

## 畳み込みニューラルネットワークを用いた オフライン毛筆書文字の定量評価

吉田直哉<sup>†</sup> 大山航<sup>†</sup>埼玉工業大学<sup>†</sup>

### はじめに

毛筆書写は基本的な文化活動のひとつである。国内では、書の芸術性とは別に形の整った読みやすい毛筆文字を書く毛筆書写教育が行われている。教育現場においては毛筆書写文字の評価は重要である一方、文字の評価は主観的な心理評価尺度によって行われることが多い。心理評価は評価者ごとに基準が曖昧になりがちであり、付与された評価に対する説明生や納得性が低下しやすい問題がある。

手書きされた書文字の定量評価は、書写学習の支援だけでなくアミューズメントにも活用できる。例えばポータブルゲーム端末 Nintendo DS では、美文字トレーニング<sup>[1]</sup>がある。しかし、オフライン型の手法については研究が少なく実用化されていない。

手書き文字の定量評価に関する研究にはいくつかの例がある。飛谷らは筆あと計測に基づく習字学習教育支援システム<sup>[2]</sup>を提案した。この手法は筆記の過程に着目したオンライン型手法であるため、オフライン情報である筆記済み手書き文字を対象とした熟練度の評価は不可能である。オフライン型手法の例として古性らの研究<sup>[3]</sup>がある。この研究では書の芸術性に注目して評価する一方、形状の整った文字に対しては評価の精度が低い結果が示された。

オフライン文字の定量評価手法として、石原<sup>[4]</sup>は字形からその文字の心理的評価値を回帰する手法を提案した。この手法は特定文字（「永」の字）を対象にした手法であり、回帰に用いる説明変数と目的変数とにそれぞれ、濃度こう配特徴と文字のバランスを考慮した字形特徴、心理実験により付与されたその文字の心理的評価値を採用した。また、この研究では、オフライン文字定量評価の実験に利用できるデータセットが整備された。

著者らによる前報<sup>[5]</sup>では、深層学習手法である畳み込み自己符号化器で字形特徴を抽出し、抽出された字形特徴を説明変数として心理評価値を回帰する手法を提案したが、小規模な実験用データセットにおいて石原らの手法に対して十分な有効性を示すことはできなかった。

本報告では、字形評価の問題設定を、回帰問題から分類問題に変更する。すなわち、組み合わせて提

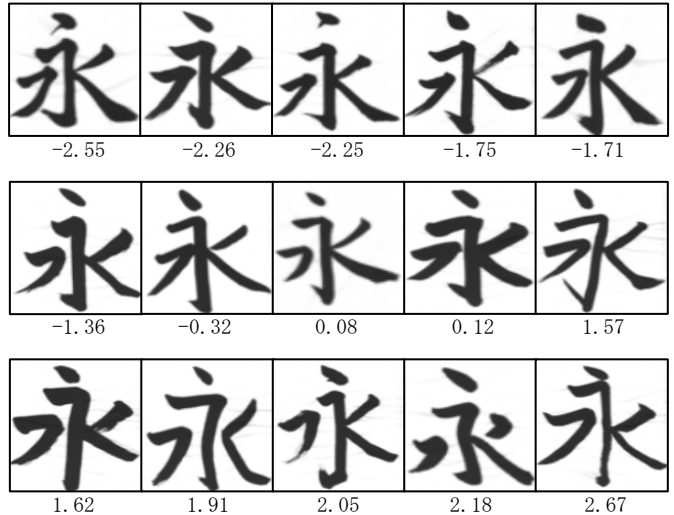


図1 書文字画像の例:

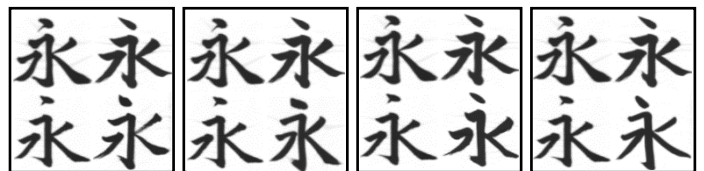


図2 画像セットAから4つ選んで並べた画像の例

示された文字のうち、最も心理評価値が高い（上手い）文字を選択する分類器を機械学習で実現する。

### 提案手法

本研究では、組み合わせて提示された文字の識別と投票に基づく毛筆書文字評価手法を検討する。実験用データセットには、毛筆書文字 30 例（15 例×2 セット）に対してサーストンの一対比較実験で算出された心理評価値が付与されている。図1に1セットの例を示す。それぞれも文字に併記されている値が心理評価値であり、心理評価値が小さいほど上手な文字であることを表す。それぞれのセットから4例の文字を取り出し、図2のように配置した画像を分類器に入力する。分類器は、4例のうち最も高い評価値を持つ文字の位置を出力、すなわち4クラス分類器となる。分類器の出力に基づき文字画像に対して投票を行い、最終的な得票数に基づき文字画像を順位づけする。

### データセット

本研究で使用するデータセットは、書道有段者から初心者までの様々な熟練度の筆記者が書写用半紙

Quantitative Evaluation of Off-line Brush Handwritten Characters using Convolutional Neural Network

