

修 士 論 文

CRAFT を用いた自由手書き文字列 領域分割の高精度化

令和 5 年度修了

三重大学大学院 工学研究科 情報工学専攻
ヒューマンコンピュータインタラクション研究室

岡島 優太

はじめに

郵便番号認識や帳票認識など枠内に描く手書き文字認識は実用化されているが、自由手書き文書に対する文字認識技術は、まだ発展の途上にある。自由手書き文書認識の実現により、情報のデジタル変換が促進され、業務の RPA 化に役立つ。また、文書管理や検索の効率が向上し、手書きのメモやノートから得られるデータを有効に活用できるようになる。日本語は、他言語に比べて文字種が豊富で類似文字が多いこと、1 文字が複数の連結成分で構成されている文字が存在することにより、文字認識が困難であるとされている。また、自由手書き文書に対する文字認識が研究段階にある理由の 1 つとして、文書からの行検出や文字の切り出しが困難であることが挙げられる。自由手書き文書の文字認識精度を向上させるためには、行検出や文字切り出しを正確に行う必要がある。本研究の目的は、文字切り出しの精度を向上させることである。

文字切り出しの手法として、伊藤 [1] は、行画像の文字画素を垂直方向に投影して作成したヒストグラムの谷で連結成分を区切る手法や、行画像の背景領域 (白画素領域) を細線化して得られた線で行画像を連結成分ごとに分割し、近くの分割領域同士を重なり率に応じて統合する手法を提案している。また伊藤 [2] の研究では、Fully Convolutional Networks (FCN) を用いた文字セグメンテーションの手法が用いられている。これらの手法の課題としては、偏と旁が離れている漢字のように、文字を構成する連結成分が複数ある場合に、切り出し精度が低いことが挙げられる。

そこで本研究では、動的計画法を用いて単語や文字列単位で文字らしさの和が最も良くなるように文字領域を分割する手法を提案し、文字切り出しの精度向上を目指す。提案手法では、文字検出手法の CRAFT(Character Region Awareness for Text detection)[3] を用いて領域分割を行った後、文字の外接長方形のアスペクト比を利用した統合処理や、文字の連結成分間の距離などを利用した事前統合により精度向上を試みた。また、分割した文字候補に対する文字らしさとして、日本語手書き文字データベースである ETL9B[11] で学習した CNN を用いて予測した認識結果の字種に対する信頼度を使用する。

実験の結果、文字の切り出し精度は 86.6% となった。伊藤 [1] の手法を同条件で行なったものは文字の切り出し精度が 66.1%、CRAFT[3] のみを利用した手法では 83.8% となり、提案手法の精度が上回る結果となった。

今後の課題として、連結成分が別の文字と繋がっているものや文字サイズが小さい促音「っ」などの切り出し、ETL9B には無い文字種類であるカタカナや数字などへの対応が挙げら

れる。また, アスペクト比が 1:1 から大きく外れる「一」や「し」のような文字に適切な対処をする必要がある。本論文では, 1 章では研究背景, 先行研究や研究目的について, 2 章では関連知識, 3 章で提案手法, 4 章で実験と考察, 5 章で結言としてまとめと今後の課題について述べる。

目次

はじめに	i
第 1 章 緒言	1
1.1 研究背景	1
1.2 先行研究	1
1.3 研究目的	2
第 2 章 関連知識	3
2.1 光学的文字認識に関する基礎知識	3
2.2 畳み込みニューラルネットワーク	3
2.3 CRAFT(Character Region Awareness for Text Detection)	4
第 3 章 提案手法	5
3.1 データセット	5
3.2 CRAFT による領域抽出	6
3.3 連結成分の事前統合	8
3.4 文字候補結合領域の文字らしさ	12
3.5 動的計画法 (DP マッチング) による文字の切り出し	17
第 4 章 実験	18
4.1 実験データ	18
4.2 評価手法	18
4.3 実験結果	19
4.4 考察	19
第 5 章 結言	21
5.1 まとめ	21
5.2 今後の課題	21
付録 A 付録	22
A.1 データファイル	22

A.2	プログラムの詳細	22
謝辞		23

第 1 章

緒言

1.1 研究背景

郵便番号認識や帳票認識など、特定の条件下での手書き文字認識は実用化されているが、自由筆記文書に対する文字認識技術は、まだ実用化の途上にある。自由筆記文書とは、ノートやメモのように自由な形式で筆記された文書を言う。自由筆記文書の文字認識が実用化されれば、紙に書かれたメモやノートのスキャン画像をコンピュータ上でテキストデータに変換することで、キーワード検索が可能となり、文書管理が容易になる。また、近年スマートフォンの普及に伴い、一般のユーザが手軽に写真を撮影することが容易となり、メモや掲示物を撮影して記録する用途にも幅広く利用されている。画像内の文字を高精度で自動検出し認識できれば、翻訳システムや看板情報を利用したナビゲーションシステムなど、多岐にわたる応用が考えられるため、自由筆記文書文字認識技術への期待が高まっている。一方、自由筆記文書に対する文字認識技術が研究段階にある理由の 1 つとして、文書からの正確な文字の切り出しが困難であることが挙げられる。したがって、文字の切り出し精度の向上は、自由筆記文書の文字認識精度向上に繋がる。

1.2 先行研究

伊藤 [1] は文字の切り出し手法として、次の (2)~(4) の手法を提案し、一般的な手法 (1) と比較している。

- (1) 行画像の文字画素 (黒画素) を垂直方向に投影して作成したヒストグラムの谷で連結成分を区切る。
- (2) (1) の後で、幅の狭い連結成分を隣の連結成分と統合する。
- (3) 行画像の背景領域 (白画素領域) を細線化した線で連結成分を分離し、近くの連結成分同士を重なり率を用いて統合する。
- (4) 行画像中の文字連結成分をラベリングし、近くにある連結成分どうしを外接矩形の重なり率などを用いて統合する。

手法 (1)～(4) の評価用画像に対する文字切り出し成功率は, (1) 80.4% , (2) 79.0% , (3) 66.1% , (4) 81.5%であった. これらの手法の課題として, 1 つの文字を構成する連結成分が複数ある場合に切り出し精度が低いことが挙げられる.

