

CenterNet を用いたくずし字の認識

陳 楽涵[†], 呂冰[†], 富山 宏之[‡], 孟 林[‡]立命館大学 理工学研究科[†] 立命館大学 理工学部[‡]

1. はじめに

多くの日本古典籍はくずし字で記述されている。くずし字は文章を迅速に書く、簡略化された書体で、現代があまり使われていないため、日本古典籍の解読が難航している。深層学習は物体認識として幅広く応用されているため、くずし文字の認識にも貢献できると考えられる。従って、我々は深層学習を用いて、くずし字の自動認識を目指している。本研究は 2019 年 4 月に提案された最新のモデルである CenterNet [1] を使用し、くずし字の自動抽出と認識を行い、評価を行った。我々は、文学オープンデータ共同利用センター (CODH) が提供した日本古典籍くずし字データセット [2] を用いて CenterNet モデルの学習を行う。また、より効率に学習するために、データを拡張し、その有効性を示す。

2. くずし字データセット

2.1 くずし字とデータセット

平安時代以来明治時代まで文書が主に印刷体と全然違うくずし字で書かれている。くずし字は、一つの平仮名と漢字は複数の字形があるだけでなく、字の幅と長さが印刷体のように同じではない。また、字を書きやすくするために、簡略化されて筆画が結んでいる。そんな特点是図 1 に示す。それにより、日本古典籍の認識が難問となり、古代歴史の研究の難しさも増やしている。

従って、研究者らは日本の古典籍を集めて、オープンデータセットを作って、くずし字認識の研究を進めている。本研究では、CODH が提供したくずし字のデータセットを用いて、深層学習で文字認識の研究を進行している。今回の実験では、虚南留別志や源氏物語などの 29 冊の古典籍から、4212 種類のくずし字を含めた、60 万枚の単個文字の画像 [1] を利用した。また、認識精度を向上するため、単個の文字データの拡張と、CenterNet 学習のデータセットの拡張も行った。

2.2 データセットの拡張

実験により、CODH の日本古典籍くずし字データセットには、各種類文字の学習データの数が異なるというデータのアンバランスの問題が存在することが分かった。表 1 にはその問題を示している。

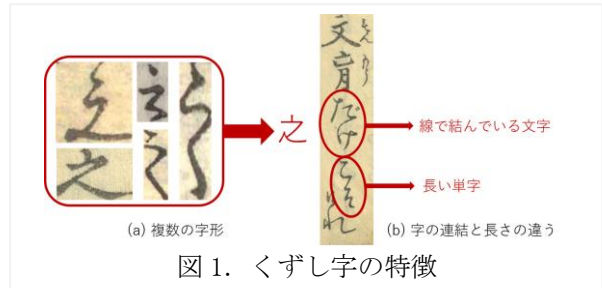


図 1. くずし字の特徴

表 1. 文字画像の統計

U+4E00	一	3114
U+4E0B	下	830
U+4E16	世	705
U+4E4D	乍	2
U+5AB3	媳	1

筆画の少ない文字、例えば量詞と単純な平仮名や助詞などの出現回数が多い、一方、筆画の多い漢字と意味の深い文字の出現回数が少ない。文字認識に対して、深層学習のモデルを訓練するための訓練画像が多くなれば多いほど良いと言えない、各文字種の訓練データ数が近いほうが良いと考えられる。

従って、我々は、まず単個文字画像の拡張を行った。本研究では、複数の画像処理方法を用いて、文字を変形せずに、画像の特徴を少し変化し、1 種類の文字の学習画像を最小 1000 枚に生成する。その方法は以下のように実現し、主に画面と位置の調整などとなる。

- (1) **色の変更**: 二つの方法を用いて色の変更を行う。方法 1 では全部画素の数値を一定の値を足す。方法 2 はグレースケールを変化し、画素の数値を一定の値足す。
- (2) **ノイズの追加**: ごま塩ノイズ又はガウスノイズを用いて、元画像にノイズを追加する。
- (3) **字の中心への移動**: 字を変形せずに、画面の中心に少し移動する。
- (4) **字の回転**: 字を時計回りに回転する
- (5) **字をリサイズする**: 字を変形せずに、一定的な比率に縮小と拡大を行う。
- (6) **画面の膨張と収縮**: 膨張と収縮フィルタを用いて、画像の膨張と縮小を行う。

データ拡張を実行するフローチャートは図 2 に示す。くずし字データセットの各単個文字画像のデータ拡張後、拡張された単個データをランダムに選んで連結することにより、一枚の画像に複数の文字を

「Kuzushi Character Recognition based on CenterNet」

† 「Lehan Chen and Bing Lyu · Graduate School of Science and Engineering, Ritsumeikan University」

‡ 「Hiroyuki Tomiyama and Lin Meng · College of Science and Engineering, Ritsumeikan University」

