

Fully-CNN を用いた近代公文書画像からの文字検出

高木裕平[†] 山田雅之[†] 目加田慶人[†] 長谷川純一[†] 中貴俊[†] 宮崎慎也[†]

[†] 中京大学

1 はじめに

省庁・自治体等が保管する近代公文書は、近世古文書の流れを汲む近代古文書であり、一般の行政職員が読むことは容易ではない。また、近代公文書は日本の史実を知るための歴史的史料でもあり、一般の国民や外国人研究者が利用できることが望ましい。そこで、我々は近代公文書の自動解読システムの開発を目指し、基礎データ収集と要素技術開発を進めている。台湾総督府文書は、終戦時に焼かれず原形のまま保管されていた近代公文書である。あらゆる種類の公文書を含み、その量は13,146 簿冊 (1 簿冊約 500 ページ× 400 字) にのぼることから、本研究はこれをデータ収集および実験の試料として利用する。この文書は不特定多数の筆記者による手書き文書であり、字体・字形バリエーションが多く、文字サイズや文字間隔も一定ではない。さらに、文字の上下が繋がっていることが多く、一文字に分割することが難しい。そこで本研究では人手により個々の文字を切り出して得られた字形データを用いて学習用データセットを作成し、文字および文字区切り位置を検出するフィルタを Fully-CNN により構成することを試みる。

本稿は学習用データセット作成方法、Fully-CNN のモデル、文字・文字区切り位置検出実験について述べる。

Character detection from modern official document images using Fully-CNN

[†]Yuhei Takagi, Masashi Yamada, Yoshito Mekada, Junichi Hasegawa, Takatoshi Naka, Shinya Miyazaki: Chukyo University

2 学習用データセット

現在、2,466 ページ分の原画像があり、その内 853 ページ分について字形データの作成が完了している。この字形データには、個々の文字に外接する矩形情報が含まれている。この矩形情報を用いて、文字画像、上区切り線画像、下区切り線画像の 3 種類の 1 チャネル画像を作成する。ここで上区切り線、下区切り線とは矩形の上辺と底辺に対応する。図 1-(c) は、(b) の矩形情報を用いて得られる 3 種類の画像を示す。学習用の教師画像は、これら 3 種類の画像をそれぞれ R,B,G に割り当てた 3 チャネル画像とする。本稿の実験では、853 ページのうち特徴の異なる 38 ページを人為的に選び、学習用データセットを作成する。入力画像、教師画像は、原画像および 3 チャネル画像を 256 × 256 に分割した大きさとし、入力画像および教師画像の組の総数は 640 である。図 2-(a),(b) に学習用の入力画像と教師画像を示す。

3 Fully-CNN

本研究は、文献 [1][2] を参考にして Fully-CNN を構築する。フレームワークには Chainer を利用する。ネットワークは、Convolution 層 (以下、Conv とする) と Deconvolution 層 (以下、Deconv とする) を持ち、各層で BatchNormalization による正規化を行う。活性化関数は、全ての層で ReLU を使用する。損失関数は 2 乗和誤差、最適化は Adam を使用する。本研究で使用するネットワークの構成を表 1 に示す。

