

深層学習を用いた二段階甲骨文字認識

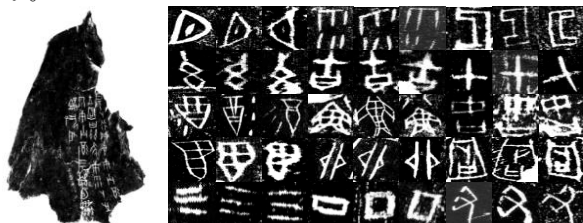
紙徳直生† 孟林‡ 山崎勝弘‡

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 3000 年以上前の象形文字で 2000 種類以上存在し、図 1 に示すような亀の甲羅や獣の骨に刻まれている。これらの解読と認識は非常に重要であるが、劣化が原因で認識が困難である。我々はテンプレートマッチングなどによる甲骨文字認識を進めてきた[2][3]。

本研究では、二段階の深層学習を用いて高精度な甲骨文字認識を目指す。認識率を高めるために、一度目の学習と認識後に、最上位の文字を削除したデータセットに更新し、再度学習と認識を行う。甲骨文字 114 種 820 枚に対して、回転・切り取り等のデータ増強により得られた約 60 万枚のデータセットで学習を行い、テスト画像 538 枚で認識を行う。



(a)甲骨

(b)甲骨文字

図 1 甲骨と甲骨文字

2. 二段階甲骨文字認識

図 2 に二段階甲骨文字認識の流れを示す。本システムは、データセットの構築、一段階目の学習と認識、二段階目の学習と認識により構成される。

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。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。また、教師画像と訓練画像にそれぞれデータ増強を行い、大量の画像を用意し、これらを学習用データセットとする。

次に、畳み込みニューラルネットワーク (convolutional neural network: CNN) を用いて一段階目の学習と認識を行う。類似度が最上位の文字が正解でない場合、最上位の文字を削除したデータセットに更新し、再び CNN を用いて二段階目の学習と認識を行う。二段階目では誤認識した文字数分繰り返して行い、学習時間は一段階目にかかった時間にデータセットを更新した回数をかけたものである。

3. 深層学習を用いた甲骨文字認識

3.1 データ増強

CNN では、大量のデータを用いて学習することで認識率が高くなることが知られているが、学習用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 のように反転・回転・切り取り・輝度値変更の順で処理を行い、学習に使用

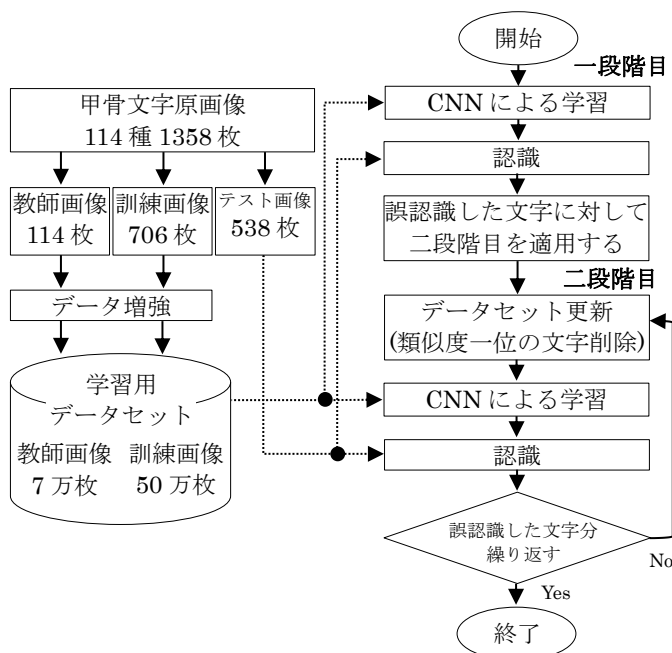


図 2 二段階の甲骨文字認識

するデータを増強する。これにより、原画像 1 枚あたり、反転(2)×回転(13)×切り取り(9)×輝度値(3)=702 枚の画像が得られる。



図 3 学習用データ増強

3.2 畳み込みニューラルネットワーク

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 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。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。その後、誤差逆伝搬法、及び確率的勾配降下法を用いて各層の重みの値を更新する。

