

近代書籍のための CRAFT を用いたレイアウト解析手法

飯田 紗也香¹ 竹本 有紀¹ 石川 由羽² 高田 雅美¹ 城 和貴¹

概要：近代書籍とは、明治から昭和初期に刊行された活版印刷の文書である。近代書籍は現代の OCR 技術に用いられるレイアウト解析処理では、正確な解析が困難である。本稿では、風景画像から文字切り出しを行う CRAFT 手法を、文書領域抽出と文字切り出し処理に分けて適用し、帝国議会議録を用いて検証を行う。

A CRAFT-Based Layout Analysis Method for Early-Modern Japanese Books

1. はじめに

明治から昭和初期にかけて刊行された、活版印刷の文書を近代書籍と呼ぶ。国立国会図書館が提供する Web サービスでは、近代書籍が公開されている。例えば、帝国議会議録検索システム [1] では、活版印刷された帝国議会議録の国会・委員会の会議録が全 92 回分公開されている。議会内では、当時の出来事に即した議論が行われており、会議録は当時の状況を知るための重要な資料である。帝国議会議録を含め、近代書籍のほとんどは画像データ形式により公開されている。戦後期以前の会議録にはテキストデータが存在しない。そのため、テキストデータを用いた本文内容の検索や音声読み上げ、翻訳ができず、誰でも容易に扱うことが難しい状況である。よって、近代書籍を便利に扱うためにテキスト化が求められている。

現在、近代書籍文字認識は様々な方法で試みられており、その精度は日々向上している [2]。近代書籍文字認識では、主に正しく切り出された文字に対する認識を想定している。そのため、文字認識を実用化するためには、正確なレイアウト解析方法が必要不可欠である。しかし、現代の光学文字認識 (Optical Character Recognition, OCR) ソフトウェアに用いられる既存のレイアウト解析手法では、近代書籍のレイアウト解析を正確に行うことは不可能である。正確なレイアウト解析を行うには、文字列の幅や図

表の有無など、書籍のレイアウトについて前提条件の設定が必要である。現代の印刷物は、フォントや印刷様式が決まっているため、正しく文字を切り出すことが容易である。近代書籍は統一された規格が存在しない時期に発行されており、出版者や時代により印刷様式が異なる。よって、書籍ごとに前提条件を設定しなければ、正しいレイアウト解析ができない。近代書籍全般のレイアウト解析には汎用的なレイアウト解析手法が求められている。

本稿では、CRAFT (Character-Region Awareness For Text detection) [3] を用いた文書領域抽出を行う。CRAFT とは、看板や標識等が写る RGB の風景画像から、任意の形状に並ぶ文字列の切り出しを行う手法である。畳み込みニューラルネットワーク (Convolution Neural Network, CNN) へ画像を入力して出力されるヒートマップに対して後処理を行うことで、任意の形状に並ぶ文字列の切り出しを行う。本稿では CRAFT 手法を用いて近代書籍からの文書領域抽出を試みる。文字切り出し用の学習済みモデルに対して、1,000 文字以上の文字を含む、高解像度の疑似近代書籍画像を追加で学習する。これにより、会議録画像から文書領域抽出を試みる。抽出された文書領域からさらに文字切り出しを行い、文字切り出しの精度を検証する。

本稿の構成は、以下の通りである。第 2 章では、CRAFT 手法について説明する。第 3 章では、CRAFT を用いた文字切り出しについて述べる。第 4 章では、会議録画像に対する文書領域抽出と文字切り出しの実験を行う。文書領域抽出と文字切り出しの実験方法について述べて、実験結果を示して考察を行う。

