

令和 6 年 6 月 14 日現在

機関番号：17501

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K12045

研究課題名（和文）手書き文字認識問題を対象とした深層学習における入力パターン内論理構造の自己組織化

研究課題名（英文）Self-Organization of Logical Structures within Input Patterns in Deep Learning for Handwritten Character Recognition Problems

研究代表者

行天 啓二 (Gyohten, Keiji)

大分大学・理工学部・准教授

研究者番号：80305028

交付決定額（研究期間全体）：（直接経費） 1,400,000円

研究成果の概要（和文）：本研究は、手書き文字認識における論理構造理解を目的として進められた。具体的には、文字のストロークの構造を制御し、ニューラルネットワークがどの程度その構造を理解しているかを解析することを目指した。線画自動生成手法を用いて文字認識データの拡張を試み、ストロークレベルでの認識精度向上と誤認識原因の特定を行った。また、文字の構成要素をボトムアップおよびトップダウンの両アプローチで解析する手法についても検討した。その結果、認識精度と説明可能性の向上に成功したが、認識率や自然な手書き文字画像生成において課題が残った。今後は汎用性の検証や定量的評価手法の開発が必要である。

研究成果の学術的意義や社会的意義

本研究は、手書き文字認識における論理構造理解を目指し、文字のストローク構造を制御してニューラルネットワークがその構造をどの程度理解できるかを解析した。また、文字内の部分構造を、ボトムアップ的またはトップダウン的に獲得することができることを示した。これにより、深層学習技術の「ブラックボックス」問題を部分的に解決し、識別結果の説明可能性を向上させたことが学術的意義として挙げられる。社会的意義としては、認識精度と説明可能性を両立させ、深層学習技術の信頼性を高めることに寄与したと考える。また、線画自動生成手法の応用によるデータ拡張の可能性も示すことができたと考える。

研究成果の概要（英文）：This study aimed to understand the logical structure in handwritten character recognition. Specifically, it sought to control the structure of character strokes and analyze the extent to which neural networks understand this structure. By using an automatic line drawing generation method, we attempted to expand the character recognition dataset and improve stroke-level recognition accuracy while identifying the causes of misrecognition. Additionally, we examined both bottom-up and top-down approaches for analyzing character components. As a result, we successfully enhanced recognition accuracy and explainability, although challenges remained in recognition rates and the generation of natural handwritten character images. Future work will involve verifying the generalizability and developing quantitative evaluation methods.

研究分野：知能情報学

キーワード：深層学習 文字認識 線画生成 文字ストローク認識 物体検出 枝刈り

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

パターン認識分野では、深層学習を用いたニューラルネットワークの研究が盛んであり、多くの成果を上げている。特に、識別器が入力パターンを学習することで優れた性能を示すことが知られている。しかし、深層学習で得られる識別器は入力パターンに対応する出力を学習するだけであり、入力パターンに内在する論理構造を理解することはできない。これは、「なぜこのような判断をしたのか」を説明することが困難であるという問題を引き起こし、深層学習技術を社会システムに組み込む際の障害となっている。

現在、研究者たちは自然画像における物体認識を対象に、「入力画像のどの特徴を抽出しているのか」「どの部分に注目して判断したのか」「識別器は認識対象をどのように捉えているのか」を明確化する研究に取り組んでいる。しかし、これらの研究は多くの場合、入力パターン内の論理構造には立ち入っていない。研究代表者は、これまで文字のストローク構造を表現する文字モデルを用いた文字認識に関する研究に携わってきた。これを基に、深層学習における文字認識の過程で文字構造モデルを用いて手書き文字画像を人工的に生成・制御し、ニューラルネットワーク内の重みや認識結果の変化を解析することで、入力パターンの構造に関する知見を得られると考えている。

2. 研究の目的

本研究の目的は、文字認識問題を通じて入力パターンに内在する論理構造を把握することである。具体的には、文字の部首構造やストローク構造を制御しながら文字を学習させ、通常深層学習によって生成される識別器との違いを明らかにすることを目指している。また、学習済みの識別器に対しても同様に文字構造を制御しながら文字を認識させ、その認識結果の変化を検証することで、深層学習による識別器が入力パターンの内在する論理構造をどの程度理解しているかを解析する。

