

## 手書き数字の筆者識別に向けた CNN の構成方式に関する研究 A study on CNN architecture for writer identification of handwritten digits

木場迫 昇二<sup>†</sup>      黒木 修隆<sup>†</sup>      沼 昌宏<sup>†</sup>  
Shoji Kobasako      Osataka Kuroki      Masahiro Numa

### 1. はじめに

一般に、筆跡鑑定は鑑定士の主観的な判断に依存しやすく、結果が安定しにくい問題がある。一方、CNN を筆跡鑑定に応用すれば、人の主観に左右されにくい安定した結果が得られる可能性がある。しかし、筆者すら意識していない潜在的な特徴を認識することは、単なる文字認識に比べて非常に難易度が高い。

そこで、本稿では筆跡鑑定を文字認識タスクと筆者識別タスクの 2 つに分け、これらを組み合わせる方法について検討を行ったので報告する。

### 2. 提案手法

本稿では 3 種類の CNN の構成方法を提案し、後の実験で比較する。

#### 2.1 手法 A

手法 A の構成を図 1 に示す。この構成では画像  $I$  を単一の筆者推定用 CNN(CNN for Writer)に入力することで、

$$\mathbf{p} = \text{CNN for Writer}(I) \quad (1)$$

のようにそれぞれの筆者の確率群  $\mathbf{p}$  を得る。この  $\mathbf{p}$  の中で最大値を持つラベルを  $I$  の筆者として出力する。

#### 2.2 手法 B

手法 B の構成を図 2 に示す。まず、画像  $I$  を数字推定用 CNN(CNN for Number)に入力することで、

$$\mathbf{n} = \text{CNN for Number}(I) \quad (2)$$

のように 0~9 の各数字に対する確率群  $\mathbf{n}$  を得る。この  $\mathbf{n}$  の中で最大値を持つラベル  $i(i=0\sim 9)$  を  $I$  の数字として出力する。

次に、数字 “ $i$ ” 専用の筆者推定 CNN である  $\text{CNN}_i$  を用いて、

$$\mathbf{p} = \text{CNN}_i(I) \quad (3)$$

のように筆者の確率群  $\mathbf{p}$  を得る。この  $\mathbf{p}$  の中で最大値を持つラベル(筆者)を  $I$  の筆者として出力する。



図 1 手法 A の構成

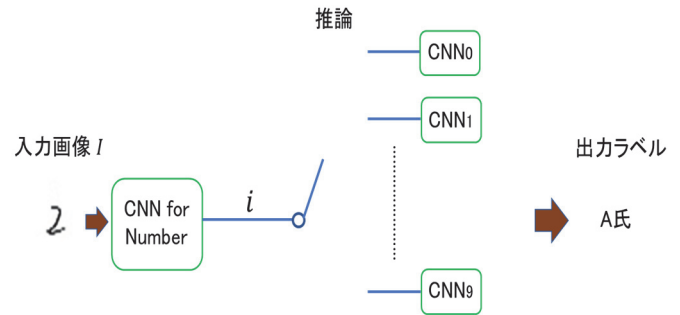


図 2 手法 B の構成

#### 2.3 手法 C

手法 C の構成を図 3 に示す。まず、手法 A と同様に式 (2) を用いて、各数字に対する確率群  $\mathbf{n}$  を得る。次に、 $I$  とすべての  $\text{CNN}_i(i=0\sim 9)$  を用いて、

$$\mathbf{p}_i = \text{CNN}_i(I) \quad (4)$$

のように筆者の確率群  $\mathbf{p}_i$  を 10 個得る。そして、確率群  $\mathbf{n}$  の  $i$  番目の要素を  $n_i$  として、

$$\mathbf{p} = \sum_{i=0}^9 \mathbf{p}_i n_i \quad (5)$$

とすることでアンサンブルによる筆者の確率群  $\mathbf{p}$  を得る。この  $\mathbf{p}$  の中で最大値を持つラベルを  $I$  の筆者として出力する。

### 3. 評価実験

#### 3.1 実験内容

手法 A, B および C を用いて、手書き数字の筆者推定を行った。データセットには 10 人分の手書き数字を用いた。データセットの画像の一例を図 4 に示す。画像 1 枚には 1 つの数字が書かれており、 $28 \times 28$  pixel で構成されている。1 人あたり “0” から “9” の数字がそれぞれ学習用画像 300 枚、評価用画像 50 枚ずつ用意されており、合計で学習用画像 30,000 枚、評価用画像 5,000 枚で構成されている。学習回数は 93,700 回(epoch:100 batchsize:32)、評価指標は認識率とする。また、すべての手法において、CNN のモデルは LeNet-5 [1] を基にしたネットワーク構造を採用した。LeNet-5 の構造を図 5 に示す。