また伊藤の研究 [2] では, Fully Convolutional Networks (FCN) を用いた文字セグメンテーションの手法が提案されている. この手法では文字認識を行っていないため, 偏と旁が離れている漢字を偏と旁で構成される漢字とするか, あるいは違う漢字同士とするかの判断ができず, 別々の文字だと判別してしまうことが課題である. 切り出し成功率は 66.1%と実用には不十分であり, 改善の余地がある.

歌川の研究 [4] では, 文字の切り出しに CRAFT[3] という文字検出手法が利用されている. 歌川は文字を外接矩形でトリミングした学習用データと, 上下左右の文字にかからない範囲の最大の余白で文字をトリミングした学習用データの 2 種類を用いた機械学習で, 手書き文字に生じる文字どうしの連結やサイズの違いに対応している. 課題として, 文字と文字の間隔が狭い場合への対応や句読点の切り出しが挙げられている.

1.3 研究目的

先行研究には, 複数の連結成分からなる文字や偏と旁が離れている文字に対して正しい文字の切り出しが困難であるという課題がある. そこで本研究では, 文字認識と動的計画法を組み合わせた文字区切りの手法を提案し, 文字区切りの精度を向上させることをこの研究の目的とする.

第 2 章

関連知識

2.1 光学的文字認識に関する基礎知識

光学的文字認識は Optical Character Recognition (OCR) とも言われ, スキャナやデジタルカメラなどによって読み取られた文書画像をコンピュータで扱うことのできるデジタルデータに変換する技術のことである. OCR の基本的な処理の流れは以下のようなものである.

1. 画像入力
2. 前処理
3. 文字セグメンテーション
4. 文字認識

前処理では, 入力画像が 2 値画像に変換され, 文字画素以外の黒画素領域を除去するノイズ除去などの処理が行われる. 文字セグメンテーションでは, 画像中の各文字の領域を求める. その後, 各領域内の文字画像を認識用辞書と比較して文字認識が行われる.

OCR の歴史は古く, 1900 年前後から研究が始まっている. 電信技術の拡張と視覚障害者が文字を読むのを支援するために研究されていたのが原点であったという [5]. 日本では, 1968 年に郵便番号が導入され, 20 世紀末には手書きの郵便番号を認識するシステムが実用化された. 最近まで, 文字認識手法は特徴抽出と統計的機械学習によるものが主流であったが, 2010 年代になると深層学習が使われるようになった. 近年では, 自然言語処理において大きな成果を挙げた Transformer[7] を文字列認識に応用した認識手法も提案されている. また, OCR の技術を応用し, ナンバープレートの認識 [8] や情景中の文字を認識する技術 [9] も研究されている.

2.2 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は Convolutional Neural Network (CNN) とも言われ, 画像を分類したり特徴を抽出したりする深層学習手法の一種である. CNN は通常, 畳み込み層, プーリング層を繰り返し配置したもので, 入力から特徴を抽出することができる. 画像認識問

題では得られた特徴ベクトルを全結合層に渡して、出力層では最も信頼度の高いクラスラベルを答えさせる。畳み込み層では、カーネルと呼ばれる重み行列 (フィルタ) を入力行列上で移動させながら、加重和を計算することで出力を得る (特徴マップ)。移動量はストライドと呼ばれ通常は 1 である。畳み込み層を複数用意することで様々な特徴が抽出できる。プーリング層では、特徴マップ内の領域ごとにダウンサンプリングを行う。全結合層では、多階層の畳み込み層やプーリング層を通して得られた特徴を元に、最終的な出力を生成する。文字認識の研究では、LeCun[6] の研究で CNN が採用され、高い精度が報告されている。

2.3 CRAFT(Character Region Awareness for Text Detection)

CRAFT[3] とは、Character Region Awareness for Text Detection の略で画像内のテキスト領域を検出する深層学習ベースの文字検出手法である。2019 年に NAVER Corp. によって CVPR2019 で発表され、情景文字の検出において優れた精度を示した。ピクセルが文字の中心にくる確率を示す Region Score と、ピクセルが隣接する文字間のスペースの中心に来る確率を示すヒートマップを出力する。情景内の文字列は、傾いている場合や変形・湾曲している場合もあり、今までのバウンディングボックスを用いた検出では正確に文字領域を抽出することは難しかった。CRAFT は、この課題に対処するために、Region Score と Affinity Score を組み合わせて使用し、多様な形状の文字列を検出することに成功した。

第 3 章

提案手法

提案手法の概要を図 3.1 に示す. CRAFT の出力である文字領域のヒートマップスコアを利用して文字列領域を分け, 動的計画法 (DP マッチング) で単語や文字列単位で 1 文字あたりの文字らしさの平均値が最大になるように領域を組み合わせる文字とする. 本研究では, 文字らしさの評価値として ETL9B で学習した CNN を用いて認識した字種の信頼度を使用する.



図 3.1: 提案手法の概要

3.1 データセット

本研究では, 文字認識のための学習用データに ETL 文字データベースの ETL9B を用いる. ETL9B は, ひらがな 71 字種と JIS 第 1 水準漢字 2965 字種の計 3036 字種を, 各字種 200 サンプルずつ格納する手書き単文字画像データベースである. ETL9B の文字画像例を図 3.2 に示す.



