

ARU-Net と LeNet を用いた日本古典籍の自動解読

呂 氷 (立命館大学大学院 理工学研究研究科)

富山 宏之・孟 林 (立命館大学 理工学部)

現在、数多くの日本古典籍が未だ解読されていない。また、多くの文字はくずし文字により記載されたため、日本古典籍の自動解読の難しさを増している。我々は、複数の深層学習モデルを用いて、日本の古典籍の自動解読を目指す。詳細には、まず、ARU-Net を用いて、文章ラインを自動的に抽出する。次に、ARU-Net による得られた文章ラインを用いて、文字サイズ毎にスライスし、文字の候補を生成する。最後に、候補である文字に対して、LeNet により認識を行い、高い信頼度の文字が認識結果とする。

Early Japanese Books automatic recognition by combining ARU-Net and LeNet

Bing Lyu (Graduate School of Science and Engineering, Ritsumeikan University)

Hiroyuki Tomiyama and Lin Meng (College of Science and Engineering, Ritsumeikan University)

Now, Lots of Early Japanese books are not be understand. Furthermore, a lot of these books are described by Kuzushi character, which increases the difficult of automatic understanding of Early Japanese Books. In this paper, we propose a deep learning based Early Japanese Books recognition method and aim to realize the automatic recognition. Firstly, the text lines of Early Japanese Books are extracted by ARU-Net, then the extracted text lines are sliced as a character size for obtaining the candidate of characters. At last the candidate characters are testing by LeNet for obtaining the recognition results and the confidence. The high confidence results are select as the final recognition results.

1. まえがき

日本古典籍は多くの情報を記録し、これらの古典籍の解読が、歴史、政治、文化の研究に非常に役に立つ。しかし、解読されていない日本古典籍は多く存在している。近年、人工知能技術の急速な発展に伴い、研究者らは日本古典籍に記録された情報を解読するために、深層学習を通じて日本古典籍に文字の認識を目指している。しかし、ほとんどの日本古典籍は、くずし文字で記録される。くずし文字は、バリエーションが多く、現在の日常に使われていなく、読める人が非常に少ない。そのため、文字の認識と日本古典籍の整理は非常に困難である。さらに、文字と文字を繋がっていること、文字のかすれが酷く、汚れ、虫食いなどにより、古典籍の自動解読の難しさを増している[1]。図1には、日本古典籍とその問題点の例を示す。

本論文は、複数の深層学習モデルを用いて、日本古典籍の自動認識を目指す。詳細について、まず、ARU-Net[1]を用いて文章ラインを抽出する。そして、抽出された文章ラインに対して、文字大きさを推定し、スライスして、文字の候補を切り取る。最後に、LeNet[2]を用いて、文字の候補から信頼度の高い文字を認識結果とする。評価実験について、立命館大学アート・リサーチセンターの古典籍データベースを用いる。

2. 先行研究

本論文は、複数の深層学習のモデルを用いて、日本古典籍の解読を目指している。古典籍の完全な自動認識のために、深層学習において、文字のセグメンテーション (character segmentation) と文字のクラシフィケーション (character classification) が両方必要である。

近年、物体のセグメンテーションとクラシフィケーションを同時に行う深層学習のモデルは提案されて、その中によく使われているのが YOLO (You Look Only Once)[4] と SSD (Single Shot MultiBox Detector) [5] となる。また、研究者らは、SSD を用いて、古代文献である甲骨文字の解読へ

