

転移学習を用いた小規模データセットで学習可能な 手書き文字と印刷文字の分類モデル

東井 俊輔[†]南葉宗弘[‡]東京学芸大学 情報教育選修[†]東京学芸大学 情報科学分野[‡]

1. はじめに

近年の画像処理の発展により、写真や映像等の画像から文字を認識し、これをテキストデータに変換する技術は広く一般に使われるようになった。しかし、手書き文字と印刷文字が混在する文書における文字認識率が低下してしまう事例が多い。この問題に対しては、文書内に含まれる文字の印字方法を手書きと印刷の二種類に分類した後、それらを別々に文字認識することが有効である[1]。

既存の研究では、1万枚を超える画像データから成る文字画像データセットで学習したモデルを使用して、データセットに含まれる文字種の分類を試みている[2][3]。しかし手書き文字と印刷文字の分類だけに限れば、文字種や言語間の特徴によらず、筆記そのものの特徴で分類が可能であると考えられる。文字種の異なるデータセットで学習したモデルで手書き文字と印刷文字の識別が可能であれば、限られた量のデータしか集められない言語の識別にも使用することができる。そこで本研究では、数字文字だけから成る小規模な画像データセットを用いて印字方法の識別の精度を向上させる方法を提案する。

2 関連技術

2.1 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (CNN) は、層間を共通重みの局所結合でつないだニューラルネットワークである。画像の特徴を抽出するための畳み込み層や、代表値を中心に特徴を抽出するプーリング層を重ねて層を形成する。

A Classification Model of hand-written and machine-printed characters with small dataset using transfer learning,

[†] SHUNSUKE TOI, Tokyo Gakugei University

[‡] MUNEHITO NUMBA, Tokyo Gakugei University

一般に入力に近い層は、入力画像の特徴を抽出し、入力から遠い深層では、分類機としての認識が行われる。

2.2 転移学習

転移学習は、最初の学習タスク用に開発された学習モデルが2番目の学習タスクの学習モデルの初期値として再利用される機械学習の手法である。一般に転移学習には、あるデータセットによる学習で得たモデルの一部を、別のデータセットで再学習させる手法と、入力層に近い部分の重みを固定し、分類機を含む出力層に近い層のパラメータを学習しなおす手法がある。本研究では後者の手法を採り、モデルの作成を行う。

3 提案モデルと実験

3.1 使用するデータセット

本研究では、MNIST から無作為に抽出した手書き数字文字の画像と、それを回転させて作成した画像計600枚に手書き文字のラベルを付与する。一方、表1に示すWindows11に標準搭載されているフォントの文字の画像を同様に回転させたものも含めて計600枚に印刷文字のラベルを付与する。手書き文字の画像と印刷文字の画像のうち、それぞれ480枚を訓練用の画像データ、120枚をテスト用の画像データとする。

3.2 提案モデル

本研究では、図1に示すモデルで学習した後、図2に示すように上層3層のパラメータを固定し、下層2層と分類機を、新たな分類機に置き換えた提案モデル1(2-a)、1層の畳み込み層と分類機に置き換えた提案モデル2(2-b)を作成した。図2の濃く塗りつぶされている層は新しく追加し再学習した層である。

3.3 実験

本研究では、提案モデルをそれぞれ用意したデータセットで学習させ、データセットのテストデータを使い推論を行う実験を行う。

汎化性を確かめる推論用のデータとして、表 1 のフォントを用いた大文字、小文字のアルファベットの画像をそれぞれ 26 枚ずつ計 676 枚用意し、それと同数の手書き文字の大文字、小文字アルファベットの画像を用意した。

本研究では、2 つの提案モデルでの推論の結果と、3 層の畳み込み層と分類器からなる再学習を行わないモデル (比較モデル 1) での推論、5 層の畳み込み層と分類器からなる再学習させないモデル (比較モデル 2) での推論の結果を比較する。

3.4 結果と考察

使用したモデル、学習方法、筆記法ごとの正答率を表 2 に示す。3 層のモデル、5 層のモデルでは出力がそれぞれすべて、手書き文字、印刷文字となり、正しく認識できていなかった。また、提案モデル 1 は最も手書き文字の正答率が高いが、総合的に見ると提案モデル 2 が最も良く認識していた。畳み込み層を追加し再学習したことでモデルの汎化性が向上したと言える。

4 まとめ

CNN の学習において、上層で特徴を抽出し、下層で分類される特性から、学習済みの上層のパラメータのみを使い新たに分類器を学習させることで、数字画像のみのデータセットでアルファベットなど異なる文字種も正しく手書き文字と印刷文字を分類できるという予測のもと本実験を行った。結果、5 層のうち上層 2 層の重みを使用し 1 層の畳み込み層と分類器を再学習させることで、認識率を向上させることがわかった。本研究の検証では、アルファベットでの実験に留まっている。今後、別の言語や記号などの文字での認識に本研究での提案モデルが有用かについて検証したい。

参考文献

[1] Feng Ziyong, Yang Zhaoyang, Jin Lianwen, Huang Shuangping, and Sun Jun, "Robust shared feature learning for script and handwritten machine-printed identification," *Pattern Recognition Letters*, vol. 100, pp. 6-13, 2017.

[2] Mahsa Vafaie, Oleksandra Bruns, Nastasja Pilz, Jörg Waitelonis and Harald Sack, "Handwritten and Printed Text Identification in Historical Archival Documents," *Archiving Conference*, vol.

19, pp. 15-20, 2022.

[3] Jinhong Katherine Guo, Matthew Y Ma, "Separating handwritten material from machine printed text using hidden markov models," *Proceedings of sixth international conference on document analysis and recognition*, pp. 439-443, 2001.

表 1 データセットに使用したフォント

arial	Trebuchet MS	Century	Comic Sans MS
MS Serif	Times New Roman	Courier	Verdana
メイリオ	MSP ゴシック	Trebuchet MS	

表 2 実験結果

モデル	学習方法	認識率 (%)		
		hand	print	総合
提案モデル 1	5 層→3 層	89.4	70.6	80.0
	転移学習	7	2	4
提案モデル 2	5 層→4 層	85.2	84.3	84.7
	転移学習	6	2	9
比較モデル 1	3 層	100	0	50
比較モデル 2	5 層	0	100	50

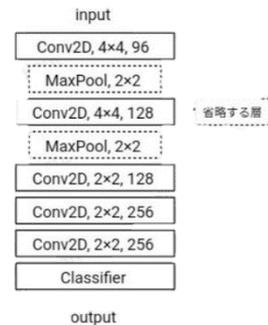


図 1 使用する転移前の 5 層モデル

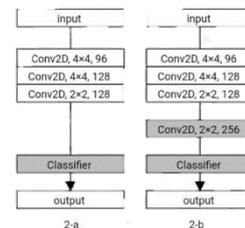


図 2 提案モデル