最初に、深層学習における学習用データ拡張のための、手書き文字画像の自動生成手法を検討する。これにより、文字認識の性能を向上させるためのデータ拡張を試みる。具体的には、文字の様々な筆跡を表現し、自然な文字画像を生成する手法の確立を目指し、文字画像を含む線画を自動的に生成することができる手法を確立する。また、生成した文字画像を用いて深層学習による文字認識性能を向上させることを目指す。

次に、文字種を識別する従来の方法に加え、ストロークレベルで文字を認識する識別器を生成することを目指す。これにより、文字の細部構造を捉えることで、深層学習による文字認識性能が、文字細部の認識結果からどのような影響を受けるかについて、詳しく検証する。

次に、通常深層学習により生成された識別器に対して遷移学習を適用し、学習の過程でニューラルネットワーク内の重みを解析したり、弱識別器を組み合わせることによって、文字内・文字間の部分的同一性を把握することを目指す。これにより、通常深層学習で生成される識別器と構造制御を行った識別器を比較し、文字構造を反映した認識が実現できているかを確認する。そして、このような構造がニューラルネットワーク内のどこで表現されているかを特定することを目指す。具体的には、部首構造やストローク構造に部分的に認識することができる弱識別器をトップダウン的、または、ボトムアップに獲得し、文字内・文字間の部分的同一性を把握ししつつ文字自体に内在する論理構造を獲得できるかどうかを確認する。これにより、本研究の目標である、入力パターンに内在する論理構造の自己組織化の可能性を検証する。

3. 研究の方法

(1) 線画の自動生成

線画を自動生成する手法を確立するために、pix2pixを用いた線画の自動生成手法を提案する。具体的には、輪郭線のみによる線画を入力として、特徴線を含む線画を自動生成するためのpix2pixモデルを構築する。このモデルの学習には、元画像から輪郭のみを抽出し、ランダムな射影変換とサイズの正規化を施した訓練データを用いるものとする。さらに、ガウシアンフィルタによるぼかし処理を適用して、生成される線画の詳細度を調整できるようにすることを試みる。

(2) ストローク構造の把握を伴う文字認識

本研究では、ストロークの構造を把握することにより、手書き文字認識における誤認識の原因を特定するための手法を提案する。従来の文字認識手法に加え、文字構造の細部を捉えるためにストロークレベルでの認識を可能にする畳み込みニューラルネットワーク(CNN)を構築する。まず、既存の文字画像を射影変換などの画像処理により拡張し、次にKanjiVGデータベース[1]を用いてストローク単位の構造を持つ文字モデルを生成する。これにより、一部のストロークが欠けた文字画像を作成し、これを学習データとして使用することで、CNNが各ストロークの存在

を認識できるようにする。最終的に、全体の文字認識の結果を出力するために密結合層を追加し、最適化を行う。

(3)文字構造を自動的に把握する文字認識

ボトムアップ的アプローチ

本研究では、Mask R-CNN を使用して漢字の構成要素（部首や簡単な漢字）を検出し、これらの要素がどのように組み合わせられて複雑な漢字を形成するかを理解することを試みる。最初に簡単な漢字の部分学習データとして学習した後、複雑な漢字の構造を検出する方針をとる。検出された部分構成要素を入力として、複数入力の CNN を用いて漢字認識を行う。この方法により、誤認識が発生した場合には、構造に基づいて誤認識の理由を説明することができることが期待される。

トップダウン的アプローチ

本研究では、手書き漢字認識における構造的な判断根拠を示すために、CNN を用いて文字画像を特徴抽出し、その後、簡単なパターンを検出する弱識別器を追加・学習する。学習中にエッジの枝刈りを行うことで、各弱識別器がシンプルなパターンに反応するように調整する。また、これらの検出器間のネットワーク接続にも枝刈りを適用し、弱識別器の階層的な接続を明示的に取得する。最終的には、弱識別器の接続関係を解析し、木構造としてパターン認識プロセスを表現することで、手書き漢字の認識を実現する。この方法により、認識率を大きく低下させることなく、複数の漢字に共通するパターンを持つ弱識別器を取得することを試みる。