図 3.2: ETL9B の文字画像例

3.2 CRAFT による領域抽出

CRAFT では, 文字列画像を入力すると図 3.3(a) のように文字領域が推定されたヒートマップが出力される. 提案手法では, 文字列画像を入力し CRAFT で出力されるヒートマップスコアを利用して文字列の領域分けを行う. 具体的には, 以下の 3 ステップで文字領域を抽出する.

1. CRAFT で, 文字領域推定ヒートマップを得る. (図 3.3(a))
2. ヒートマップからしきい値 t_h 以上の領域の外接枠を得る. (図 3.3(b))
3. 外接枠が文字画素全体を囲むように拡大する. (図 3.3(c))
4. 検出されていない文字連結成分の外接枠を追加する. (図 3.3(d))



(a) 文字領域推定ヒートマップ



(b) ヒートマップから外接枠生成



(c) 文字連結成分が囲まれるように外接枠を修正



(c) 囲まれていない文字連結成分の外接枠を追加

図 3.3: CRAFT を用いた領域抽出方法

3.3 連結成分の事前統合

動的計画法での計算量を減らすため事前にある程度連結成分を統合しておく。本研究では、以下の 5 つの方法で連結成分の事前統合を行う。3.3.1 は、連結成分を囲む外接枠どうしが、 x 軸方向および y 軸方向にどれだけ重なっているかにより統合する手法、3.3.2 と 3.3.3 は、2 つの各外接枠の辺どうしの距離を基に事前統合する手法、3.3.4, 3.3.5 は、それぞれの外接枠の辺上の点間の最短距離を基に事前統合する手法である。

3.3.1 外接枠の x 軸方向 y 軸方向の重なり

この手法では、連結成分を囲む外接枠どうしが x 軸または y 軸方向にどれだけ重なっているかに基づいて事前に統合を行う。具体的な手順は以下の通りである。

1. 3.2 節で得た外接枠の面積が式 (3.3) で計算した t_s 以下の連結成分を求める。
2. 1 で求めた連結成分の外接枠と x 軸方向に 80%以上かつ y 軸方向に 80%以上重なっている外接枠があれば統合する。 x 軸方向については図 3.4 において式 (3.1) または式 (3.2) を満たす場合となる。 y 軸方向についても同様である。ただし、統合後の 2 つの外接枠を含む外接枠の幅が式 (3.4) の t_w 以上になる場合は統合しない。
3. 統合対象が無くなるまで 1, 2 を繰り返す。

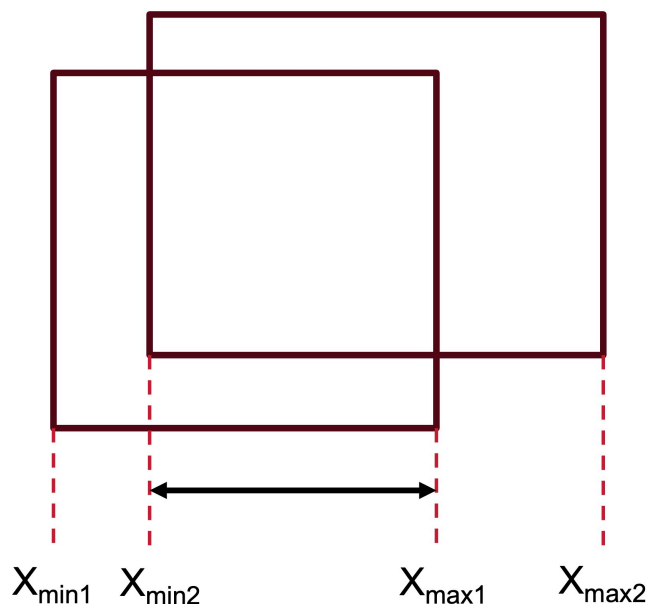


図 3.4: x 軸方向の重なり条件の説明図

$$(X_{max1} - X_{min1}) \times 0.8 \leq X_{max1} - X_{min2} \quad (3.1)$$

$$(X_{max2} - X_{min2}) \times 0.8 \leq X_{max1} - X_{min2} \quad (3.2)$$

$$t_s = \mu_s - \sigma_s \quad (3.3)$$

t_s : 外接枠の面積が小さいと判定する時の閾値

μ_s : 外接枠の面積の平均値

σ_s : 外接枠の面積の標準偏差

$$t_w = \mu_w + \sigma_w \quad (3.4)$$

t_w : 外接枠の幅が大きいと判定する時の閾値

μ_w : 外接枠の幅の平均値

σ_w : 外接枠の幅の標準偏差

3.3.2 外接枠間の辺の距離

この手法では、連結成分を囲む各外接枠の上下左右の辺の距離に基づいて統合するか否かを判定する。具体的な手順は以下の通りである。

1. 3.2 節で得た外接枠の面積が式 (3.3) で計算した値以下の連結成分を求める。
2. 1 で求めた連結成分の外接枠とその他の外接枠の x 軸方向, y 軸方向の辺間の最短距離が式 (3.5) を満たす場合に統合する。ただし、統合後の 2 つの外接枠を含む外接枠の幅が式 (3.4) の t_w 以上になる場合は統合しない。
3. 統合対象が無くなるまで 1, 2 を繰り返す。

$$\max(d_x, d_y) < \mu_w \times 0.4 \quad (3.5)$$

d_x : 2 つの外接枠の x 軸方向の辺間の最短距離

d_y : 2 つの外接枠の y 軸方向の辺間の最短距離

μ_w : 外接枠の幅の平均値

ただし, d_x, d_y は, 外接枠の重なりがある場合には 0 とした.