¹ 奈良女子大学
Nara Womens University

² 滋賀大学
Shiga University

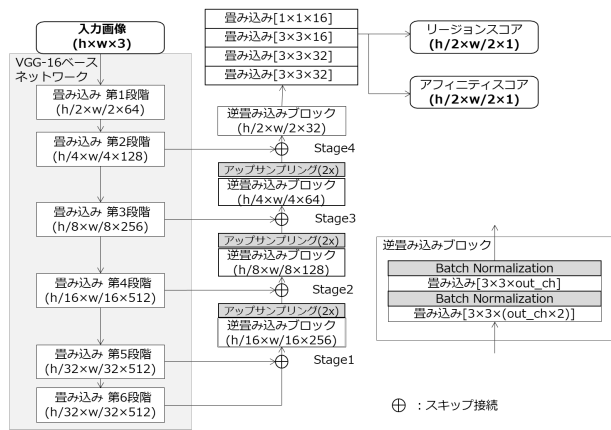


図 1: CRAFT に用いられる CNN

2. CRAFT

CRAFT (Character-Region Awareness For Text detection) [3] とは、看板や標識など、文字列を含む風景画像から文字列切り出しを行う手法である。CRAFT では、ヒートマップで表現される検出対象における中心位置の確率を CNN を用いて出力する。出力されたヒートマップに対して後処理を行うことで、文字列の切り出しを行う。CRAFT では、デザイン性のある看板やポスターにみられる、任意の形状に並ぶ文字列の切り出しが可能である。

2.1 CRAFT における CNN

CRAFT に用いられる CNN は、Semantic Segmentation[4] 等に用いられる、完全畳み込みニューラルネットワーク (Fully Convolution Network, FCN) [6] である。FCN とは、全結合層を持たず、畳み込み層のみで構成されている CNN のことである。図 1へ CRAFT に用いられる CNN を示す。この CNN はエンコーダ部分は VGG-16[7] と類似する構造をしており、U-Net[8] と同様にデコーダ部分へスキップ接続を持つ。U-Net は Semantic Segmentation を行う FCN の 1 種である。スキップ接続とは、特徴抽出を行う畳み込み層の特徴マップを保存しておき、対応する逆畳み込み層に足し合わせる処理のことである。スキップ接続により、プーリング層で失われる位置情報がデコーダ部分で復元される。CRAFT の CNN は文字列を含む画像と、それに対応する文字の位置と文字間の位置をラベルとしたデータセットを学習することでヒートマップの出力を行う。出力されるヒートマップを用いて後処理を行い、文字矩形および単語矩形を取得する。

CRAFT に用いられる学習データは、文字を含む風景画像と、ヒートマップによるラベルのセットである。ラベルデータには、図 2 に示すリージョンスコアとアフィニティスコアの 2 つのヒートマップを用いる。リージョンスコアは、図 2a のように与えられた画素が文字の中心である確率を表す。アフィニティスコアは、図 2b のように、与え



(a) リージョンボックス
およびリージョンスコア
(b) アフィニティボックス
およびアフィニティスコア

図 2: ラベルデータの矩形とヒートマップ [10]

られた画素が隣接する 2 文字間の中心である確率を示す。学習データのラベル作成手順について述べる。まず、リージョンボックスとアフィニティボックスを作成する。図 2a に示す赤色枠線のような矩形領域をリージョンボックスとする。リージョンボックスは文字を囲む 4 点の座標により構成されている。図 2b に示す黄色枠線のような、隣接文字間の矩形領域をアフィニティボックスとする。隣接する 2 文字それぞれのリージョンボックスについて、図 2b における緑色枠線のように、対角線を結ぶことで 2 つの三角形を作成する。文字列の方向が縦方向に並ぶ場合は左右に並ぶ三角形、横方向の場合は上下に並ぶ三角形を用いる。このとき、隣接する 2 文字に含まれる合計 4 つの三角形の重心を結んだ矩形領域をアフィニティボックスと定義する。次に、ガウス分布に従いマッピングした画像を作成する。この画像をガウスマップ画像と呼ぶ。最後に、リージョンスコアとアフィニティスコアを作成する。これらのヒートマップ画像は、先に作成されたガウスマップ画像を用いて作成する。それぞれの矩形に従い、ガウスマップ画像を線形変換して張り付ける。書籍画像 1 ページに含まれる文字は非常に多く、その全てに対して個別にガウス分布を計算することは非常に計算コストがかかる。あらかじめ作成されたガウスマップ画像を矩形に合わせて線形変換することで、ヒートマップを作成する。