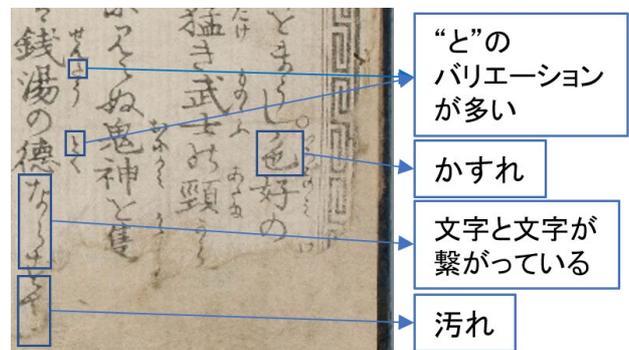


図1. 日本古典籍例とその問題点

の応用も行っている[6]。しかし、200枚ぐらいの画像で、40クラスの学習セットに対して、Nvidia CeForce GTX1080ti ×3, memory 64 GB のハードウェアリソースでの学習時間が2日間となる。日本古典籍を自動認識する場合は、文字のクラスと学習データ量は、甲骨文字よりはるかに超えているため、SSDでの日本古典籍の認識に、学習時間が非常にかかるとの問題点がある。

SSD、YOLOなどのモデルは、物体の場所が特定されていないことを想定し、実現したモデルである。一方、日本の古典籍は文章のラインが特定できる場合が多いため、我々は、文字のセグメンテーションとクラシフィケーションを二つに分けて実行することにする。

ARU-Netは、古文章の文章ラインを検出するための2段階深層学習手法である。

また、クラシフィケーションについて、非常に多くの深層学習モデルが提案された。それらのモデルは、既に高い精度を達成している。その中に、有名なのは、LeNet, AlexNet[7], GoogLeNet[8], VGG[9], ResNet[13]などが挙げられる。しかし、層が増えている傾向があり、学習時間が長くなる問題点が生じる。本論文は、一番シンプルなモデルであるLeNetを使用し、高精度な日本古典籍の自動識別を目指す。

深層学習は、古くから提案されたニューラルネットワークの発展である。なぜ、現在は非常に速いスピードで発展し、また、応用範囲が拡大しているの一つの理由は、ビックデータである。学習に必要なデータを十分に満たされているわけである。深層学習を用いて、日本の古典籍の自動解読も、複数のデータベースが実現されたことにより、まい進している。その中に、非常に貢献しているのは、人文学オープンデータ共同利用センターで公開されている日本古典籍くずしデータセ

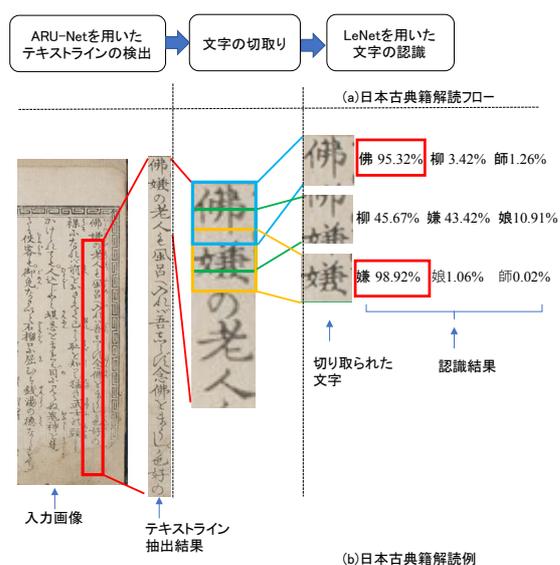


図2 古典籍解読の流れ

ットである[10]。日本古典籍くずしデータセットには、28冊の日本古典籍の4,645文字種で684,165文字画像を収集し、その上、各文字が画像内での座標、文字サイズなどの情報を纏めている。さらに、立命館大学のARC古典籍ポータルデータベースでは、膨大な日本古典籍をデジタル化し、閲覧と検索できる[11]。

3. 深層学習を用いた日本古典籍の自動解読

本論文は、ARU-NetとLeNetを結合し、日本古典籍文章ラインの抽出から、文字認識までを自動化していることを目指す。図2. は認識の流れとその例を示す。

入力画像に対して、ARU-Netを用いて、テキストラインに示された文章が得られる。そして、文字サイズを事前に決め、文字候補を切り取る。最後に、文字候補をLeNetにより認識を行う。図2の右にはLeNetによる各文字候補の信頼度を示す。それらの信頼度により、高いものが文字として正しく認識された。