3.3.3 外接枠間の縦辺の最短距離 (絶対値)

この手法では, 連結成分を囲む外接枠間の x 軸方向の距離に基づいて統合を行う. 具体的な手順は以下の通りである. 図 3.5 に統合の様子を示す.

1. 3.2 節で得た外接枠の面積が式 (3.3) で計算した t_s 以下の連結成分を求める.
2. 1 で求めた連結成分の外接枠との縦の辺の x 軸方向の距離 d (絶対値) が最小となる外接枠を統合する. ただし, 統合後の 2 つの外接枠を含む外接枠の幅が式 (3.4) の t_w 以上になる場合は統合しない.
3. 統合対象が無くなるまで 1, 2 を繰り返す.

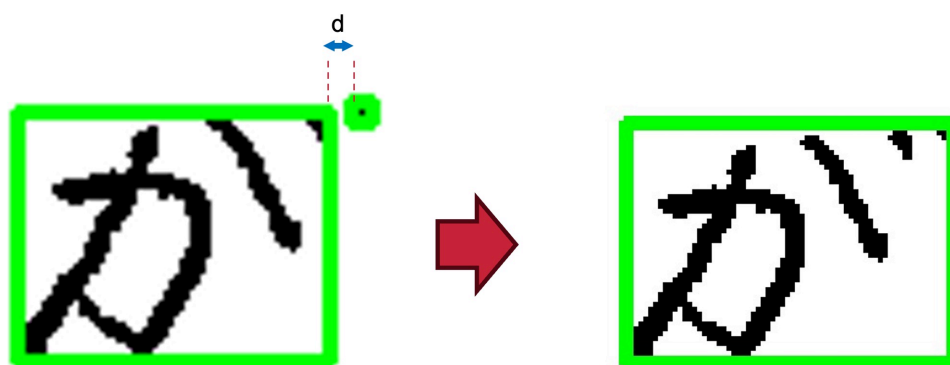


図 3.5: 縦の辺の x 軸方向距離による事前統合の例

3.3.4 外接枠間の midpoint の最短距離

この手法では, 連結成分を囲む外接枠間の辺の midpoint の距離が最短になるものを統合する. 図 3.6 に統合の様子を示す. 具体的な手順は以下の通りである. 以下の処理は統合する外接枠がなくなるまで繰り返す.

1. 3.2 節で得た外接枠の面積が式 (3.3) で計算した値以下の連結成分を求める.
2. 1 で求めた連結成分の外接枠の辺の 4 つの midpoint うちのうち任意の 2 点間の距離が最小となる外接枠を統合する. ただし, 統合後の 2 つの外接枠を含む外接枠の幅が式 (3.4) の t_w 以上になる場合は統合しない.
3. 統合対象が無くなるまで 1, 2 を繰り返す.

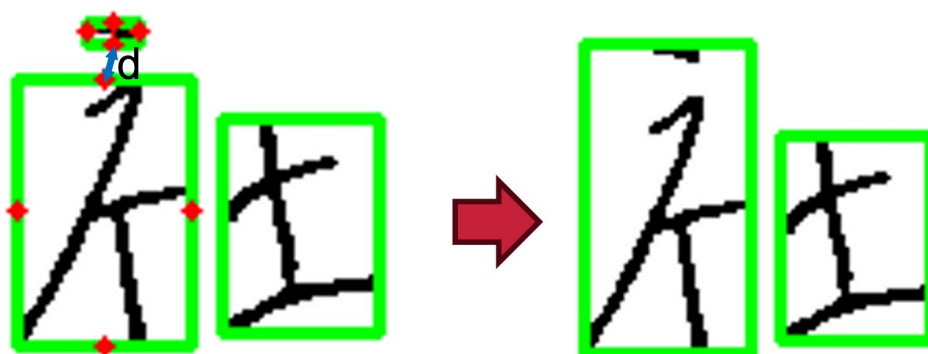


図 3.6: midpoint の距離による事前統合の例

3.3.5 外接枠の 4 つの辺上の 16 点のうち任意の 2 点間の最短距離

この手法では, 連結成分を囲む外接枠間の辺上の 16 点うちのうち任意の 2 点間の距離が最短になるものを統合する. 16 点とは, 図 3.7 に示すように, 上下左右のそれぞれで辺の始点, 終点と, 辺を 4 等分する点のことである. 具体的な手順は以下の通りである.

1. 3.2 節で得た外接枠の面積の平均と標準偏差を計算し, 外接枠の面積が式 (3.3) で計算した値以下の連結成分を求める.
2. 1. で求めた連結成分の外接枠の 4 つの辺上の 16 点うちのうち任意の 2 点間の距離が最小となる外接枠を統合する. ただし, 統合後の 2 つの外接枠を含む外接枠の幅が式 (3.4) の t_w 以上になる場合は統合しない.
3. 統合対象が無くなるまで 1, 2 を繰り返す.



図 3.7: 外接枠の辺上の 16 点

3.4 文字候補結合領域の文字らしさ

3.3 節の事前統合で生成した文字領域を 1 セグメント, または隣り合う 2, 3, 4 セグメントを結合した文字候補を対象とし, すべての文字候補について文字らしさを求める. 文字らしさは, 3.4.1 で後述する CNN を用いて文字候補を認識し, 認識字種に対する信頼度に ETL9B の文字アスペクト比分布 (図 3.11) に応じてペナルティを与えた数値を使用する. 事前統合後の文字領域が図 3.8 であるとき, その文字候補の例は図 3.9(a)~(d) のようになる.