4. 研究成果

(1)線画の自動生成

本研究では、提案した手法を用いて靴の画像データセットを対象に実験を行った。結果として、輪郭線のみによる線画から特徴線を含む線画を自動生成し、さらにそれを基に自動着色することができるとも確認した。実験結果を図1に示す。特に、ガウシアンフィルタのカーネルサイズを調整することで、生成される線画およびカラー画像の詳細度を制御できることを確認した。具体的には、カーネルサイズを大きくすることで、ノイズの少ない自然な線画を生成することが可能となり、最終的なカラー画像も高品質なものとなった。この成果は、特にイラスト制作の分野での応用が期待される。

次に、本手法に基づいて文字画像を自動的に生成し、学習用文字画像データを拡張することによる、深層学習による文字認識の性能向上を試みた。しかしながら、従来のデータ拡張手法の性能を大幅に上回るような結果を得ることはできなかった。この理由については今後明らかにしなければならないが、人間が表現し理解することができる多様な筆跡を表現しきれなかったものと考えている。

本研究の今後の展望としては、まずは本手法を他の多様な画像データセットに適用し、その汎用性を検証することが挙げられる。特に、対象を限定しない、より実用的な対象画像での評価が必要である。また、生成された線画の質を定量的に評価するための基準や手法の開発も重要である。

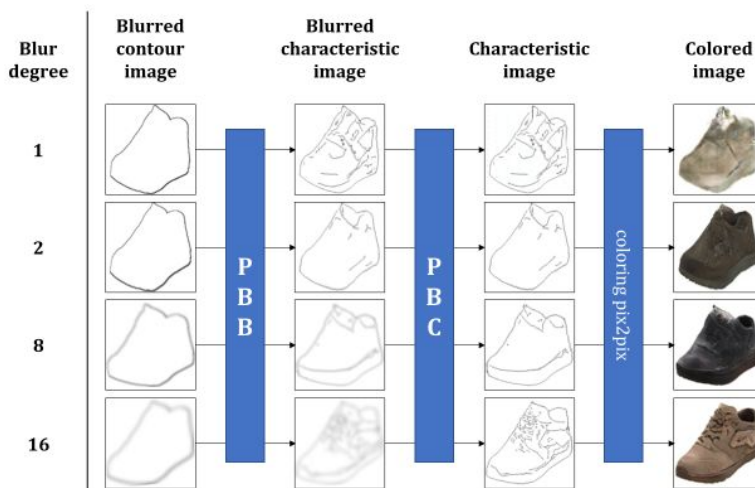


図1 線画自動生成結果 ([2]より引用)

(2)ストローク構造の把握を伴う文字認識

実験結果から、提案手法は誤認識の原因特定に一定の有効性を示した。従来の CNN と比較した場合、提案手法は、各ストロークの認識精度に基づいて文字全体を識別する能力を持つことを確認した。しかし、認識率においては、データ拡張のみを用いた従来の CNN (認識率 99.7%) に対して、提案手法は若干劣る結果 (認識率 99.5%) となった。この原因の一つとして、生成された文字画像が自然でない点が考えられる。より自然な文字画像を生成するための改良が今後の課題となる。

また、誤認識の原因を特定する際、多くの場合で全ストロークに対応するノードが発火しない

現象が観察された。このことから、全ての誤認識原因を特定するには、ストロークレベルの認識精度向上や CNN の構造に更なる改善が必要であると考えられる。

今後の展望としては、CNN のハイパーパラメータの最適化や、ストロークの認識精度向上のための新しい学習データセットの開発が挙げられる。また、異なる筆跡や書き方に対応するためのデータ拡張手法の改善も必要となる。さらに、解釈可能なニューラルネットワークの構築に向けた研究を進めることで、誤認識の原因特定を超えて、より高度な文字認識技術の開発が期待される。

(3)文字構造を自動的に把握する文字認識

ボトムアップ的アプローチ

まず、提案手法により文字認識そのものが実現可能であることを検証した。単純な漢字の構成要素を検出するために、30 文字の漢字を 16 フォントで学習し、複雑な漢字を認識するための複数入力 CNN を構築した。結果として、構造を正確に把握し、高い精度で漢字認識を行うことができることを確認した。具体的には、テストデータに対する認識率は 99.9% に達し、構造に基づく誤認識の理由説明も可能となった。さらに、学習用画像にランダムなノイズを背景に追加したり、学習用画像に複数の漢字を含む画像を挿入することにより、より安定して漢字の構成要素を検出することができるようになることを確認した。複雑な漢字中に存在する、単純な漢字の構成要素を検出した例を、図 2 に示す。