3.1. 前処理

画像を直接にARU-Netで文章ラインを検出する場合は、枠ノイズが発生する。そのため、文章ラインを検出する前に、枠外の部分を切り除く必要がある。

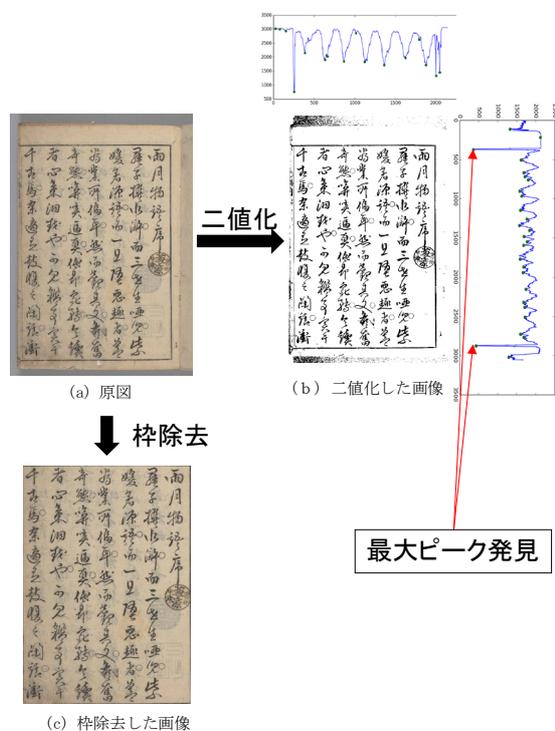


図3. 画像の枠を取り除き

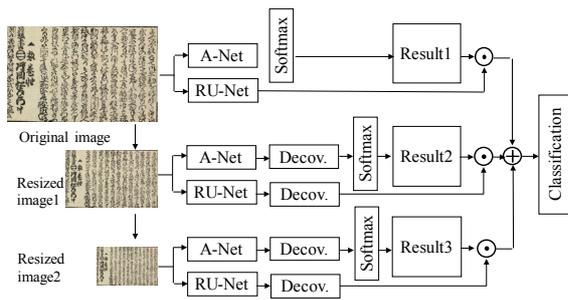


図 4. ARU-Net

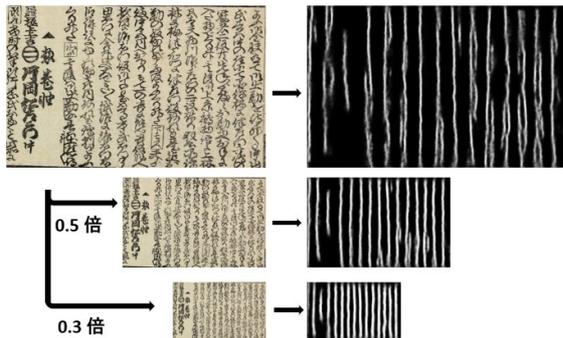


図 5. 縮小画像と文字列のセグメンテーション

図 3. は画像の枠ノイズを取り除くの流れを示す。

我々が、画像の 2 軸にピクセルの数を計算し、ピークとなる黒い点は、当該軸での場所となることにより、枠の場所を取得し、枠を除去する。具体的には、まず、図 3 (a) に対して、大津法により閾値を決定し、原画像の二値化処理を行う。そして、縦軸と横軸での白ピクセル数 (文字とノイズ部) をカウントする。両軸の黒ピクセルの数について、枠のラインでは非常に低いとの特性がある。従って、両軸のピクセルのピーク値 (3 つ以上がある) を探し、隣のピーク値間で微分演算を行う。各軸において、大きさが上位二つの微分結果を持つ座標は、枠となる。最後に、両軸で見つける 4 つの上位座標によって、画像を図 3 (c) のようにカットする [9]。

3.2. ARU-Net による文章ラインの抽出

ARU-Net は、英語の手書き手紙などのテキスト行を検出するための深層学習モデルである。図 4 は、ARU-Net の構成を示し、A-Net と RU-Net により構成される。A-Net は明暗領域を分別し、RU-Net はピクセルのラベリングを行う。詳細について、入力画像三回リサイズを行い、それぞれ、A-Net と RU-Net により、特徴量を抽出する。原画像の処理では、A-Net を経由し、softmax で得られた結果を、RU-Net で得られた画像を統