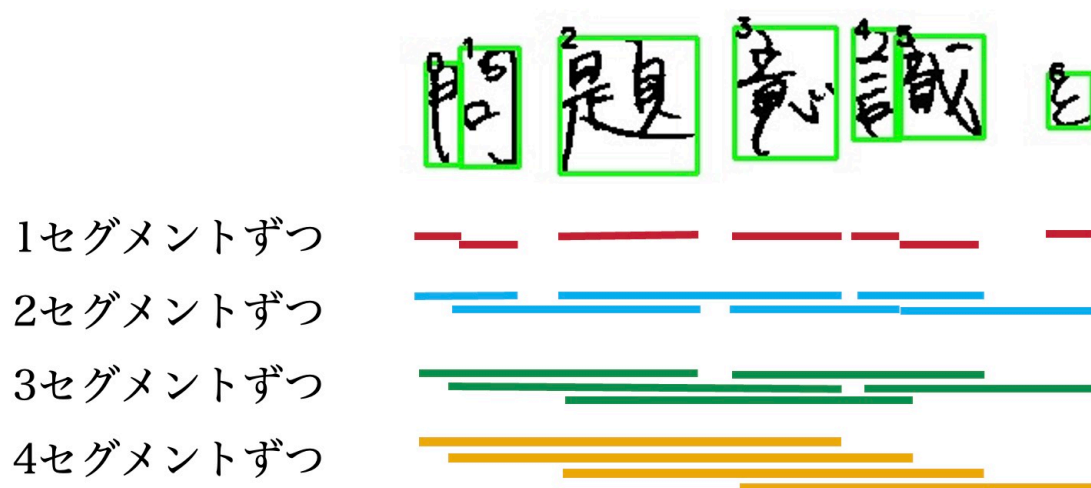


図 3.8: 事前統合後の文字領域とセグメントの組み合わせ

(a) 1 セグメントの組み合わせの例

(b) 2 セグメントの組み合わせの例

(c) 3 セグメントの組み合わせの例

(d) 4 セグメントの組み合わせの例

図 3.9: 組み合わせた文字候補の例

3.4.1 CNN の学習

実装した CNN の構成を図 3.10 に示す。まず畳み込み層 1 (C1) には, 入力画像として ETL9B の日本語の手書き文字を与え, 5×5 のフィルタで畳み込みを行い 64 枚の特数量マップを得る。次にプーリング層 2 (P2) では, C1 の出力を入力として, 小領域内の最大値を選択するマックスプーリングにより画像のサイズを小さくする。そして, この動作を 2 度繰り返すことでプーリング層 6 (P6) の出力は 5×5 のサイズとなり, これを全結合層 7 (F7) に入力して識別させる。出力層の活性化関数にはソフトマックス関数を用いる。CNN の構成の詳細は表 3.1 に示す。

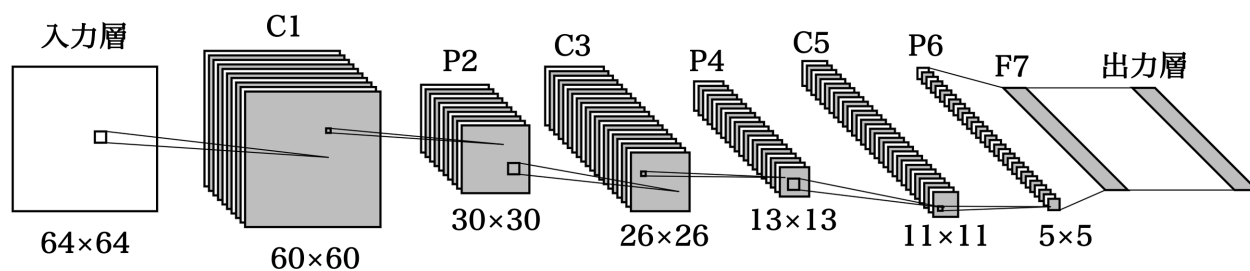


図 3.10: CNN の構成

表 3.1: CNN の構成の詳細

特徴量マップ	縦×横×枚数	フィルタサイズ
入力層	$64 \times 64 \times 1$	—
C1	$60 \times 60 \times 64$	5×5
P2	$30 \times 30 \times 64$	2×2
C3	$26 \times 26 \times 96$	5×5
P4	$13 \times 13 \times 96$	2×2
C5	$11 \times 11 \times 128$	3×3
P6	$5 \times 5 \times 128$	5×5

3.4.2 ETL9B の文字アスペクト比分布に応じた重み付け

文字らしいアスペクト比を持つ文字候補の信頼度を高め、文字らしくないアスペクト比をもつ文字候補の信頼度を低くするため ETL9B の文字アスペクト比分布に応じた重み付けを行なった。CNN で学習したデータセットである ETL9B の全文字のアスペクト比分布を図 3.11 に示す。度数の自然対数を取り最大値が 1 になるように正規化したグラフを図 3.12 に示す。正規化した数値を用い、文字認識結果に各文字候補のアスペクト比に応じた重み付けを行う。重み付けは以下の条件で計算した。アスペクト比は式 (3.6) で定義する。