Semantic Segmentation を用いた近代書籍に対するレイアウト解析 [5] では、各画素に対してそれぞれラベル付けを行う。一方、CRAFT では中心である確率をガウス分布に従うヒートマップによりラベル付けする。学習データにおけるラベルのヒートマップ表現は、厳密な境界を持たない領域を扱う際に高い柔軟性を発揮する。そのため、Dark Pose[9] などの姿勢推定を行う CNN モデルに用いられることが多い。これらのヒートマップは、関節の位置などの中心位置そのものを求めるために用いられている。CRAFT では、ヒートマップを単語や文字を囲む矩形領域を求めるために用いている。矩形領域は CNN の出力に対する後処理により求められる。

2.2 CRAFT における後処理

CRAFT は、出力されるヒートマップに対して後処理を

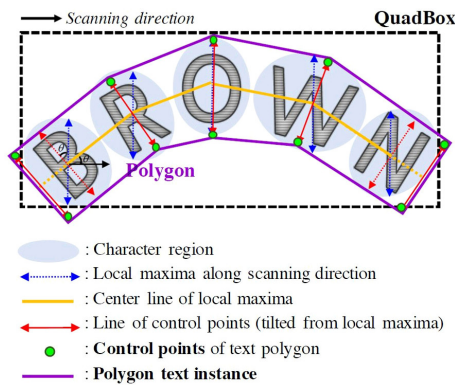


図 3: CRAFT における後処理 [3]

行うことで、文字列切り出しを行う。CRAFT の後処理について述べる。出力されるヒートマップ 2 枚に対して、閾値処理による 2 値化を行う。それらを加算したものから単語領域を検出する。閾値処理されたリージョンスコアは、図 3 における Character region のように示される。単語領域から、図 3 の点線で示されるような、単語領域を囲む矩形を取得する。この時、得られる単語領域を囲む矩形のことを、Quad Box と呼ぶ。得られた単語領域と Quad Box から、Polygon の作成を行う。Polygon とは、単語の形状に沿った領域のことである。Character region から、図 3 に示される青色矢印線のように、文字の高さを取得する。これらを Local Maxima 線と呼ぶ。Local Maxima 線の中から最も長い線を単語矩形全体の高さとして、各文字の Local Maxima 線の長さを統一する。各文字における Local Maxima 線の中点を隣接する文字同士で結び、これを図 3 における黄色線に示す Center Line とする。両端の文字を通る Center Line の端は Character region の端まで延長する。図 3 に示す赤色矢印線のように、Center Line に垂直になるように Local Maxima 線を回転して、Center Line の端と回転された Local Maxima 線から図 3 の紫色枠線のような Polygon を作成する。この後処理により、CRAFT では湾曲した文字列の検出が可能となる。

CRAFT が対象とする文字は、風景画像中に含まれる看板や標識、ポスター等に印刷された活字文字である。そのため、書籍画像のように小さな文字が大量に含まれる画像や、白黒 2 値画像には対応していない。本稿では、CRAFT 手法を用いて、近代書籍画像から文書領域抽出が可能か検証を行う。疑似近代書籍画像を学習することで、文書領域の抽出および文字切り出しを試みる。

3. CRAFT を用いた文字切り出し

CRAFT を用いた文字切り出しの検証 [10] で提案されている文字切り出し手法について述べる。この文字切り出し手法では、CRAFT と同じ CNN モデルを用いて、疑似的に生成される近代書籍風の画像を学習する。本稿では、学習データとして生成される画像を疑似近代書籍画像と呼ぶ。

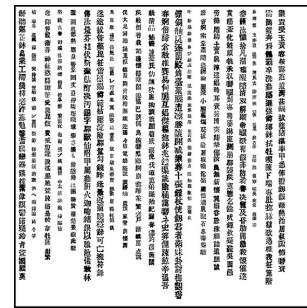


図 4: ベースモデルに用いられる疑似近代書籍画像

疑似近代書籍画像は、近代書籍から切り出された文字が縦書き方向に貼り付けられた白黒 2 値画像である。CNN へ疑似近代書籍画像を学習することで、CRAFT 手法による近代書籍に対する文字切り出しを行う。