合する。同じように、リサイズされた画像を実行し、三種類の結果を用いて、分類を行う。

ARU-Net を用いて、オリジナルの日本古典籍のスキャン画像を処理すると、ページの枠はノイズとなり、セグメンテーションに悪影響となる。従って、前処理での枠除去が必要となる。前節で説明したとおり、原画像の枠ノイズを取り除くことができる。

入力された切り取った画像が大きいため、十分に成功を果たせないため、図 5 のように画像を元の画像にそれぞれ 0.5 倍と 0.3 倍に縮小し、0.3 倍に縮小すると結果が最良であることがわかる [12]。しかし、文献 [12] では、ARU-Net の結果により手動でテキストの切り取りを行うため、実用化するのが難しい。そのため、我々は、画像を 180 度回転し、上下二回の文字ラインを抽出することにより、テキストのセグメンテーションを実現した。

図 6. は ARU-Net を用いた文字例のセグメンテーションの流れと結果を示す。枠ノイズを取り除いた画像図 6 (a) に対して、ARU-Net により文章ラインを検出し、その結果は図 6 (b) に示す。そして、図 6 (a) を 180 度回転させ、再び ARU-Net により文章ラインを検出する。その結果は、図 6 (c) に示す。次に、図 6 (b)(c) を二値化し、それぞれ、図 6 (e)(f) を得られる。最後に、図 6 (e) の白線ごと

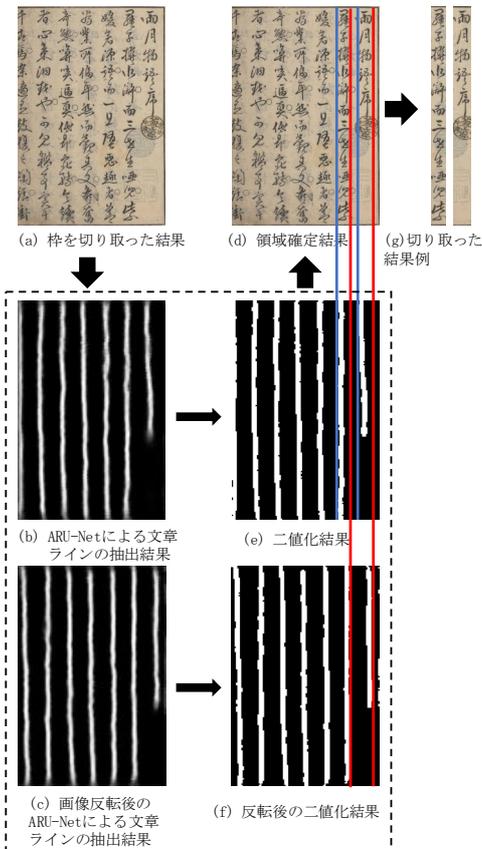


図 6. ARU-Net を用いた文字列のセグメンテーション

の最左座標(青線)と図6(f)の白線ごとの最右座標(赤線)により、その間の文章ラインを切り取って、図6(g)に示すように文章ラインが正しく得られる。

3.3. LeNet を用いた文字の認識

本研究では、LeNetを用いて、ARU-Netにより切り取ったくずし文字の認識を目指している。LeNetは、手書き文字認識として、よく知らせるモデルで、畳み込み層、プーリング層、全結合の5層により構成されている。しかし、LeNetは、一枚ずつの画像しか対応できないため、抽出された文章ラインでの自動認識ができない。そのため、我々は、文字のサイズを推定し、文章ラインでスライスし、文字の候補を切り取る。文字の候補をLeNetで認識し、確信度の高い文字候補は、認識結果とする。