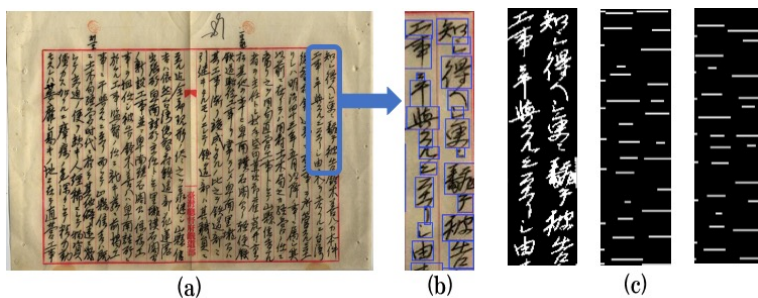


図 1: (a) 原画像, (b) 文字の外接矩形, (c) 文字画像, 上区切り線画像, 下区切り線画像

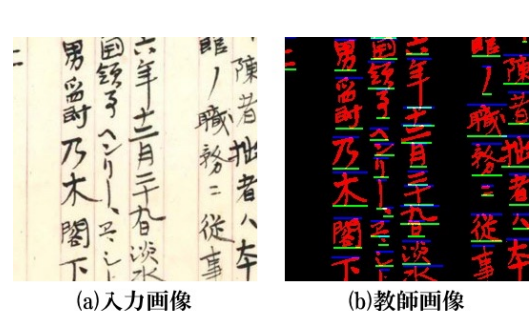


図 2: 学習用データ

表 1: ネットワーク構成

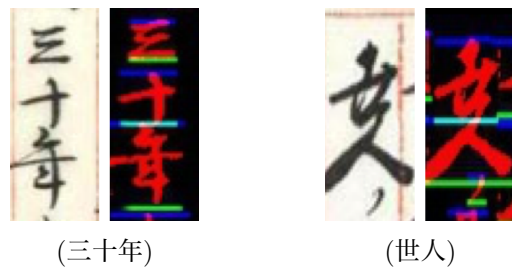
type	kernel	stride	padding	output size
input	-	-	-	$3 \times H \times W$
Conv	3	1	1	$32 \times H \times W$
Conv	4	2	1	$64 \times H/2 \times W/2$
Conv	3	1	1	$64 \times H/2 \times W/2$
Conv	4	2	1	$128 \times H/4 \times W/4$
Conv	3	1	1	$128 \times H/4 \times W/4$
Conv	4	2	1	$256 \times H/8 \times W/8$
Conv	3	1	1	$256 \times H/8 \times W/8$
Conv	4	2	1	$512 \times H/8 \times W/8$
Conv	3	1	1	$512 \times H/8 \times W/8$
Deconv	4	2	1	$256 \times H/4 \times W/4$
Conv	3	1	1	$256 \times H/4 \times W/4$
Deconv	4	2	1	$128 \times H/2 \times W/2$
Conv	3	1	1	$128 \times H/2 \times W/2$
Deconv	4	2	1	$64 \times H \times W$
Conv	3	1	1	$32 \times H \times W$
Conv	3	1	1	$3 \times H \times W$



図 3: 入力画像 (上) と出力画像 (下)

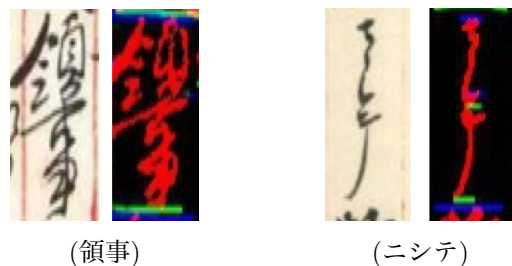
4 実験結果・考察

学習済みフィルタに原画像 2,466 ページを入力して得られる出力画像の例を図 3 に示す。出力画像からは、文字間隔が大きい場合でも小さい場合でも適切な区切り線が得られたことが分かる。図 3 の総文字数 367 に対し、正確に区切れた文字数は 99.2% の 364 であった。文字区切りが成功する割合は一般には入力画像に依存する。成功例と失敗例を図 4, 図 5 に示す。漢数字の三は一文字として正しく区切られている。また、多少の上下のつながりであれば正しく区切られている。しかし、大きく崩れ、上下が繋がった字は、区切り線を入れられていない。これは、学習データにこのような字形列がないことが原因としてあげられる。また、学習済みフィルタの内部パラメータは未調査であるが、水平方向の文字画素の有無だけでなく、垂直方向の文字画素の分布を考慮して、区切り線を入れていると推測できる。この結果から、大きく崩れた字に対して区切り線を入れる精度は低いが、そうでない字に関しては、文字区切りの検出が可能であり、文字自動切出しなどに有効である。



(三十年) (人)

図 4: 成功例



(領事) (ニシテ)

図 5: 失敗例

5 おわりに

本稿は Fully-CNN を用いた文字・文字区切り位置検出実験について述べた。今後、得られた区切り線に基づいた文字自動切出しの方法を検討する。

謝辞

本研究は第 45 回三菱財団人文科学助成, JSPS 科研費 JP17K03049 の助成を受けた。

参考文献

- [1] J.Long, E.Shelhamer, and T.Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, Vol. abs/1411.4038, 2014.
- [2] E.Simo-Serra, S.Iizuka, K.Sasaki, and H.Ishikawa. Learning to Simplify: Fully Convolutional Networks for Rough Sketch Cleanup. *ACM Trans. on Graphics (SIGGRAPH)*, Vol. 35, No. 4, 2016.