

ニューラルネットワークを用いた 古典籍画像からの文字検出

西村 遼平[†]宮崎 智^{††}菅谷 至寛^{††}大町 真一郎^{††}[†] 東北大学工学部電気情報物理工学科^{††} 東北大学大学院工学研究科

1. はじめに

過去の文書に記載されている情報は今後起きることの予測など、貴重な情報源となるため、古典籍に書かれている情報を活用する試みが行われている。しかし、現在の文字とは字体が異なること、情報量が膨大であることなどから、有効に活用するのは難しい。そこで本稿では、入力したキーワードを検索するシステムの実現を目的とし、物体検出の手法を用いた文字検出について検討する。

2. 関連手法

2.1 YOLO YOLOとは、物体の検出とクラス分類を同時に行うことで、高速な物体検出を可能にしたネットワークである[1]。入力画像の特徴量を計算した後画像を分割し、bounding boxを用いた物体検出と、分割された領域内のクラス分類を並行して行い、物体の検出・認識をしている。

2.2 U-Net U-Netとは、画像のセグメンテーション(画像中のクラス検出をピクセル単位で行うこと)に特化したネットワークである[2]。CNNによる畳み込みの際に、深い層で得られる局所的な特徴量と浅い層で得られる正確な位置情報を組み合わせることによって、精度の高いピクセル単位での分類を可能にしている。

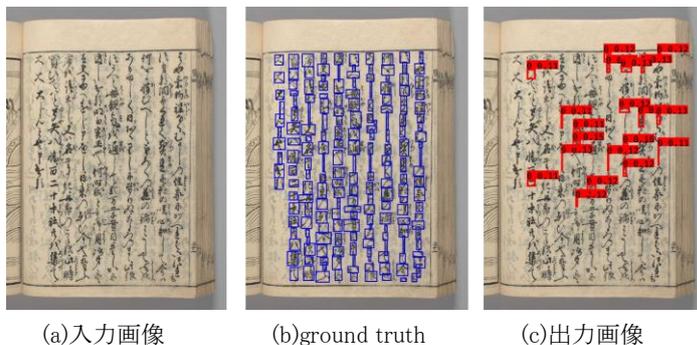
3. 古典籍画像中の文字検出

上述の物体検出の手法を用いて、古典籍画像から1文字ずつ文字の検出を行う。学習・テストには人文学オープンデータ共同利用センターの日本古典籍くずし字データセットを用いる[3]。本稿では、文字の検出のみを目的とし、すべての文字を同じクラスで学習を行う。学習用のデータとして、YOLOでは入力画像と文字を囲む矩形の座標、U-Netでは入力画像と文字を囲む矩形が白く塗りつぶされたマスク画像を与える。

4. 検出結果

YOLOを用いて学習・検出を行った結果を図1に示す。文字の検出精度は低く、多くの検出漏れが見られた。入力画像をいくつかの領域に分割するYOLOの性質上、文字と文字の間に詰まっている部分に対して検出精度が低くなっていると考えられる。

続いて、U-Netを用いて学習・検出を行った結果を図2に示す。文字の部分がマスクされており、ground truthに近い出力を得ることができた。検出結果の中には文字を囲む矩形が繋がっていたり、矩形が乱れていたりする部分が見られたため、さらなる精度向上が求められる。



(a)入力画像

(b)ground truth

(c)出力画像

図1. YOLOを用いた検出結果の一例



(a)入力画像

(b)ground truth

(c)出力画像

図2. U-Netを用いた検出結果の一例

5. まとめ

既存の物体検出の手法を用いて古典籍画像中から文字の検出を行った結果、YOLOでは検出精度が悪かったが、U-Netでは望ましい出力を得ることができた。

今後はネットワーク構造や学習方法を検討し、検出精度を向上させる予定である。また、検出とともに文字認識を行う。

参考文献

- [1] Joseph Redmon, Ali Farhadi. YOLOv3: An Incremental Improvement.
- [2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation
- [3] 日本古典籍くずし字データセット(国文研ほか所蔵/CODH加工), doi:10.20676/00000340, 人文学オープンデータ共同利用センター, <http://codh.rois.ac.jp/>