図7は、抽出された文章ラインを用いて、LeNetによる文字認識例を示す。現時点で、文字は正方形と想定しているため、文字の縦軸と横軸のサイズが、図6で示された当該文字領域の赤線と青線の間の距離となる。図7の四角は、切り出された文字を示す。そして、切り出した文字列の先頭から(黄色四角)文字列幅の1/5をストライド長として切り出す。次に、切り出した文字をLeNetで文字を認識する。ここで、文章ラインをスライスし、上の二文字を三文字候補となる。信頼度の閾値は90%と定義する。信頼度が90%以上の文字は、認識結果とする。

図7では、それぞれ、LeNetで認識された第一番目の候補は、90.2%、50.4%、46.2%、90.5%、100%であるため、90.2%、90.5%と100%の高い確信度文字が認識結果となる。

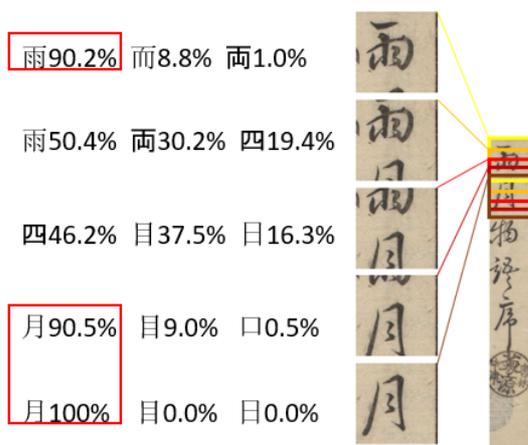


図7. 文字認識例

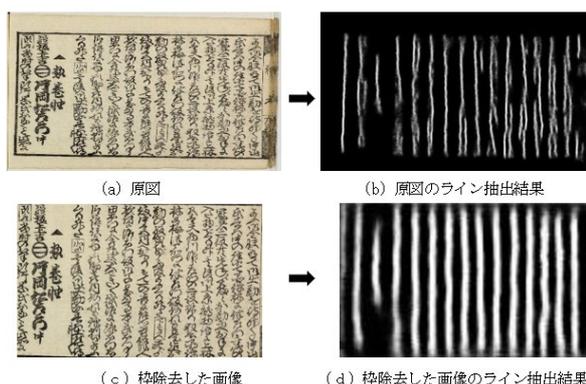


図8. ARU-Netによる文章のラインの抽出結果

4. 実験

4.1 ARU-Netによる文章のラインの抽出評価

本論文は立命館大学アート・リサーチセンター[1]所蔵の「役者神事競」(資料番号: arcBK04-0072、Book1とする)と「絵本義経一代実記」(資料番号: ナ04-0752、Book2とする)を用いて、ARU-Netによる文章ラインの抽出評価を行う。各本から10ページを選択し、再現率、適合率、F値を用いて、評価を行う。再現率は正しく抽出された文章ラインが全てのラインの割合で、適合率は、正しく抽出された文章ラインが抽出されたライン(正しく抽出された文章ラインと間違った抽出された文章ラインの和)の割合である。

図8は、ARU-Netによる文章のラインの抽出結果を示す。原図をARU-Netを用いてテキスト抽出を行った結果、枠の問題と1行の文字が2行に認識されてしまう問題が多くの問題点が見られた。そこで枠を除去し画像を縮小した後に再びARU-Netでテキスト抽出を行ったところ、結果が図8(d)のように良くなった。

Book1において、枠を除去する前に、再現率は99.1%、適合率は78.4%、F値は87.4%である。図9は、Book1とBook2の枠処理後の評価結果である。

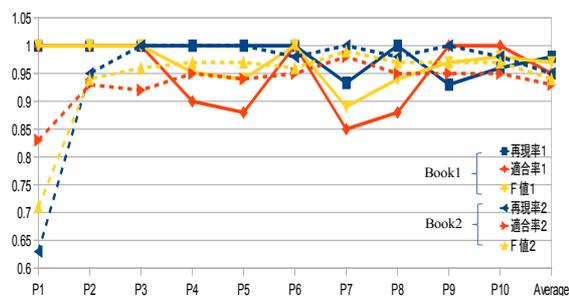


図9. ARU-Netでのライン抽出評価

U+306E の 2016	U+3044 い 1514	U+304B か 1503	U+306A な 1336	U+306F は 1330
U+3089 ら 1160	U+307E ま 1079	U+306B に 1073	U+300C 「 1072	U+3068 と 1043
U+3064 つ 952	U+3078 へ 945	U+308B る 912	U+308A り 884	U+3060 だ 869

