

電気・電子情報工学専攻	学籍番号	M183270	指導教員氏名	市川 周一
申請者氏名	村井 健			

論文要旨 (修士)

論文題目	物体検出によるくずし字認識の検討
------	------------------

近年、古文書をデジタルアーカイブ化して公開することが進められている。古文書は「くずし字」で書かれており、文字の区切りが不明確な形で表現されているため、一般の利用者は理解することができない。また、多くの古文書は手書きであり、個人の癖なども強いため、手書き文字認識技術による翻刻の完全自動化は現在も困難である。そのため、これまでは1文字ごとに切り分けられた画像をデータセットとして学習し、文字認識をしていた。

その後、2017年に開催された第21回PRUMアルゴリズムコンテストでは1文字の認識だけでなく、レベル別に3文字やそれ以上の文字数の認識を行った。また、2019年には人文学オープンデータ共同利用センターが主催のKaggle Kuzushiji Recognitionが開催された。ここで優勝した tascj のモデルは MMDetection を用いた物体検出によるくずし字の認識を行っていた。物体検出という方法は、これまでの文字認識の精度向上とは別の方向からアプローチしており、精度も高いことが示された。そこで本研究では、これまで文字の認識精度向上を図ってきた方法に対し、物体検出の方向から精度が改善されるのか検証する。また、自作した文字認識システムと物体検出での方法を比較する。

はじめに、文字認識によるくずし字認識を検証した。文字認識のネットワーク構造は、3層のCNNと2層のMLPを組み合わせたものを使用し、1文字ごとに切り分けたトレーニング画像683464枚を4212クラスに分類するように学習させた。1サイクルごとにテスト画像を入力して検証をしていた際には400サイクル(4エポック)後の精度は0.8660であった。

次に、物体検出によるくずし字認識を検証した。物体検出にはKaggle Kuzushiji Recognitionで優勝した tascj が使用していた MMDetection を用いた。物体検出方法はFaster R-CNNを用い、トレーニング画像3605枚を512*512のサイズに分割した133000枚を学習させた。10エポック学習後のモデルでテスト画像1736枚の画像を縦横0.5倍にした画像を入力したときの精度は0.6846、F1スコアは0.8128となった。また、テスト画像1736枚を1文字ずつ切り分け、学習済みモデルに入力したときの精度は0.7209、F1スコアは0.8378となった。

しかし、先行研究と比べると、今回作成した2つのモデルでは、どちらも先行研究のレベルには到達できなかった。また、Kaggle Kuzushiji Recognitionでは物体検出によるくずし字認識の方が認識精度が高かったのに対し、作成したモデルでは文字認識の精度の方が良くなっていた。文字認識では判定した文字がクラス分けできているのみを判定しているのに対し、物体検出では文字として検出をしたうえでクラス分けができていないと正解とならないためと考えられた。

今後の課題としては、文字認識と物体検出の性能の定量的比較方法の検討、学習手法(画像の前処理・使用する画像の取舍選択・ネットワーク構造)を変えた場合の比較、ハイパーパラメータの調整が上げられる。