- 文字候補のアスペクト比が 2 を超える場合、マイナス無限大の重み付けをする。
- 文字候補のアスペクト比が 0.4 以下、もしくは 1.4 以上の場合、文字認識結果の 2 乗に文字候補のアスペクト比に対応した ETL9B のアスペクト比分布を正規化した数値を乗算する。
- 文字候補のアスペクト比が 0.4 以上 1.4 以下の場合、文字認識結果の数値をそのまま用いる。

$$\text{アスペクト比} = \frac{W}{H} \quad (3.6)$$

W : 文字候補の幅

H : 文字候補の高さ

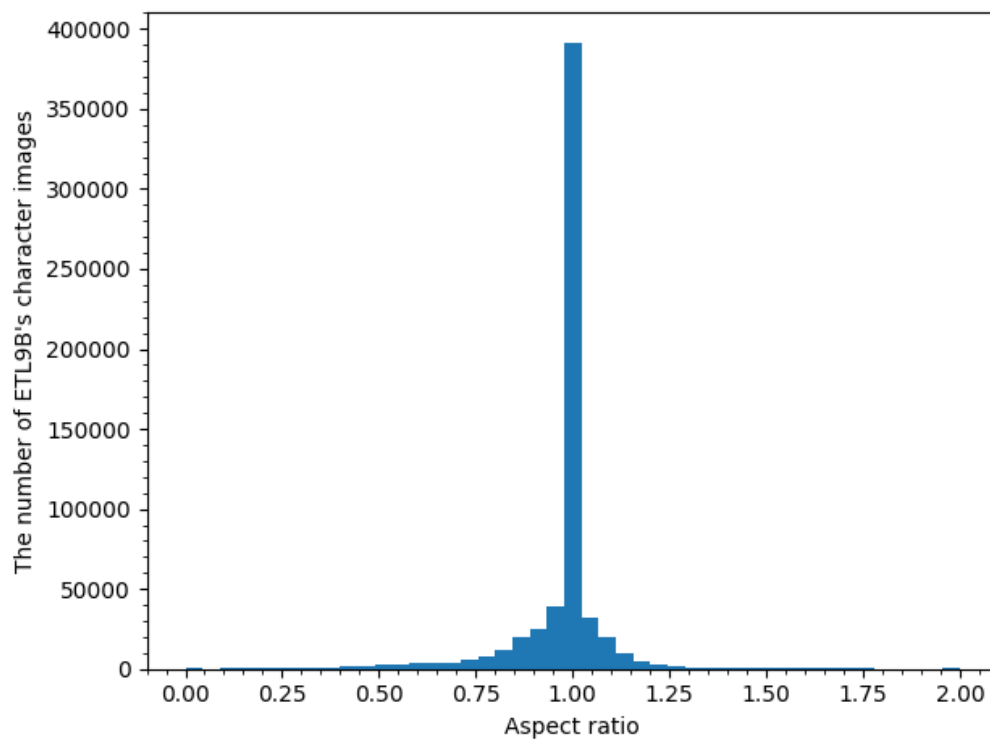


図 3.11: ETL9B のアスペクト比分布

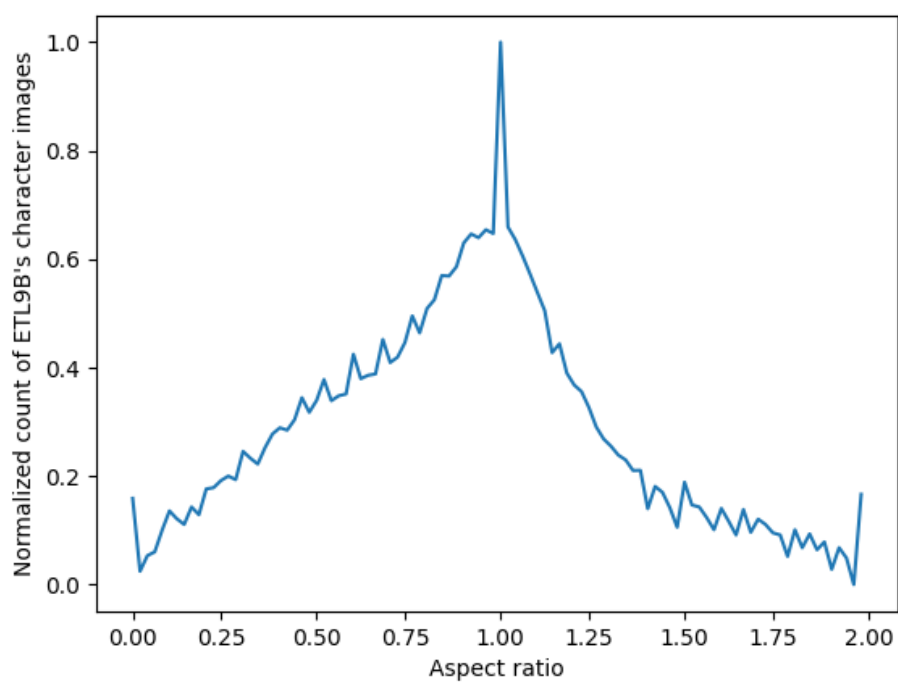


図 3.12: ETL9B のアスペクト比分布の度数の対数を取り最大値が1になるよう正規化したグラフ

3.5 動的計画法 (DP マッチング) による文字の切り出し

文字列全体で 3.4 節で求めた 1 文字あたりの文字らしさの平均値が最大になる組み合わせを動的計画法により求める。動的計画法による類似度の和の計算式を以下に記述する。

3.5.1 4 セグメントまで考慮する場合

$$\begin{aligned}
 f(0, 0) &= d_1(0, 0) \\
 f(0, 1) &= d_2(0, 1) \\
 f(0, 2) &= d_3(0, 2) \\
 f(0, 3) &= d_4(0, 3) \\
 f(1, 1) &= f(0, 0) + d_1(1, 1) \\
 f(1, 2) &= \max(f(0, 1) + d_1(1, 2), f(0, 0) + d_2(1, 2)) \\
 f(1, 3) &= \max(f(0, 2) + d_1(1, 3), f(0, 1) + d_2(1, 3), f(0, 0) + d_3(1, 3)) \\
 f(2, 2) &= f(1, 1) + d_1(2, 2) \\
 f(2, 3) &= \max(f(1, 2) + d_1(2, 3), f(1, 1) + d_2(2, 3)) \\
 f(3, 3) &= f(2, 2) + d_1(3, 3)
 \end{aligned}$$