本研究は、漢字認識における構造的な理解を深めることにより、従来の方法では説明が難しかった認識結果の合理的な説明を可能にした。この点で、認識結果の説明が求められる分野において、本研究の成果は大きなインパクトを与える可能性があると考えられる。従来の研究では、視覚的に認識結果を示すことが主流であったが、本研究は、構造に基づく説明を提供する点で一線を画する。これにより、漢字認識の精度と信頼性が向上し、実世界での応用範囲が広がることが期待される。

今後の課題として、少数の未知の構成要素を学習データから取得する方法の実装が挙げられる。本研究では、小学生の初期に学習する漢字に限定して実験を行ったが、より複雑な漢字や未知の構成要素を認識するためには、新しい構成要素を自動的に取得し、学習するメカニズムの確立が必要となる。また、現時点では構造に基づく説明が主観的な比較に留まっているため、定量的な評価方法の考案が求められる。これにより、認識結果の説明の客観性が高まり、さらに広範な応用が可能となると考える。



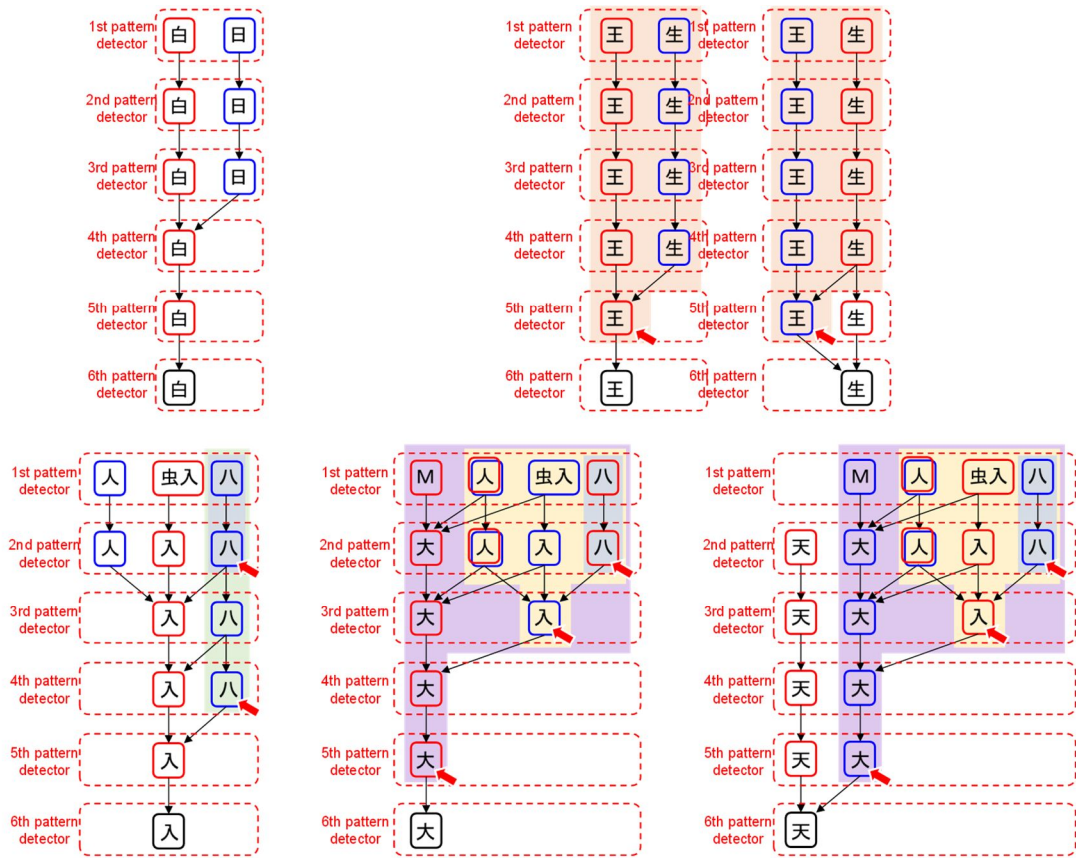
図 2 漢字構成要素抽出結果 ([3]より引用)