<sup>†</sup> 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

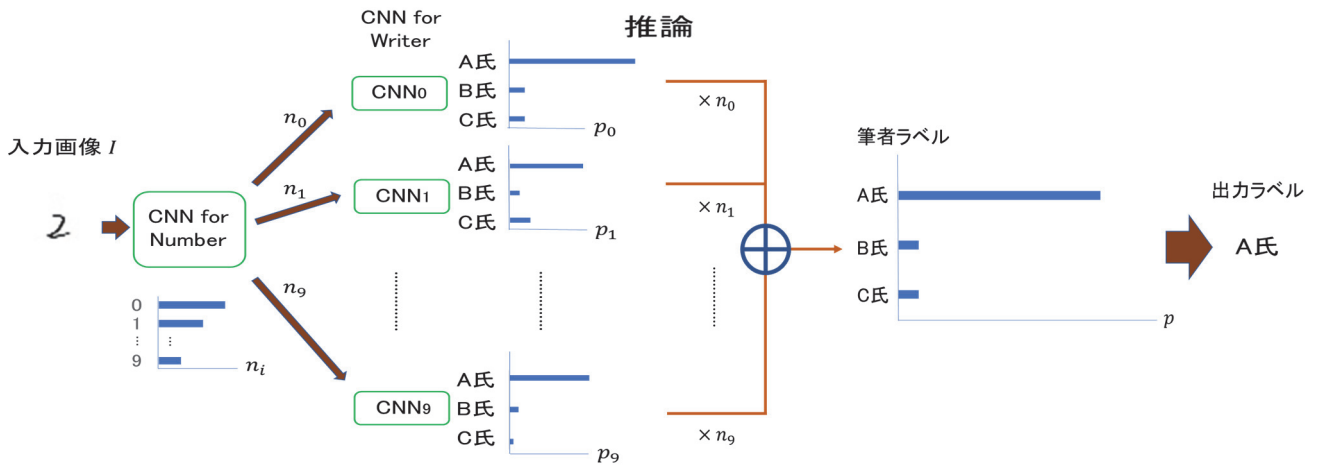


図 3 手法 C の構成

筆者	数字									
A氏	0	1	2	3	4	5	6	7	8	9
B氏	0	1	2	3	4	5	6	7	8	9
C氏	0	1	2	3	4	5	6	7	8	9
D氏	0	1	2	3	4	5	6	7	8	9
E氏	0	1	2	3	4	5	6	7	8	9
F氏	0	1	2	3	4	5	6	7	8	9
G氏	0	1	2	3	4	5	6	7	8	9
H氏	0	1	2	3	4	5	6	7	8	9
I氏	0	1	2	3	4	5	6	7	8	9
J氏	0	1	2	3	4	5	6	7	8	9

図 4 データセットの画像の一例

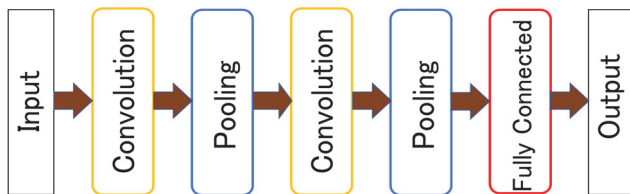


図 5 LeNet-5 の構造

### 3.2 実験結果と考察

筆者および数字の認識率を表 1 に示す。手法 B, C における数字の認識率は 99.8 % となった。そのもとで、後段の筆者識別を行った所、認識率が手法 A と比べて、手法 B では 3.8 pt, 手法 C では 3.9 pt 向上した。

次に、数字別の筆者の認識率を表 2 に示す。手法 A と比べて手法 B, 手法 C ではすべての数字において筆者の認識率が向上した。これらの結果から、先に数字推定を行うことの有用性を確認した。ただし、後段の筆者推定では手法 C が 10 個の CNN のアンサンブルを導入したにも関わらず、手法 B に対する改善が 0.1 pt であった。

一方、表 2 より、数字別の認識率では“1”のみすべての手法において 90 % を割る結果となった。これは、“1”が他の数字に比べて単純な形をしており、筆者ごとの特徴が現れにくかったためであると考えられる。

表 1 筆者および数字の認識率

評価手法	手法 A	手法 B	手法 C
筆者の認識率[%]	92.8	96.6	<b>96.7</b>
数字の認識率[%]	-	99.8	99.8

表 2 数字別の認識率

数字	認識率[%]		
	手法 A	手法 B (手法 A からの 増加幅)	手法 C (手法 B からの 増加幅)
0	87.6	93.6 (+6.0)	93.6 (±0)
1	<b>80.6</b>	<b>87.0 (+6.4)</b>	<b>87.0 (±0)</b>
2	94.6	97.2 (+2.6)	97.2 (±0)
3	91.2	97.6 (+6.4)	97.6 (±0)
4	96.8	99.2 (+2.4)	99.2 (±0)
5	98.4	99.8 (+1.4)	99.8 (±0)
6	94.0	97.2 (+3.2)	97.2 (±0)
7	95.2	97.8 (+2.6)	97.8 (±0)
8	95.2	99.0 (+3.8)	99.0 (±0)
9	94.8	98.0 (+3.2)	98.2 (+0.2)

### 4. まとめ

本稿では、手書き数字の筆者の認識率の向上を目的とし、3 種類の CNN の構成方法を提案した。

実験の結果、単一の CNN しか用いない手法 A と比べて手法 B では 3.8 pt, 手法 C では 3.9 pt 向上した。これより、一段目に数字認識を行うことの有用性を確認できた。

今後は、受験番号や金額などの連続した数字列に対して複数回の判定を行うことで、より高度な筆者識別を目指す。

### 参考文献

- [1] Y.LeCun, L.Bottou, Y.Bengio, P.Haffner, “Gradient-Based Learning Applied to Document Recognition”, *Proceeding of the IEEE*, 1998.