一般の場合:

$$\begin{aligned}
 1 \leq i \leq M - 1, \quad 4 \leq j \leq N - 1 \\
 f(i, j) &= \max \left(f(i - 1, j - 1) + d_1(i, j), f(i - 1, j - 2) + d_2(i, j), \right. \\
 &\quad \left. f(i - 1, j - 3) + d_3(i, j), f(i - 1, j - 4) + d_4(i, j) \right)
 \end{aligned}$$

$d_1(i, j)$ 、 $d_2(i, j)$ 、 $d_3(i, j)$ 、 $d_4(i, j)$ はそれぞれ以下のことを示す:

- $d_1(i, j)$: 第 j セグメントを第 i 文字目としたときのある字種に対する類似度
- $d_2(i, j)$: (第 $j - 1$ セグメント + 第 j セグメント) を第 i 文字目としたときのある字種に対する類似度
- $d_3(i, j)$: (第 $j - 2$ セグメント + 第 $j - 1$ セグメント + 第 j セグメント) を第 i 文字目としたときのある字種に対する類似度
- $d_4(i, j)$: (第 $j - 3$ セグメント + 第 $j - 2$ セグメント + 第 $j - 1$ セグメント + 第 j セグメント) を第 i 文字目としたときのある字種に対する類似度

$f(i, j)$ の計算結果から、文字候補ごとのある字種への類似度を求め、全文字候補で 1 文字あたりの平均をとったものが最大になるような文字数とセグメンテーションを求める。

第 4 章

実験

4.1 実験データ

実験用のデータとして, 独自に収集した日本語自由手書き文書の文字列画像を用いる. これは三重大学学生のレポートの画像から ETL9B に存在する文字種を含む 5 文字からなる文字列画像を切り出したものである. 文字列の長さを 5 文字に統一したのは, 問題を単純化してアルゴリズムの有効性を確認するためである. 実験には, 切り出した文字列画像を 500 枚, 計 2500 文字を用いる. CRAFT のヒートマップの閾値 t_h は 0.43 に設定し文字の領域分割を行った. CRAFT による領域抽出後の処理で連結成分の最適な組み合わせを求めたいため, 閾値 t_h はあらかじめ細かく領域が分かれるように設定した.

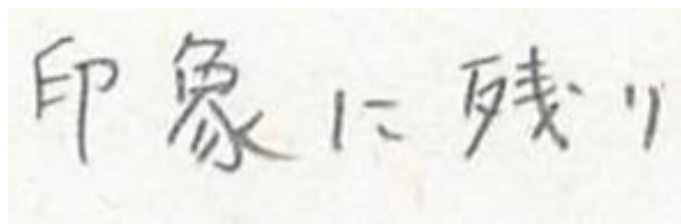


図 4.1: 文字列画像の例

4.2 評価手法

文字切り出しの精度は, 正しく切り出された文字数を N_c , 全行画像に含まれる文字数を $N_{all} = 2500$ として, 式 (4.1) で求める. 文字の一部が欠けていたり, 別の文字の連結成分が含まれているものは失敗と判定する.

$$\text{文字切り出し成功率 (\%)} = \frac{N_c}{N_{all}} \times 100 \quad (4.1)$$

4.3 実験結果

実験結果を表 4.1 に示す. 手法は以下に記述する. (1),(2) は伊藤 [1] の研究で用いられた手法, (3)~(6) は提案手法である. 各手法では使用した実験データは 4.1 節で説明したもので統一している.

- (1) 白画素細線化と統合処理を用いた文字切り出し
- (2) 連結成分の外接矩形と統合処理を用いた文字切り出し
- (3) 外接枠の x 軸方向 y 軸方向の重なりを利用した事前統合による文字切り出し
- (4) 外接枠の辺の距離を利用した事前統合による文字切り出し
- (5) 外接枠の辺の距離 (絶対値) を利用した事前統合による文字切り出し
- (6) 外接枠の辺の 4 つの中点どうしのうち任意の 2 点間の距離を利用した事前統合による文字切り出し
- (7) 外接枠の辺上の 16 点どうしのうち任意の 2 点間の距離を利用した事前統合による文字切り出し

表 4.1: 実験結果

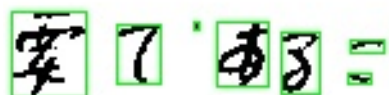
実験手法	切り出し成功率
(1) 白画素細線化と統合処理	65.4%
(2) 連結成分の外接矩形と統合処理	66.1%
(3) 外接枠の x 軸方向 y 軸方向の重なりを利用	85.4%
(4) 外接枠の辺の距離を利用	86.0%
(5) 外接枠の辺の距離 (絶対値) を利用	86.3%
(6) 外接枠同士の辺の 4 つの中点どうしのうち任意の 2 点間の距離を利用	86.5%
(7) 外接枠同士の辺上の 16 点どうしのうち任意の 2 点間の距離を利用	86.6%

手法 (7) が最も高い精度を達成し、86.6%の正確さを示した. 外接枠の辺上の 16 点どうしのうち任意の 2 点間の距離を利用して文字候補を事前統合することが, 提案手法の文字切り出しにおいて効果的であった.

