1ZJ-01

ディープラーニングを使用した光源方向に依存しない点字文字認識

白川　太地†　　八木　悠河‡　　山下　晃弘‡　　松林　勝志‡

東京工業高等専門学校　機械情報システム工学専攻†　東京工業高等専門学校　情報工学科‡

1. はじめに

点字は，視覚障害者が情報を取得する貴重な手段として広く用いられている．一方で，盲学校や自治体では，視覚障害者が書いた手紙やテストの回答を確認する場合など，点字で記載された書類を健常者が読む必要がある場面が存在する．このような場面において，自動で点字をテキストに翻訳する需要が拡大している．

点字の自動認識に関する研究は，多様なアプローチで展開されているが，近年は，点字文書の凹凸を光学的手法を用いて認識する技術が主流になっている．文献[1]は，CNN(Convolutional Neural Network)の一種である Retina Net[2]を用いた物体検出技術を用いて，光学的点字認識システム AngelinaReader を開発し，オープンソースとして提供している．このシステムは高い認識精度を誇っている一方で，光の照射方向を固定する必要があり，上方からの光を受けていない点字文書は精度良く認識できない．

そこで本研究では，光の照射方向に依存しない高精度な点字認識技術の開発を目指す．具体的には，ディープラーニングを応用した新たな点字認識モデルの開発と，その効果の実証に取り組む．

2. 点字文書の画像認識と AngelinaReader

2.1 点字の基礎

点字は，1 文字が 6 点で構成されている．これらの点は凹凸で表現され，63 種類の文字を表現可能である．同じ 6 点の配置であっても，言語によって異なる文字として解釈される．また，一般に点字は一枚の紙の両面に印刷されることが多く，本研究で開発するシステムも両面点字に対応する．

Braille Character Recognition Independent of Light Source Direction with Deep Learning

Taichi Shirakawa†, Yagi Yuga‡, Yamashita Akihiro‡, Matsubayashi Katsushi‡

† Mechanical Information System Engineering, National Institute of Technology, Tokyo College

‡ Computer Science, National Institute of Technology, Tokyo College

2.2 AngelinaReader の概要

本研究では AngelinaReader が採用したモデルを改良する．このモデルの開発者は，紙の歪みやパースの変化にもロバストであることが特徴であると述べている[1]．AngelinaReader は上方からの光源を用いた画像を学習データとして使用しているため，認識を行う際にも同様の光源条件での撮影が必要である．

認識された点字は，6 点がそれぞれ 1～6 の数字に対応しており，凸点の数字の組み合わせで表現される．その後，定義された点字と通常の文字との対応表を用いて変換が行われ，変換された文字は元の画像上にラベルとして出力される．

2.3 物体検出技術と Retina Net

AngelinaReader は，Facebook AI Research (FAIR) が 2017 年に提案した Retina Net[2]を採用している．このモデルは Focal Loss という損失関数を用いた 1 段階モデルで，物体検出とクラス分類を同時に実施する．Focal Loss は，クラスの不均衡を考慮して設計された損失関数であり，分類が容易なクラスの影響を減少させ，難しいクラスに焦点を当てることができる．この結果として，Retina Net は一段階のモデルの認識速度を持ちながらも，二段階のモデルに匹敵する高い精度を持つ．

3. 本研究のディープラーニングモデル

3.1 ネットワークとフレームワークの選択

本研究は，AngelinaReader と同様に RetinaNet を採用した．実装には，Facebook AI Research (FAIR)が開発した PyTorch ベースの物体検出ライブラリである Detectron2 を使用した．Detectron2 の利点は，様々なモデル構造や学習手法を容易に実装・評価できることである．

3.2 データセット

本研究では，毎日新聞社が発行する両面点字の「点字毎日」を利用して独自のデータセットを作成した．各ページに対し光源を 30 度ずつずらしながら 12 回撮影し，合計で 86 ページ分，1,

032 枚の画像データセットを作成した．アノテーションは，AngelinaReader による認識結果を元に，アノテーションツールを用いて大きな誤りを手作業で修正し，データの精度を向上させた．最終的に，Detectron2 で使用するために，データ形式を COCO 形式に変換した．

