

文字の形状情報の学習を抑制することで個人性を強調する CNN 単漢字筆者識別  
CNN single-character writer identification that emphasizes individuality information by suppressing learning of character shape information

上松 和史<sup>†</sup> 荒井 秀一<sup>†</sup>  
Kazushi Uematsu Shuichi Arai

### 1. はじめに

人間の手書き文字には字画構成や誤字、運筆など様々な固有筆癖が存在する。筆者に内在する固有筆癖を図 1 に示す。



図 1 様々な筆者の「城」という文字

図 1 はそれぞれ異なる筆者が書いた文字画像である。本来ならば線が繋がらない部分が運筆によっては繋がっていたり、はねるべき収筆に波勢がない筆者など、人によって様々な固有筆癖が存在する。この固有筆癖を用いて筆者識別の研究が行われており、古文書の筆跡鑑定やオンライン署名の自動化など様々な分野への応用が期待されている。従来、収筆や斜勢など人間が特徴量を定義した筆者識別法が提案されていたが、文字に現れる特徴量があまりにも多いため、人間では定義しきれないという問題が存在していた。近年、深層学習の発達によりこれまで人間が定義していた特徴量をコンピュータが自動的に定義する手法が研究されている。しかし、従来の深層学習における筆者識別法では、物体認識モデルで用いられたモデルを使用しているため、識別器は文字の詳細な形状を表すのに十分なマップサイズを持っている。従来研究では、文字認識としての学習を避けるために入力文字画像を複数化し、文字の詳細な形状をぼかす手法が提案された。しかし、入力文字画像の複数化は文字に含まれる固有筆癖を増加させているため、識別器の精度が向上するのは当然である。本稿では、単一文字画像入力 CNN(Convolutional neural network)の深層化により詳細な形状を表現できないほどマップサイズを縮小する手法を提案し、CNN 識別器が文字認識のような学習から固有筆癖を特徴とした筆者識別を行うように変化したことを報告する。

### 2. 従来研究

深層学習を用いた筆者識別法として、CNN による筆者識別法が提案された[1]。深層学習は、コンピュータが自動的に特徴抽出を行うため、これまで人間が定義しきれなかった多くの特徴量を自動的に定義できるようになった。しかし、この筆者識別法は物体認識としてのモデルを利用しているため、文字の詳細な形状を認識する文字認識としての学習を行ってしまうという問題が存在する。CNN モデル

<sup>†</sup> 東京都市大学(Tokyo City University)

の判断基準箇所を可視化する grad-CAM(grad-Class Activation Map)によって可視化した結果を図 2 に示す[2]。

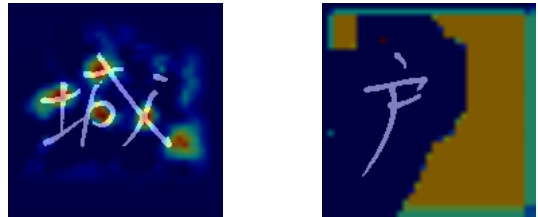


図 2 grad-CAM による可視化結果

図 2 はヒートマップであり、図中の赤い部分が特徴抽出箇所を表している。図 2 から、識別器は文字の局所的な部分の特徴抽出箇所としている場合と、文字そのものではなくその周囲から特徴抽出をしている 2 つのパターンで識別を行なっていることがわかる。この結果から、識別器は文字に内在している固有筆癖を識別しているのではなく、概形や輪郭といった大まかな形状を特徴とする文字認識のような学習をしており、これは人間の識別方法とは大きく異なる。

文字の概形からの特徴抽出を避ける筆者識別法として、CNN に入力する漢字画像を複数化する筆者識別法が提案されている[3]。入力画像の複数化により、文字の詳細な形状をぼかし文字認識のような学習の回避を目的としているが、入力する文字画像が増加すれば内在する固有筆跡も増加する。そのため、精度が向上するのは当然であり、固有筆癖を特徴とした筆者識別を行なっているとは言い難い。本稿では文字の概形を認識する文字認識としての学習を抑制し、固有筆癖を特徴として筆者識別を行う識別器の提案を目的とする。

### 3. 提案手法

識別器が概形からの学習ではなく固有筆癖を特徴抽出するために、本稿では単漢字入力 CNN 識別器の深層化を行う手法を提案する。CNN の深層化によるマップサイズの変化を図 3 に示す。図 3 上部の従来手法は、単一入力 CNN[1]の畳み込み層から出力される特徴マップのサイズを表している。図 3 中の従来手法では、最終畳み込みによって得られるマップサイズが、ドットフォントで表せる最小画素 16×16 よりも大きいため、文字に対して詳細な形状が表現可能であった。そのため、識別器は文字に対して形状を記憶してしまうのではないかと考察した。そこで、本稿では畳み込み層を 2 層追加することで深層化を行い、最終畳み込み層によって得られるマップサイズをドットフォントで表せる最小画素 16×16 ピクセルよりも縮小した。マップサイズの縮小により、単漢字入力 CNN において文字の詳細な形状の表現を明示的に抑制する。

