

# 深層学習を用いた書道作品における書体翻訳

二ノ宮 梢平<sup>†</sup>中山 雅紀<sup>†</sup>宮澤 篤<sup>‡</sup>藤代 一成<sup>†</sup><sup>†</sup>慶應義塾大学 理工学部情報工学科<sup>‡</sup>慶應義塾大学 大学院理工学研究科

## 1 背景と目的

書道の作品に用いられる書体は篆書、隸書、草書、行書、楷書の5種に大別される。これらの多くは識字に専門の知識が必要とされる。これは書道作品の鑑賞を初心者にとって困難なものにする最も大きな要因のひとつである。

こうした難読文字の識字を目的とした研究の例として、日本の古文書などの書物におけるくずし字を対象とした研究 [1] が挙げられる。この研究にも採用されているように、識字のアプローチとして深層学習を用いることは有効であると考えられる。

本研究では、書道作品の鑑賞支援として、書道文字の識字と他の書体への変換を提供するシステムの設計を試みる。本稿では、識字処理へのアプローチとして畳込みニューラルネットワークを用いた手法を提案する。漢字をパーツに分解し、パーツごとに識字処理を施すことで、書道作品に用いられる難読文字の識字を可能にする。ここでのパーツとは、偏旁冠脚による構成部分となる部首の候補をさす。なお、本稿では篆書を識字の処理対象とする。篆書は偏旁冠脚が移動し、パーツの組合せによって別の文字になることから、本手法の有用性の調査に適しているためである。

## 2 概要

本手法は、漢字がパーツから構成されるという特徴を利用し、漢字に対してパーツ単位での識字処理を施す。また、識字に至るまでの流れは、入力画像に対する識字処理の対象領域の選択、画像の前処理、識字処理、処理結果の提示、文字検索からなる。その処理の概略を図1に示す。なお、本稿では識字処理に注力し書体変換には既存の書道古典の画像を利用したが、今後書道作品の特徴を考慮した変換手法を検討する予定である。

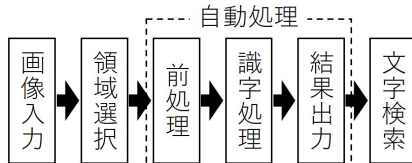


図1: 識字までの処理の概略。パーツを対象に識字処理を施す。そこで、入力画像の対象となるパーツの領域を選択し、前処理をしたうえで識字処理を実行する。その結果から漢字を推定する

## 3 提案手法

識字処理の詳細とその前後の処理について説明する。

### 3.1 識字処理

識字を行うための学習アルゴリズムとして、畳込みニューラルネットワークを用いる。また、少ない学習データに対して、識字の精度を上げるために学習済みモデルであるVGG16[2]を微調整する。まず、VGG16モデルは出力層が1,000クラスあるため、本手法におけるクラス数に変更した。損失関数に交差エントロピー誤差を、最適化手法にMomentum SGDを用いた。また、各層のパラメタの学習率は、入力層に近い順に16個の畳込み層で0.01%、2個の全結合層で0.05%、そして最後の全結合層で0.1%と設定した。ここで、入力に近い層は汎用的な特徴を抽出するため学習率を小さくし、出力に近い層は専用の特徴を抽出するため学習率を大きくした。

### 3.2 データセット

本手法では書体を篆書に限定し、表1のパーツを含むデータセットを作成した。ここで、複数の分類に所属するパーツもひとつのクラスとする。よって、9個のクラスからデータセットが構成される。また、篆書の古典は現存するものが少ないため、データ拡張としてノイズを付与した画像とトリミングを施した画像をモデルの訓練時に使用した。データ拡張後の学習用画像の枚数を表2に示す。

表1: データセットの内容

分類	内容
偏	いとへん, おんなへん, きへん, てへん, さんずい
旁	いと, おんな
冠	くさかんむり, たけかんむり
脚	おんな, き, こころ, みず
繞	しんによる

表2: データセットの画像枚数

クラス	画像枚数
いと(へん)	524
おんな(へん)	272
き(へん)	476
くさかんむり	1,281
こころ	432
さんずい・みず	544
たけかんむり	602
てへん	284
しんによる	212

Script translation in calligraphic works using deep learning  
Shohei Ninomiya<sup>†</sup>, Masanori Nakayama<sup>†</sup>, Atsushi Miyazawa<sup>‡</sup>, Issei Fujishiro<sup>†</sup>

<sup>†</sup>Department of Information and Computer Science, Keio University

<sup>‡</sup>Graduate School of Science and Technology, Keio University

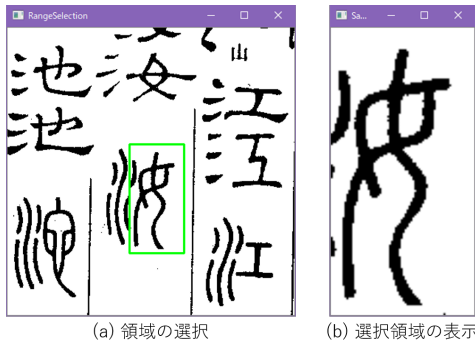


図 2: 対象領域の選択と表示. 入力画像に対してマウスをドラッグすることで識字処理の対象となる領域を選択する. 入力文字画像の典拠: <https://www.benricho.org/Unchiku/5taijirui/>

### 3.3 対象領域の選択

ユーザによって入力画像における識字処理の対象領域が選択, 保存される. ここでの対象領域とは, 1 パーツ以上を含む領域である. 図 2 (a) に示すようにマウスのドラッグで作成された矩形が対象領域となる. また, 図 2 (b) に示すように保存の前に選択領域だけを表示, 確認することも可能である.