(a) 文字種と文字数 (一部)



(b) トレーニング画像例

図 10. 「浮世風呂」文字例

る。処理後の平均再現率は 98.3%に達し、適合率は 95.2%に達し、F 値は 96.6%に達した。未処理の画像と比較して、Book1 の F 値は 10%近く増加し、Book2 の F 値は 2%増加する。

再現率が 1 に近いことは、文章のラインが正しく抽出されていることを意味する。しかし、適合率と F 値は、1 に 5%ぐらいの差がある。それは、ノイズや文章ではない絵は文章として抽出されてしまったケースが存在することである。従って、ノイズを除去した後は大幅に改善されたが、1 には達しないため、画像のノイズを完全に除去するのは将来の課題である。

4.2 LeNet を用いた文字認識

本論文では、人文学オープンデータ共同利用センター[3]の中から本(「浮世風呂」、文書番号:200015779)を用いて、LeNet の性能を評価した。図 10(a)は「浮世風呂」の 1 文字あたりの数であり、この本は合計 1817 文字種、総文字数は 60381 である。図 10(b)はトレーニングデータセットの例である。LeNet のトレーニング画像とテスト画像は 28*28 のグレースケール図であるため、トレ



図 11. トレーニング画像拡張例

ーニング画像とテスト画像に対して 28*28 のグレースケール図になるように前処理を行う。実験 OS は Ubuntu、プログラム言語は Python である。

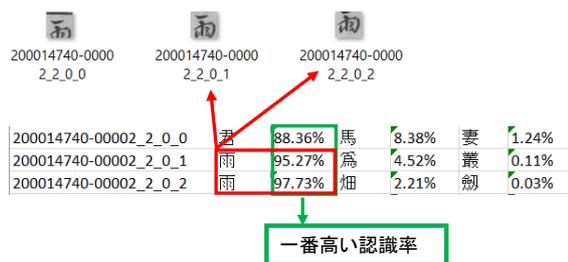
まず、学習データについて、クラス 10 から 50 まで、10 個ずつ増加して、評価を行った。学習データは、日本古典籍くずし字データセットを使用する。

表 1 には、クラス毎のトレーニング画像数、テスト画像数、学習時間と認識率を示す。それぞれ 90%ぐらいの識別率に達成した。

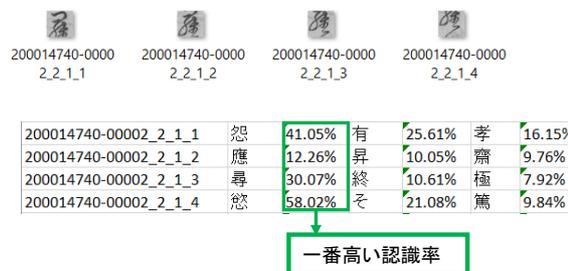
表 1 LeNet 認識率結果

クラス	トレーニング画像数	テスト画像数	エポック	時間	認識率
10	19323	1000	200	0:10:35	0.930
20	28412	2000	200	0:32:28	0.884
30	35165	3000	300	1:06:44	0.899
50	41425	5000	500	1:49:01	0.890

LeNet の認識率が高いため、LeNet を用いて「浮世風呂」の文字認識を行った結果例を図 12 に示す。トレーニングデータが不足しているため、データセットを増加する。図 11 のように 1 枚のトレーニング画像の文字を上、下、左、右計 4 方向に、それぞれ 25 と 50 ピクセル移動させている。1 枚は 9 枚に拡張する。図 12 から分かるように、“雨”の認識率はそれぞれ 95.27%と 97.73%であり、この文字が“月”であると推測される。しかし正解が認識できない場合もあり、図 12(b)のよ