CRAFT 手法による文字切り出しは、前処理、CNN による認識、後処理の流れで行う。まず、前処理として会議録画像の分割を行う。画像の分割は、画像に含まれる文字数が学習データに含まれる文字数の範囲内になるように行う。次に、分割された画像を CNN で認識してヒートマップを作成する。最後に、後処理を行う。2 つのヒートマップのうちリージョンスコアのみを後処理に用いる。リージョンスコアに対して閾値処理を行い、背景と文字領域に 2 値化する。2 値化されたリージョンスコアから領域抽出を行い、文字領域矩形を取得する。分割された全ての画像に対して、取得される矩形情報と画像の統合を行う。以上の流れで文字切り出しを行う。

この文字切り出し手法に用いる CNN には、図 4 のような疑似会議録画像を学習に用いている。この画像の解像度は 1,024x1,024px であり、最大約 700 文字がランダムに張り付けられている。この疑似近代書籍画像 10,000 種類を学習するモデルを、ベースモデルと呼ぶ。ベースモデルを用いて、会議録画像 3 枚に対して文字切り出しを行う。その結果、平均 93.7 % の文字が切り出されている。ベースモデルの検証では、会議録画像を分割せず入力した場合、図 5 のように、ほとんどの文字におけるヒートマップが正しく出力されない。帝国議会会議録検索システムで公開されている画像は解像度が約 4,600x3,200px であり、1 ページにつきおよそ 1,000 から 2,000 個の文字が含まれている。そのため、ベースモデルの検証では、会議録画像を分割して入力を行っている。この時、会議録画像に含まれる文字数は 700 文字以内になるように、重複する領域を含めながら分割している。しかしながら、同じ文字を二重に認識したり、認識結果を統合する際に正しく切り出された文字ではなく切り出しに失敗した文字が優先される場合がある。そのため、文書領域の抽出が求められている。

文書領域の抽出方法として、CNN を用いた Semantic Segmentation による文書領域抽出方法 [5] がある。この文書領域抽出手法には、CRAFT に用いられる CNN と同様

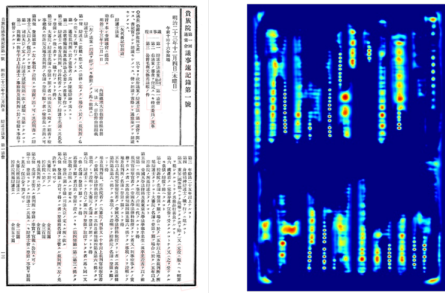


図 5: ベースモデルへ会議録画像をそのまま入力する場合のリージョンスコア出力

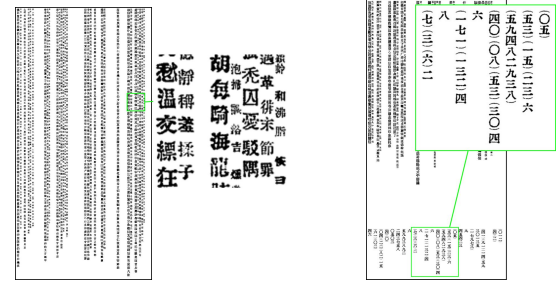
に FCN の構造を持つ CNN が用いられている。そこで、本稿では CRAFT を用いた文書領域抽出を試みる。

4. 実験

4.1 実験方法

本実験では、CRAFT を用いた文書領域抽出方法について検証する。文書領域抽出に用いるモデルを文書領域抽出モデルと呼ぶ。実験では、文書領域抽出モデルを用いて、5 枚の会議録画像に対する文書領域抽出精度を検証する。