トップダウン的アプローチ

本研究では、小学校 1 年生で学ぶ 80 種類の比較的シンプルな手書き漢字画像を対象として実験を行った。CNN による特徴抽出器をまず訓練し、次に弱識別器を順次追加・学習した。さらに、エッジの枝刈りを適用しながら弱識別器間の不要な接続を削減した。結果として、6 層の弱識別器群を接続した場合、最高の認識率 98.0% を達成した。また、枝刈りにより、各パターン検出器が複数の漢字に共通するパターンに反応することが確認された。これにより、認識対象の漢字の構造的な説明が可能となった。本実験により得られた、弱識別器の接続関係の例を、図 3 に示す。

本研究は、手書き漢字認識におけるニューラルネットワークの説明可能性を向上させる新しいアプローチを提案した点で重要であると考えている。従来の研究では、画像中の重要な領域を特定する方法が主流であったが、本研究はパターンの構造に基づく説明を可能とする可能性がある。これにより、機械学習モデルの判断根拠を人間が理解しやすくすることで、社会への受容性を高めることが期待される。

今後の課題としては、枝刈りされたネットワークが具体的にどのようなパターンに反応しているのかを可視化する方法の導入が挙げられる。本研究では、各弱識別器がどの部分に反応しているのかを直感的に理解することが難しいため、既存の可視化手法（例えば、Activation Maximization[4]）などを取り入れることが必要となると考える。また、取得した弱識別器間の木構造の客観的な評価方法の確立も重要である。本研究では主観的な評価が中心であったが、定量的な評価指標を導入することで、モデルの信頼性と説明可能性をより一層高めることができる。



* "M" indicates that there were many character types.

図3 弱識別器接続関係の獲得例 ([5]より引用)

< 引用文献 >

- [1] KanjiVG : <https://kanjivg.tagaini.net/>
- [2] Kazuki Yanagida, et al.:"Automatic Characteristic Line Drawing Generation using Pix2pix", Proceedings of 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM2022) pp.155-162, (2022)
- [3] 滋野友哉:"Mask R-CNN を用いた構造把握に基づく活字漢字認識", 大分大学大学院工学研究科博士前期課程工学専攻修士論文, (2024)
- [4] Feature Visualization : <https://distill.pub/2017/feature-visualization/>
- [5] Keiji Gyohten, et al.:"Character Structure Analysis by Adding and Pruning Neural Networks in Handwritten Kanji Recognition ", Proceedings of the 7th Asian Conference on Pattern Recognition (ACPR2023) no.2 pp.129-142, (2023)

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Keiji Gyohten, Hidehiro Ohki and Toshiya Takami
2. 発表標題 Character Structure Analysis by Adding and Pruning Neural Networks in Handwritten Kanji Recognition
3. 学会等名 Proceedings of the 7th Asian Conference on Pattern Recognition (ACPR2023) (国際学会)
4. 発表年 2023年

1. 発表者名 滋野友哉, 行天啓二, 大城英裕, 高見利也
2. 発表標題 物体検出を用いた漢字の構造把握
3. 学会等名 第75回電気・情報関係学会九州支部連合大会
4. 発表年 2022年

1. 発表者名 Kazuki Yanagida, Keiji Gyohten, Hidehiro Ohki and Toshiya Takami
2. 発表標題 Automatic Characteristic Line Drawing Generation using Pix2pix
3. 学会等名 The 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM2022) (国際学会)
4. 発表年 2022年

1. 発表者名 柳田一樹, 行天啓二, 大城英裕, 高見利也
2. 発表標題 pix2pixを用いた線画生成における特徴的な線の自動描画に関する研究
3. 学会等名 第73回電気・情報関係学会九州支部連合大会
4. 発表年 2020年

1. 発表者名 柳田一樹, 行天啓二, 大城英裕, 高見利也
2. 発表標題 pix2pixを用いた線画生成における特徴的な線の自動描画に関する研究
3. 学会等名 火の国情報シンポジウム2021
4. 発表年 2021年

1. 発表者名 Keiji Gyohten, Hidehiro Ohki and Toshiya Takami
2. 発表標題 A Method to Identify the Cause of Misrecognition for Offline Handwritten Japanese Character Recognition using Deep Learning
3. 学会等名 9th International Conference on Pattern Recognition Applications and Methods(ICPRAM2020) (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------