### 3.4 前処理

前項の操作により選択された領域のサイズは不揃いであるが, 3.1 項のモデルに適したサイズは  $224 \times 224$  画素である. したがって, 選択領域のアスペクト比を維持したまま, 縦幅と横幅で大きい方の画素数が 224 画素となるようにサイズを変更し, 他を余白とする. また, 学習用の画像は白と黒の二値画像を用いたため, 選択領域に対しても二値化処理を施す.

### 3.5 文字検索

図 3 に示すように, ユーザは選択領域の処理結果をもとにパーツごとに検索をかけることで, そのパーツを含む漢字を絞り込むことができる. さらに, その漢字について検索をかけることで, 他の書体での表現が提示される.

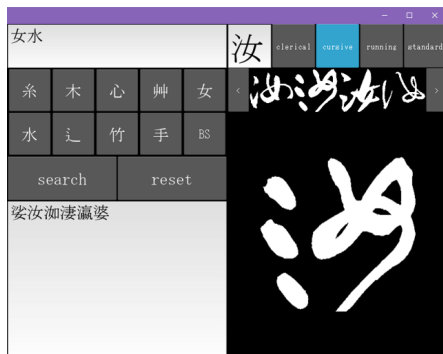


図 3: 文字検索機能. パーツを入力するごとにそれらを含む漢字の絞り込みが可能である. また, 漢字を入力することで他の書体での表現を確認できる.

出力文字画像の典拠: <http://www.shufazidian.com/>

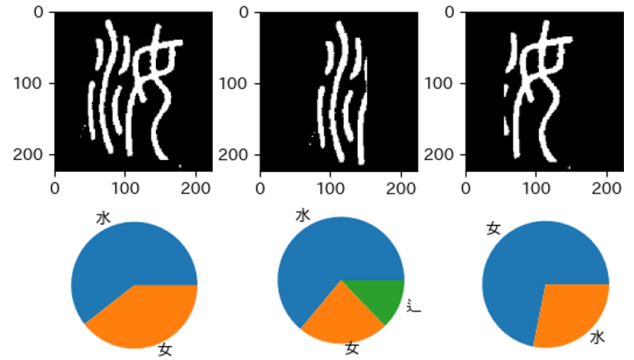


図 4: 識字処理の出力例. 選択領域ごとの処理結果を表示する

## 4 結果と評価

3.1 項におけるモデルの識字の精度を評価したところ 97.0%を示した. ここでは, 各クラスにおいてデータ拡張以前の 10%の画像を検証目的で使用した. また, システムにおける処理結果を図 4 に示す. 入力画像として図 2 と同様の画像を用いた. 識字の候補が 4 個以上の場合は上位 3 個を結果として採用し, それ以下を切り捨てたうえでパーツを含む割合を提示する. この例では, すべての画像において正確な処理結果を示した. しかし, 処理結果として提示された文字パーツが画像のどこに対応するかが示されていないため, 出力された候補のうち, どれが実際に存在するパーツであるかを判断するために必要な要素が不足している.

## 5 結論と今後の課題

本稿では, 文字画像に対して畳込みニューラルネットワークによる識字処理をパーツごとに行う方法を提案し, それに伴う識字の一連の流れを示した. これにより, 1 漢字 1 クラス対応の識字処理に対して, 少ないデータ数とクラス数でより汎用的な識字の処理環境が実現できた.

課題としては, データセットの拡張とそれに伴う処理精度の検証, 篆書以外の書体に対する識字処理の適用, 単一の文字ではなく書道作品全体としての識字処理の実装, 領域選択や文字検索におけるユーザインタフェースの改良, 処理結果の提示方法の改良が挙げられる. また, 書体の変換としては, 現段階では既存の書道古典の画像を使用しているが, 敵対的生成ネットワークを利用することにより多様性を向上させることを検討する.

### 謝辞

本研究の一部は, 令和元年度科研費基盤研究 (A)17H00737 の支援により実施された.

### 参考文献

- [1] 北本 朝展, カラーソフト タリン, 宮崎 智, 他: “文字データの分析-機械学習によるくずし字認識の可能性とそのインパクト-”, 電子情報通信学会誌, Vol. 102, No. 6, pp. 563-568, 2019
- [2] K. Simonyan and A. Zisserman: “Very deep convolutional networks for large-scale image recognition,” *arXiv*, 2014.