(a) 認識した結果例



(b) 認識しなかった結果例

図 12. LeNet で文字認識結果例

うに文字が複雑な場合もあるため、文字を正しく認識できない。

4.3 議論

文字認識を行う際に枠外に切り出される余白は時間とメモリのスペースを浪費するため、枠を取り除くことで文字認識のスペースを節約するだけでなく、文字の認識率も向上する。しかし、90%が認識率の閾値として適切であるか、また、文字の切り出し形状が正方形であるかの調整が必要であるか、更なる実験の証明と最適化を求める必要がある。さらに、LeNet の性能実験では現在の種類が 50 種類しかないという問題もあり、今後の課題である。

5. おわりに

本論文では、日本の古典籍文字の自動認識システムを提案し、一連の実験を行った。まず、ARU-Net を用いて日本の古典籍の文章の自動的抽出を紹介し、その有効性を示した。評価において、再現率は 1 に達成できたが、適合率と F 値は 1 に達成できなかった。つまり、文章の部分が正しく抽出されたが、手書きの絵が文章部として判定されたケースが存在する。次に LeNet の性能を評価したところ、50 クラスに増加した場合の LeNet の認識率は 89%と、LeNet の認識率が高いことが示された。しかし、50 個の種類が少ないため、種類を増やすことを今後の重要な課題となる。また、LeNet は「浮世風呂」中画像に簡単な文字しか認識できず、複雑な文字に対して、LeNet はうまく認識できないため、ネットワークモデルの更なる改造を行う必要がある。

6. あとがき

本研究は、立命館大学のアート・リサーチセンターの助成で行うので、ここで謝意を表す。また、助言を頂いた、立命館大学の文学の赤間亮先生に謝意を表す。本研究の一部は JSPS 科研費 18K18337 の助成を受けたものです。

参考文献

- [1] 北本 朝展, “日本古典籍字形データセットの公開と活用への期待”, <http://agora.ex.nii.ac.jp/~kitamoto/research/publications/codh17b-ppt.pdf> (2019.11.22 accessed)
- [2] T. Gruning, G. Leifert, T. Strauß, J. Michael and Roger Labahn, “A Two-Stage Method for Text Line Detection in Historical Documents,” *International Journal on Document Analysis and Recognition (IJ DAR)*, Vol.22, No.3, pp.285-302, 2019.
- [3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document

recognition,” *Proceedings of the IEEE*, Vol.86 No.11, pp.2278–2324, 1998.

- [4] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu and A. C. Berg, “SSD: Single Shot MultiBox Detector,” *Computer Vision – ECCV 2016*, pp.21-37, 2016.
- [6] L. Meng, B. Lyu, Z. Zhang, C. V. Aravinda, N. Kamitoku and K. Yamazaki, “Oracle Bone Inscription Detector Based on SSD,” *New Trends in Image Analysis and Processing – ICIAP 2019*, pp.126-136, 2019.
- [7] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet classification with deep convolutional neural networks,” In: *Advances in Neural Information Processing Systems (NIPS 2012)*, vol. 25, 2012.
- [8] C. Szegedy, W. Liu, Y.Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going Deeper With Convolutions,” *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.
- [9] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for Large-scale Image Recognition,” *International Conference on Learning Representations (CLR2015)*, 2015.
- [10] 日本古典籍くずし字データセット, 人文学オープンデータ共同利用センター, <http://codh.rois.ac.jp/char-shape/book/http://codh.rois.ac.jp/> (2019.11.22 accessed)
- [11] 古典籍ポータルデータベース, 立命館大学アート・リサーチセンター: <https://www.arc.ritsumei.ac.jp/database.html> (2019.11.22 accessed)
- [12] B. Lyu, R. Akama, H. Tomiyama, and L. Meng, “The Early Japanese Books Text Line Segmentation,” 2019 International Conference on Advanced Mechatronic Systems, 2019.
- [13] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.