<sup>†</sup>Saitama Institute of Technology

本研究は JSPS 科研費 21K11942 の助成を受けたものです。

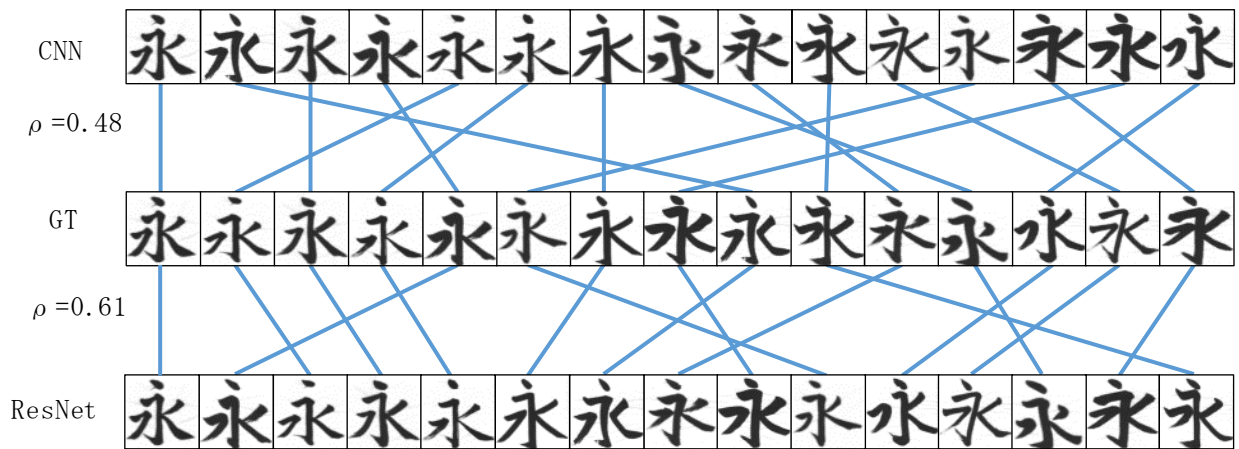


図 3 CNN による識別・投票に基づく順位予測

に毛筆で書写した「永」の字を 300dpi でスキャンした画像のデータセットである。30 例の画像を 15 例ずつ 2 つのセット (画像セット A, 画像セット B) に分ける。画像セット A, 画像セット B にはそれぞれ心理評価値が付与されている。心理評価値はサーストンの一対比較法で算出した。

本研究では、15 例の「永」の字から 4 例を選んで並べるすべての順列で組み合わせ画像を作成した。画像データセット A, B それぞれにおいて 32760 例の組み合わせ画像を生成した。

### 評価実験

画像セット A を用いて、畳み込みニューラルネットワーク (CNN) の学習を行った。CNN は、表 1 に示す構成のものと ImageNet によって事前学習された ResNet50<sup>[6]</sup>を用いた。学習が完了した CNN を用いて画像セット B の組み合わせ画像を識別、投票し、得票数に応じて文字画像の順位予測を行なった。

順位予測の性能評価基準にはスパイマンの順位相関係数を用いた。スパイマンの順位相関係数  $\rho$  は次式で定義される。

$$\rho = 1 - \frac{6 \sum D^2}{M^3 - M}$$

ここで、 $M$ ,  $D$  はそれぞれ、データセットのサンプル数、真の順位と予測順位の差である。

### 結果及び考察

表 2 に、それぞれの CNN の識別成功率とスパイマンの順位相関係数を示す。事前学習済 ResNet を用いた場合の方が識別成功率および順位相関ともに大きくなった。

図 3 に、画像セット B における真の順位 (GT) と、各 CNN による予測順位を示す。中央が GT、上段、下段がそれぞれ CNN, ResNet による予測順位である。

図中左側ほど上手な字である。識別成功率が高かった ResNet の方が文字の順位変動が小さく、順位相関も高くなった。

表 1 本研究で用いた CNN の構成

層の種類	特徴マップのサイズ
出力層	4
プーリング	32×32×16
畳み込み	64×64×16
プーリング	64×64×16
畳み込み	128×128×16
プーリング	128×128×16
畳み込み	256×256×16
入力層	256×256×1

表 2 識別成功率とスパイマンの順位相関係数

	識別成功率	順位相関
CNN	0.65	0.48
ResNet	0.77	0.61

### おわりに

本研究では、毛筆で書かれた文字に対して字の熟練度 (上手さ) を定量評価する手法を提案した。今後は CNN の識別性能の向上、GradCAM<sup>[7]</sup>等の CNN 内部解析手法を用いた識別根拠の可視化に基づく考察などを行う予定である。

### 参考文献

- [1] NintendoDS “美文字トレーニング” <http://www.nintendo.co.jp/ds/avmj/index.html>
- [2] 飛谷 他, 習字教育支援システムのための Horizon View Camera を用いた筆あと計測手法の提案と実装, 電気学会論文誌 C, Vol.129, No.5, pp.786-791, 2009
- [3] 古性 他, 感性情報に基づく毛筆漢字文字の美的評価, 信学技報, Vol.101, No.201, pp.81-88, 2001
- [4] 石原 他, オフライン毛筆書写文字の定量的質評価, 画像センシングシンポジウム, 2018
- [5] 吉田 他, 畳み込み自己符号化器を用いた毛筆書写熟練度の定量評価, MIRU2019
- [6] He et al., Deep Residual Learning for Image Recognition, CVPR2016
- [7] Selvaraju et al., Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, ICCV2017