識率を向上させることができた。今後の課題として、甲骨文字認識における最適ネットワーク構成の提案、推論過程の可視化、及び説明機能の実現などが挙げられる。

参考文献

[1] 左:上海博物館蔵甲骨文字, 上海辞典出版社, 2009.
 [2] L.Meng, Y.Fujikawa, A.Ochiai, T.Izumi, and K.Yamazaki: Recognition of Oracular Bone Inscriptions Using Template Matching, International Journal of Computer Theory and Engineering, Vol.8, No.1, pp.53-57, 2016.
 [3] Krizhevsky, Alex, et al: "Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012.
 [4] 紙徳, 孟, 山崎: 深層学習を用いた二段階甲骨文字認識, FIT2017, F-020, 2017.