有する CenterNet 訓練用の画像データを作成する。画像データの拡張では、一枚画像から大量な画像を生成できる。効率的な実験を行うために、できる限り元画像データを使用し、各種類の文字の画像を1000枚までに生成する。さらに、単個文字画像をランダムに連結し、画素は512×512のCenterNet用のデータ画像を生成する。本実験で、連結で生成された画像は66000枚となり、訓練データとテストデータと分けて、くずし字を認識する。

3. 深層学習によるくずし字認識

3.1 CenterNet

CenterNet とは、選択された画面中心部の特徴量を抽出し、物体を検測と分類するニューラルネットワーク (CNN) である。Faster-RCNN と R-CNN などの既存モデルは大体 two-stage 分類器で構成されている。これらの分類の過程は、まず疎に目標らを検測して目標エリアが枠 (bounding box) で囲まれ、そして枠内の物体を分類することである。その方法は精度が高められるが、第一段階の処理後類似の枠が削除しなければならないから (普通は Non-maximum Suppression、NMS と呼ばれる)、時間上の効率が低くなるだけでなく、end to end の分類も困難になると思う。

CenterNet と YOLO や SSD などと同じ one-stage 分類器であり、NMS 処理が必要ではなく、効率ももっと高いと言われる。その上で、YOLO 又は SSD と比べて、CenterNet はただ毎種クラスのヒートマップによって、枠内エリア中心ポイントの特徴量を抽出し、物体を検測する。

4. 実験

4.1 実験内容

元データ画像は、CODH で提供されている 3881 ページを含めた古典籍 29 冊である。座標で分割した後、4212 個のくずし文字クラスを分けて全部 683,464 枚単個文字画像を取得する。

表 2 には、実験で使用された画像の条件と認識結果を示す。元画像は、スキャンされた古典籍の画像を示す。また、画像のサイズは、訓練およびテストで使用された画像のサイズで、モデルの必要に応じ

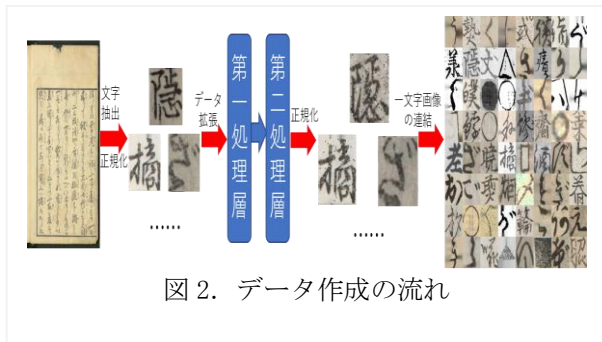


表 2. 実験条件と認識結果

画像種類	画像サイズ	モデル入力画像サイズ	訓練画像数	テスト画像数	認識率 (%)
元画像	2048 × 1024	320×320	800	56	17.8
		640×512	800	40	37.4
		768×480	3881	56	17.1
データ拡張画像	512 × 512	320×320	1000	40	33.6
		480×480	1000	40	37.4
		512×512	6600	100	83.7%

て、画像をモデル入力画像サイズにリサイズされる。我々は、画像サイズと訓練画像数を変更しながら、実験を行った。

動作環境は OS : Ubuntu 16.04、GPU : NVIDIA GeForce GTX TITAN X、Driver Version : NVIDIA-418、CUDA-10.1、Pytorch-1.2.0、OpenCV-3.6。

4.2 実験結果

CenterNet を用いて、認識結果が表 2 の右に示す。元画像の実験結果は、画像サイズ 640×512 の場合は画像 320×320 より、認識率の改善が見られた。データ拡張化した場合も同じ傾向がある。それにより、画像サイズは認識に影響することがわかった。また、データ拡張された場合は、現画像のみの使用より、大幅に改善していることも確認できた。したがって、提案した画像拡張の手法は、CenterNet の認識率の改善に意義があるとわかった。

5. おわりに

本研究は、深層学習を用いてくずし字の認識を目指す。CODH から入手したくずし字データセットを用いて、学習データセットの作成を行う。データ数のアンバランス問題に対して、六種類の画像処理手法を用いて、データの拡張を行った。今回の実験で生成した画像はデータ集としてモデルを訓練し、認識率が 85%以上に達するが、テスト集は元画像を使用するときに、認識率が非常に低い。今後、認識率が低い原因を分析し、データセットとアルゴリズムの調整が重要な課題となる。

参考文献

[1] 古典籍くずし字データセット
<http://codh.rois.ac.jp/char-shape/book/>
 [2] X.Y.Zhou, D.Q. Wang, P. Krähenbühl.: Objects as Points, ArXiv, 2019.
 [3] J. Yang, H. Zhang, M. Dencler, C. Lu.: Comparison of Shape-based and Stroke-based Methods for Segmenting Handwritten Chinese Characters, IEEE Int. Con. Conference on Computer and Information Science, 2005
 [4] 紙徳, 孟, 山崎.: 深層学習を用いた二段階甲骨文字認識, FIT2017, 2017.