4.4 考察

文字の切り出し成功率は先行研究 (1)(2) に比べて大幅に向上した. しかし, 偏や旁が分かれている漢字のように複数の連結成分で構成されている文字の切り出し精度は向上したが, 図

4.2 の「で」の濁点部分のように連結成分が文字要素から離れているものや、図 4.3 の「し」や「て」のように外接枠のアスペクト比が 1:1 から大きく外れているものは切り出しに失敗する例が多く見られた。現在、文字認識の学習は ETL9B のデータのみで行っているが、崩れた文字も学習に加えることや、DP マッチングを行う際に組み合わせるセグメント数を増やすことにより、さらなる精度向上につながると考えられる。また、文字列の前後関係の情報をを用いた知識処理を行うことで、さらなる精度の向上が期待できる。

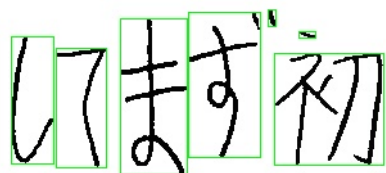


(a) 事前統合後の文字候補領域

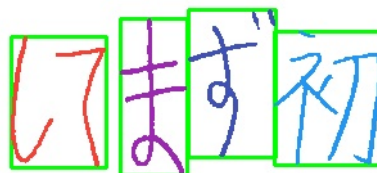


(b) 文字切り出し結果

図 4.2: 連結成分が離れているものの例



(a) 事前統合後の文字候補領域



(b) 文字切り出し結果

図 4.3: アスペクト比が極端なものの例

第 5 章

結言

5.1 まとめ

本研究では、動的計画法により単語や文字列単位で文字らしさの平均値が最大になるように文字領域を分割する手法を提案した。文字候補の領域分割方法として CRAFT を用いた処理を導入し、事前統合手法を比較した。実験は日本語の手書き文書である三重大学学生レポート中の、ETL9B の文字種からなる 5 文字の文字列画像 500 枚を使用した。本研究では、外接枠の重なり率、外接枠の辺の x, y 軸方向の距離、外接枠同士の辺の 4 つの midpoint うちの任意の 2 点間の距離、外接枠同士の辺上の 16 点うちの任意の 2 点間の距離を利用した事前統合を比較した。実験の結果、(7) が最も良く 86.6% となった。提案した手法は全て、CRAFT のみを利用した場合よりも切り出し成功率が向上している。また、先行研究に比べ 20.5% 精度が向上した。

5.2 今後の課題

今後の課題として、文字を構成する連結成分が文字要素から離れているものや外接枠のアスペクト比が 1:1 から大きく外れているものの対処が挙げられる。また、本研究では ETL9B に含まれる 3036 字種のみで学習・実験を行なっているため、日本語の自由筆記文書に対応させるためには句読点などの記号、数字、アルファベットなどのデータセットの拡充が必要である。更なる精度向上には、文字切り出しの際に文字認識を行いながら、前後の文字との関係などの知識処理を用いることが必要であると考え。本手法では、自由筆記文書中の文字を切り出した画像を用いているが、実用化するには文書中からの正確な行の検出、情景中の文字やコントラスト比の少ない画像といったあらゆる画像での文字検出の精度が求められる。

付録 A

付録

A.1 データファイル

本研究に関するデータファイルはすべて以下のディレクトリに存在する.

ETL9B

- /net/nfs2/export/home/okazima/workspace/ETL/ETL9B

実験データ

- /net/nfs2/export/home/okazima/workspace/CRAFT/data

A.2 プログラムの詳細

本研究に関するプログラムはすべて以下のディレクトリに存在する.

- /net/nfs2/export/home/okazima/workspace/CRAFT

プログラムの使用方法, データの詳細についてはディレクトリの README に記述する.

謝辞

本研究を進めるにあたり，多くの助言をいただき研究のご指導を賜りました若林哲史教授，ミーティングなどで助言をいただいた盛田健人助教，白井伸宙助教に深く感謝致します。また，これまで発表スライドの添削や研究に対して多くのアドバイスを与えてくださった研究室の先輩方，ともに切磋琢磨した同期の皆様に心から感謝を申し上げます。最後になりましたが，これまで私を支えてくれた家族に感謝の意を表して，本論文の結びといたします。

参考文献

- [1] 伊藤 蒼平, “深層学習によるオフライン自由筆記文書の 認識に関する研究,” 三重大学修士論文, 2021.
- [2] 伊藤 悠青, “オフライン自由筆記文書の文字認識に向けた文字セグメンテーション,” 三重大学卒業論文, 2020.
- [3] Baek, Youngmin, et al. ”Character region awareness for text detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
- [4] 歌川 重樹, ”自由手書き文字認識のための精度向上に有効な文字切り出し手法,” 会津大学短期大学部 2022 年度卒業研究論文要旨集, 2023.
- [5] H. Schantz, “The History of OCR. Manchester Center,” VT: Recognition Technologies Users Association, 1982.
- [6] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, ”Gradient-based learning applied to document recognition,” Proceedings of the IEEE 86.11, 1998.
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. ”Attention is all you need,” Advances in neural information processing systems, 2017.
- [8] Shashidhar, R., et al. ”Vehicle Number Plate Detection and Recognition using YOLO-V3 and OCR Method,” 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNBC). IEEE, 2021.
- [9] Singh, Amanpreet, et al. ”Textocr: Towards large-scale end-to-end reasoning for arbitrary-shaped scene text,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021.
- [10] 佐々木俊介, 陳奎廷, 馬場孝明, ”畳み込みニューラルネットワーク (CNN) を用いた手書き日本語文字認識システムの試作,” 電気関係学会九州支部連合大会講演論文集 2015(0), 348-348, 2015.
- [11] 電子技術総合研究所, Japanese Technical Committee for Optical Character Recognition, ”ETL 文字データベース,” 1973-1984.