文書領域抽出モデルは、3 章で述べるベースモデルに対して、新たに作成する疑似近代書籍画像を追加学習して作成する。文書領域抽出モデルに用いる学習データとして、図 6 のような疑似近代書籍画像を生成する。学習データに用いる疑似近代書籍画像は、最大で 1,350 文字を含み、解像度は 1,024x2,048px である。この文字数は、等間隔に文字を配置して解像度を正規化した際に、文字として読める限界の文字数である。解像度は、本実験に用いる GPU1 枚で一度に学習可能な最大サイズである。学習データの生成方法について述べる。学習データは、文字画像の配置、リージョンボックスとアフィニティボックスの作成、ヒートマップの作成という流れで生成する。アフィニティボックス生成方法とヒートマップの作成方法は、2.1 節で述べる CRAFT 手法のものと同様である。配置される文字画像には、近代書籍画像から収集された文字画像を用いる。使用する文字種は漢字、カタカナ、ひらがなの 3,044 種類である。近代書籍は国立国会図書館デジタルコレクション [11] に公開されている物を用いる。疑似近代書籍画像のレイアウトは、縦書き方向上詰めとする。文字画像の配置について説明する。まず、最大行数と行に含まれる最大文字数を指定する。指定された最大行数と文字数の範囲内で、行数と文字数を乱数によりランダムに指定する。次に、行と行内の文字の間隔を指定する。行と行内の文字の間隔は、乱数を用いてランダムな幅に変更する。最後に、文字を貼り付ける。貼り付ける文字画像は行ごとにランダムな大きさに縮小する。文字を貼り付ける際に、文字を囲む矩形の座標を取得して、リージョンボックスを作成する。生成される画像によっては、画像の半分が余白となる場合がある。



(a) 最大 1350 文字を指定して生成される疑似近代書籍画像

(b) 余白に漢数字と括弧を貼り付けられた疑似近代書籍画像

図 6: 文書領域抽出モデル用学習データ

その場合、図 6b のように画像下部分へかっこ漢数字を張り付ける。この時、用いる文字画像は会議録画像から収集されるかっこ漢数字の画像 12 種類を用いる。行数は画像の幅から算出し、行内の文字数は乱数によりランダムに指定する。行の幅と文字の間隔は一定である。

本実験の環境について述べる。実行する OS は Ubuntu18.04、計算機は Intel® Core™ i9-7960X CPU @ 2.80GHz × 32 で GPU は NVIDIA GeForce RTX 2080Ti を 4 枚使用する。ベースモデルの学習には、同じ実験環境で 217 時間 11 分が必要とされる。文書領域抽出モデルを 1 から学習するには、膨大な時間が求められる。よって、文書領域抽出モデルは追加学習により作成する。ベースモデルに対して、新たに生成されるデータセットを追加学習する。使用する学習データの枚数は 200 種類で、ミニバッチ学習により学習を行う。バッチサイズは 4 である。学習回数は 1,000 Epoch で、計算時間は 24 時間 52 分である。

CRAFT を用いた文書領域抽出方法の流れを述べる。まず、会議録に対する前処理として、枠線除去などのノイズ除去を行う。次に、会議録画像を CNN へ入力してヒートマップを出力する。このとき、学習モデルは文書領域抽出モデルを用いる。最後に、出力されるリージョンスコアとアフィニティスコアに対して CRAFT の後処理を適用する。抽出された領域に重複箇所がある場合、それらの領域を 1 つに統合する。統合された領域を、抽出される文書領域として、会議録画像から切り出す。抽出された文書領域ごとに、文字切り出しを行う。文字切り出し手法には、3 章で述べる手法を用いる。その結果、会議録画像 1 ページに含まれる全ての文字から、正しく切り出される文字の割合を用いて評価を行う。枠線などの誤認識はノイズ除去処理における精度の問題なので、本稿では評価に用いない。

4.2 実験結果

本実験では、会議録画像 1 ページに含まれる文字のうち、切り出された文字の割合を切り出し精度の評価に用いる。帝国議会会議録画像に対して文書領域抽出と文字切り出し



図 7: 実験に用いる会議録画像

を行い、それぞれの画像における切り出し精度を算出する。評価には図 7 に示す 5 枚の会議録画像を用いる。これらの会議録画像には見出し文字が含まれており、本文と異なるサイズの文字が含まれている。5 枚の会議録に対して、文書領域抽出と文字切り出しを行う結果を表 1 へ示す。文書領域抽出精度は平均で 95.2%，最大 98.4%である。文字切り出し精度は平均で 90.8%，最大 97.0%である。