学習フェーズでは，この「点字毎日」のデータセットと，AngelinaReader が学習に使用している DSBI データセット及び AngelinaDataset を組み合わせ，合計 1,430 枚の画像を含む統合データセットを使用した．

3.3 実験設定

本研究では，学習において，事前学習モデルとして Detectron2 が提供する retinanet_R_101_FPN_3x.yaml を使用し，ファインチューニングを実施した．データセットの 8 割を訓練データとし，残りの 2 割を評価データとして使用した．点字が重ならないことを優先的に考慮し，IOU の閾値を 0.02 に設定して NMS (Non-Maximum Suppression) を適用し，重なるバウンディングボックスを除去した．学習においては，Adam オプティマイザを使用し，学習率を 1e-4，バッチサイズを 32 とし，合計で 1500 エポックを実施した．この過程で，ReducelrOnPlateau を使用し，loss の改善が停止した場合に学習率を下げた．

データオーグメンテーションにおいては，入力画像に対してリサイズ，回転，クロップを適用した．リサイズは，入力画像の縦を 1200 ピクセルの±30% (840 ピクセル～1560 ピクセル) の範囲内でランダムに調整した．回転では，入力画像を±5 度の範囲内でランダムに回転させた．さらに，画像を 416×416 ピクセルの大きさにクロップして入力した．

アンカーボックスの設定は，データセット内の点字の大きさとデータオーグメンテーションによるリサイズを考慮し，最適なアンカーの大きさを計算した．その結果，アンカーのサイズを 25，35，45 とし，アスペクト比を 0.6 に設定し，これを全ての層で固定した．

4. 実験結果

4.1 評価方法と結果

点字の認識精度の評価には，mean Average Precision 50(mAP50)を使用した．この指標は，検出された点字と正解の点字との間の IoU(Intersection over Union)が 0.5 以上である場合に，正しく検出されたとみなす．

提案手法を評価データに適用した結果，評価データで mAP50=98.6%を達成した．これは，AngelinaReader の mAP50=99%には届かないものの，光源方向の変化に対しても提案手法による認識精度は一定であり，その有効性は確認された．

4.2 画像分割による精度向上

モデルの精度をさらに高めるために，入力画像を分割して予測する新しい方法を開発した．このプロセスでは，入力画像を特定の枚数に切り分け，その切り分けられた各画像をそれぞれ予測する．このアプローチにより，画像の詳細な予測が可能になり，物体検出の精度の向上が期待できる．さらに，NMS を用いて重複や誤検出を減らすことで，全体の精度が向上する．

本研究で学習したモデルに本手法を適用した結果，mAP50=98.8%までわずかに上昇した．これは，画像内の物体をより詳細に捉えることに成功し，わずかながらも性能の向上に貢献したことを示している．提案手法による予測結果の例を図 1 に示す．

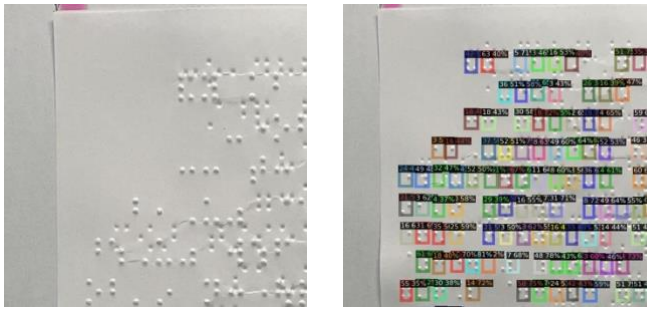

図 1: 予測前の点字画像と本モデルによる予測結果

5. 結言

本研究では，ディープラーニングを駆使して，光源方向に依存しない点字文字認識技術を開発した．具体的には，RetinaNet を基盤として，「点字毎日」から得たデータセットを用いて学習を進めた．現段階での認識精度は mAP50=98.8%であり，従来の手法と比較して劣る部分があることから，その向上が今後の課題となる．

参考文献

[1] Ilya G. Ovodov, Optical Braille Recognition Using Object Detection CNN, arXiv:2012.12412, 2020.

[2] L. Tsung-Yi, G. Priya, G. Ross, H. Kaiming, D. Piotr, Focal Loss for Dense Object Detection, arXiv:1708.02002, 2017.