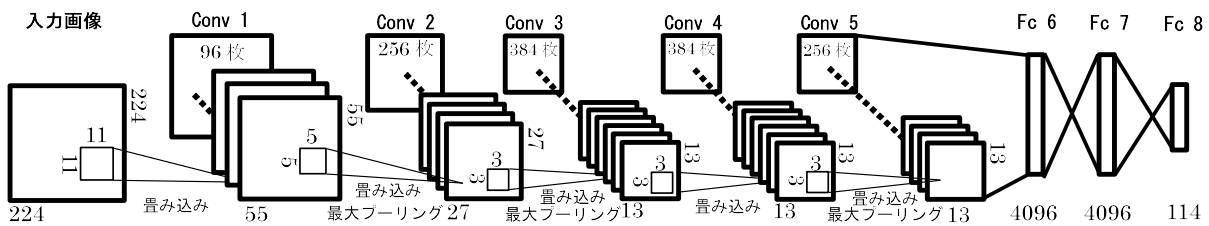


図 4 Alexnet

3.3 二段階の学習・認識

学習済み CNN にテスト画像を入力した際の二段階認識結果の例を図 5 に示す。入力画像”貞”(a)に対して、一段階目の認識結果(b)は類似度最上位の文字が入力画像とは異なっている。ここで”辰”を削除したデータセットに更新して、再度学習と認識を行った結果、(c)のように”貞”の類似度が最上位となった。また、一段目の類似度上位 5 位以内に入力画像の文字が存在しない場合でも、二段階目を適用することで認識できるようになったケースも存在する。よってこの手法で、一段階目の二位まで含めた認識率よりもさらに認識率の向上が期待できる。



(a)入力：貞 (b)一段目認識結果 (c)二段目認識結果
図 5 二段階認識結果の例

4. 実験

4.1 実験内容

拓本[1]から切り出した 538 枚の画像を用いて一段階目の認識を行った。また、一段階目で認識できなかった 53 枚の画像について、それぞれ最上位の文字を削除したデータセットに更新し、二段階目の学習と認識を行い、一段階目の二位まで含めた認識率と比較する。さらに、教師画像について 3.1 に示す通りデータ増強を行った場合、半分程度行った場合、なしの場合を評価する。

CNN の実装には、caffe を使用した。入力画像サイズは 224x224、学習率を 0.01 とし、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) x2, メモリ: 64GB である。

4.2 実験結果・考察

表 1 に一段階目と二段階目の認識率と学習時間を示す。二段階目は 95.91% の認識率を達成しており、一段階目二位までの認識率より高いことが確認できた。一段階目で誤認識した文字が 27 種類あったことから、二段階目の学習時間は一段階目の 27 倍となった。表 2 に二段階目で認識できるようになった画像を示す。入力”癸”の一段階目では、”牛”が類似度最上位として認識されている。これは”癸”と”牛”の特徴である斜め線と交差点の位置が似ているからと考えられる。しかし、”牛”を削除したデータセット

を作成し二段階目の学習と認識を行った結果、正しく認識できるようになった。同じく、入力”三”の一段階目では横線が一本多い”四”を認識しているが、二段階目を適用することで、誤認識されなくなった。二段階目を適用して認識できるようになった画像の多くは、一段階目の類似度最上位の文字が入力画像と似るケースが多いことが確認できた。

また、教師画像に対して、データ増強の有無を見てみると、すべて行った場合が最も認識率が良くなっていることが確認できた。この結果から、教師画像に対してのデータ増強は有効であると考えられる。

表 1 認識結果と学習時間

データ増強	段数	認識枚数	認識率(%)	学習時間
Full	一段階目	485/538	90.14	2h37m
	一段階目 二位まで	503/538	93.49	
	二段階目	516/538	95.91	
Half	-	478/538	88.84	50m
なし	-	474/538	88.10	1h20m

表 2 二段階目で認識できた画像

入力	認識の結果		入力	認識の結果	
	一段	二段		一段	二段
	癸	牛	癸	癸	癸
	王	王	王	王	王
	甲	甲	甲	甲	甲
	王	王	王	王	王
	卯	卯	卯	卯	卯
	卯	卯	卯	卯	卯
	卯	卯	卯	卯	卯
	卯	卯	卯	卯	卯

5. おわりに

本稿では、甲骨文字の認識率の向上を目指して、二段階深層学習による認識手法を提案した。一段目の学習と認識で誤認識した場合、最上位の文字を削除したデータセットに更新し、再度学習と認識を行った。114 種類 820 枚の甲骨文字から大規模なデータセットを構築し、二段階の深層学習で 538 枚の甲骨文字を認識した結果、一段階目で 93% を第二位以内に認識し、二段階目では 96% の認識率を達成できた。

今後の課題として、学習と認識に使用する文字数を増やすこと、新たな CNN の構造の検討、及びデータ増強の最適化などが挙げられる。

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]岸, 石井, 孟, 山崎:近傍ベクトルを用いた甲骨文字の特徴点と線の検出, 情報処理学会第 79 回全国大会, 2N-01, 2017.
- [3]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.