3 章で述べるベースモデルを用いた文字切り出し手法の検証と精度を比較する。こちらの実験では、本実験における画像 A, B, C の 3 枚を評価に用いており、表 2 に示すように平均で 93.7%の文字が切り出される。本実験における画像 A, B, C の平均文字切り出し精度は 89.9%である。ベースモデルの検証と比較して、本実験における平均文字切り出し精度は低くなる。その原因として、文字切り出し精度は、文書領域の精度に依存することが挙げられる。

ヒートマップの出力による文書領域抽出失敗例について述べる。図 7a の画像 A における文書領域抽出時のリージョンスコアに着目する。図 8 に画像 A のリージョンスコアと、後処理で切り出される文書領域を赤色枠線で示す。図 8 の拡大部分では、カタカナや漢数字の「一」などの画数が少ない文字部分について、ヒートマップが十分な濃さで出力されていないことが確認できる。これは、学習データに含まれる文字種が原因と推測される。疑似近代書籍画像に含まれる文字種のほとんどは、画数の多い漢字である。そのため、学習データ 1 枚に含まれる文字数が多くなるほど、画数の多い文字が多く含まれる傾向がある。また、画像 1 枚に含まれる文字数が多くなると、画素の密度が上がる。これにより、画素の密度が低い画数の少ない文字は認識しづらくなると考えられる。

後処理が原因である文書領域抽出の失敗例について述べる。図 7e に示す、画像 E における文書領域抽出モデルが出力するリージョンスコアに着目する。図 9 に示す赤色枠線が画像 E における切り出される文書領域である。見出し文字がある位置に、ヒートマップが正しく出力されているにも関わらず、文書領域が正しく切り出されていないことがわかる。これは、後処理が原因ではないかと推測される。大きい文字の場合、文字における中心以外にあたる領域が

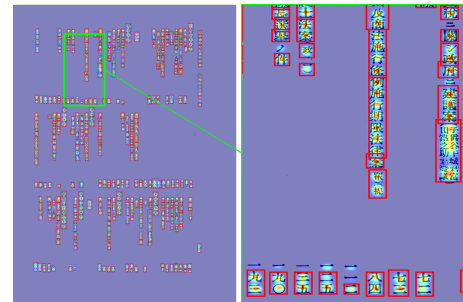


図 8: 画像 A のリージョンスコア出力

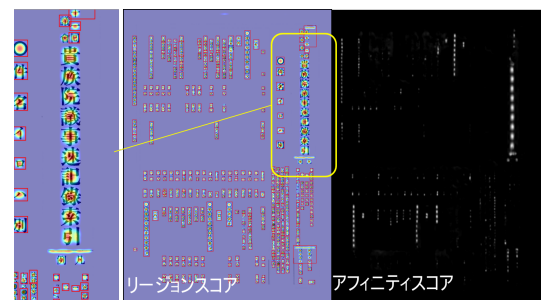


図 9: 画像 E のヒートマップ出力

大きくなる。そのため、大きい文字はヒートマップが薄い領域が増える。薄いヒートマップは後処理における閾値処理で背景に分類される。検出されるヒートマップから正しく領域を切り出すためには後処理の改良が必要である。

本実験に用いる会議録画像の全てには、見出し文字が含まれている。しかし、一般的に書籍で用いられるほとんどの文字は同じ大きさである。これは、近代書籍においても同様である。また、CNN モデルを全ての文字サイズに対応させるには、さらに学習データの枚数を増やし、その解像度を上げる必要がある。これは、本実験に用いる環境以上の膨大な計算リソースを必要とするため、現実的ではない。よって、本文と大きさが異なる文字は、個別に処理を行うことで精度向上の見込みがあると考えられる。

出力されるヒートマップについて、出力される位置は正しいが薄く出力されるヒートマップは、CRAFT の後処理で正しく切り出されない。これは、ヒートマップに対する閾値処理を用いた 2 値化が原因であると考えられる。薄く