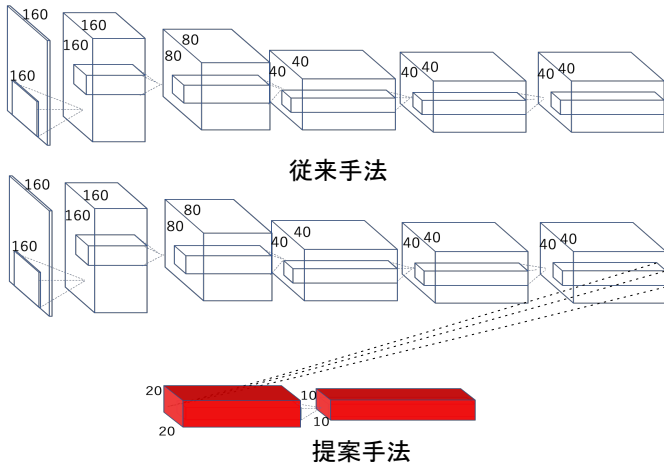


図3 アーキテクチャの変更によるマップサイズの変化

4. 実験条件

実験で用いるネットワーク構造を図4に示す。

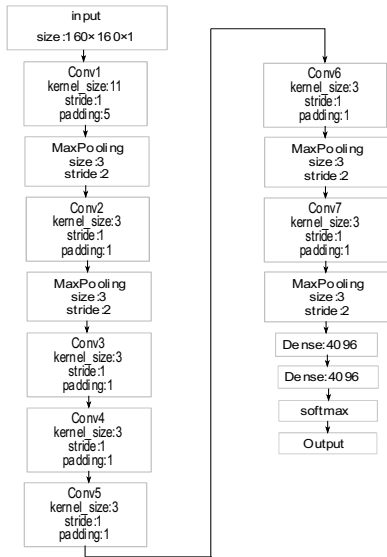


図4 ネットワーク構造

学習および評価に用いるデータセットは100人の筆者による100字種の文字を20回ずつ書いた計200,000枚の文字画像データセットを用いる。識別は100人の筆者で行い、90字種、20回分を用いた計180,000枚の文字画像で学習を行い、残りの10字種の漢字が書かれた20,000枚の画像を評価に使用する。評価条件として、softmax関数で最も値の大きいtop-1の精度を指標とする。

5. 実験結果

実験結果を表2に示す。比較のため、マップサイズが16×16より大きい場合との比較も示す。

表2 マップサイズの変化による精度比較

最終マップサイズ	精度(%)
20×20	72.3
5×5	75.3

最終マップサイズが20×20の場合と比較して、3.0pt精度が向上していることから、この手法の有効性が確認できた。また、今回の提案手法における特徴抽出箇所の違いをGrad-camによる可視化を行い図4、図5で示す。

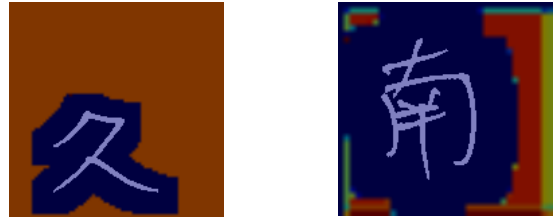


図4 最終マップサイズ20×20の特徴抽出箇所の可視化

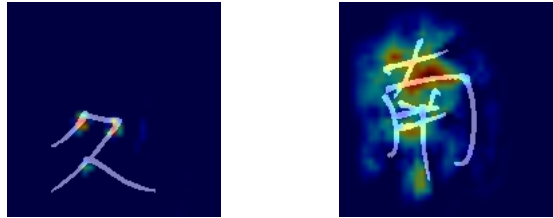


図5 最終マップサイズ5×5の特徴抽出箇所の可視化

可視化結果から、詳細な形状の表現の抑制によりこれまで概形から特徴抽出を行っていた文字画像に対して文字の局所的な部分から特徴抽出を行なっていることが明らかになった。

6. さいごに

本稿では、識別器が文字の詳細な形状を学習してしまうという問題を回避するためにCNNの深層化を行い明示的に詳細な形状の学習を抑制する手法を提案した。また、提案手法と従来手法の精度を比較し、性能の向上に関して基本的な有効性を確認できた。また、抑制により、識別器の特徴抽出箇所が、文字の概形から固有筆癖への変化したことを確認した。

謝辞

謝辞と参考文献の見出しは

「FIT謝辞と参考文献の見出し」に設定済み

参考文献

[1] Nasuno, Ryosuke, and Shuichi Arai. "Writer identification for offline japanese handwritten character using convolutional neural network." *Proceedings of the 5th IIAE (Institute of Industrial Applications Engineers) International Conference on Intelligent Systems and Image Processing*. (2017)

[2] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. (2017)

[3] Nguyen, Hung Tuan, et al. "Text-independent writer identification using convolutional neural network." *Pattern Recognition Letters*121 (2019): 104-112.