表 1: 本実験における実験結果

	総文字数 (個)	領域抽出	文字切り出し	文字切り出し精度	文書領域内の	文書領域切り出し精度
		失敗文字数 (個)	失敗文字数 (個)		文字切り出し精度	
画像 A	960	66	159	83.4%	85.3%	93.1%
画像 B	1,404	41	66	95.3%	96.7%	97.1%
画像 C	1,916	81	170	91.1%	92.1%	95.8%
画像 D	1,346	21	41	97.0%	98.5%	98.4%
画像 E	847	72	109	87.1%	89.3%	91.5%

表 2: ベースモデルにおける実験結果 [10]

	評価文字数 (個)	抽出文字数 (個)	文字切り出し精度
画像 A	938	848	90.4%
画像 B	1,388	1,358	97.8%
画像 C	1,895	1,761	92.9%

出力されるヒートマップが背景領域として置き換えられるため、領域抽出の際に正しく切り出されないという問題がある。また、ガウス分布に従うヒートマップは、中心部分が濃く、外側に向かうにつれ薄くなる。そのため、文字の端を示すヒートマップが閾値処理により背景領域に置き換わり、文字の一部がはみ出る場合がある。これらの問題は、CRAFT における後処理での 2 値化処理を閾値処理から改良することで改善が見込める。

本実験により、疑似近代書籍を学習することで、CRAFT の手法を用いた近代書籍からの文書領域および文字切り出しの有用性が示唆される。大きさの異なる文字に対する個別の処理と、CRAFT 手法における後処理の改良により、切り出し精度の向上を目指す。

5. まとめ

本稿では、近代書籍のテキスト化に必要な汎用レイアウト解析のために、CRAFT を用いたレイアウト手法の検討を行っている。CRAFT とは、看板や標識が写された RGB の風景画像から、任意の文字列を切り出す手法である。近代書籍文字を縦書き方向上詰めのレイアウトでランダムに張り付けた画像を疑似近代書籍画像と呼ぶ。CRAFT と同じ CNN に対して疑似近代書籍画像を学習させることで、近代書籍からの文書領域抽出と、文字切り出しを試みる。評価は、帝国議会会議録画像を対象に行う。

疑似近代書籍画像が学習されたモデルを用いて実験を行う。学習データとして、最大 700 文字がランダムに張り付けられた画像 10,000 種類が学習されたモデルをベースモデルと呼ぶ。ベースモデルに対して追加学習を行うモデルを、文書領域抽出モデルと呼ぶ。文書領域抽出モデルの学習には、最大 1,350 文字がランダムに張り付けられた画像 200 種類を用いる。文書領域抽出モデルを用いて文書領域抽出を行う。また、抽出された文書領域に対して、ベースモデルを用いた文字切り出しを行う。会議録画像に対する

文書領域抽出と文字切り出しの結果について検証を行う。

実験の結果、5 枚の会議録画像について、平均 95.2% の精度で文書領域が抽出される。また、平均 90.8% の精度で文字が切り出される。実験に用いたモデルでは、本文とサイズの異なる文字をうまく切り出すことができない傾向がみられる。今後は、認識画像に含まれる見出し文字等、サイズの異なる文字を個別に処理することで、文字切り出しの精度向上を目指す。

参考文献

- [1] 帝国議会会議録検索システム (online), 入手先 (<https://teikokugikai-i.ndl.go.jp/>) (2022.01.30).
- [2] Yasunami, S., Koiso, N., Takemoto, Y., Ishikawa, Y., Takata, M., & Joe, K. Applying CNNs to Early-Modern Japanese Printed Character Recognition.
- [3] Baek, Youngmin, et al. "Character region awareness for text detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
- [4] Thoma, Martin. "A survey of semantic segmentation." arXiv preprint arXiv:1602.06541 (2016).
- [5] Iida, Sayaka, et al. "Layout analysis using semantic segmentation for Imperial Meeting Minutes." Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2019.
- [6] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- [7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
- [8] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
- [9] ZHANG, Feng, et al. Distribution-aware coordinate representation for human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 7093-7102.
- [10] 飯田紗也香, et al. 近代書籍における文字切り出し手法の検討. 研究報告数理モデル化と問題解決 (MPS), 2021, 2021.4: 1-6.
- [11] 国立国会図書館デジタルコレクション (online), 入手先 (<https://dl.ndl.go.jp/>) (2022.01.30).