Recognition of Kuzushi-ji with Deep Learning Method: A Case Study of Kiritsubo Chapter in the Tale of Genji

Xiaoran Hu[*1] Mariko Inamoto[*2] Akihiko Konagaya[*1]

[*1] Tokyo Institute of Technology [*2] Keisen University

Reading ancient documents is one of fundamental works on the study of national literature. However, due to the use of kuzushi-ji (classical cursive handwriting characters) in the ancient documents, it requires a lot of knowledge and labor to read. This paper uses an End-to-End method with attention mechanism to recognize the continuous kuzushi-ji in phrases. Compared with the traditional recognition model with Connectionist Temporal Classification (CTC), our approach can get higher accuracy of recognition. The method can recognize phrases written by kana (47 different characters) with the accuracy 78.92% and recognize phrases containing both kanji (63 different characters) and kana with accuracy 59.80%.

1. Introduction

Kuzushi-ji (classical cursive handwriting characters) is general name for the outdated hiragana and kanji that are not used in school education of Japan since 1900. However, the works of Japanese classical literature, especially those before the Edo period, were almost written by kuzushi-ji. As a result, the classical literature can only be read by experts and most literary works are buried without being digitized. In order to deal with this problem, it is necessary to develop the method that aims to reprinting old ancient documents automatically. Recently, some methods were proposed to recognize kuzushi-ji, especially three-character string. The model that combines neural network with connectionist temporal classification won the challenge of recognizing kuzushi-ji (21th PRMU Algorithm Context) [1] in 2017. However, the performance of those models with various length continuous kuzushi-ji is not so good. This paper uses an End-to-End method that can recognize continuous kuzushi-ji with any length in phrases. The method uses the convolutional layers of VGG [2] and BLSTM [3] as encoder, using LSTM with attention mechanism [4] as decoder. To explain the result of two model, this paper use Grad-CAM [5] to visualize the network.

2. Method

2.1 Dataset

The training dataset of kuzushi-ji is from the Center for Open Data in the Humanities [6]. We select 47 kana and 63 kanji characters from 15 books as training dataset. The original training dataset is a set of images with single kana as shown in Fig1. In order to build arbitrary length character dataset, the original dataset does image banalization and combine single character images to form phrases. The model sets the input size of training images as width 32, height 300 pixels. To ensure kana on images are not deformed, the padding place of image is white.

The test dataset of various length continuous kuzushi-ji phrases is from a chapter of ancient roles, Kiritsubo, in the tale of Genji, which is written by unknown writer in Edo period. Separating images into many phrase images have been finished by hands. There are 137 images containing only kana and 159 images containing both kana and kanji.

The book of training dataset and test dataset are written by different persons. So the handwriting style is different between two datasets, which can decrease the accuracy of prediction. So we uses the phrases of the tale of Genji as labels of training dataset to reduce the differences between training and testing of our neural network models.

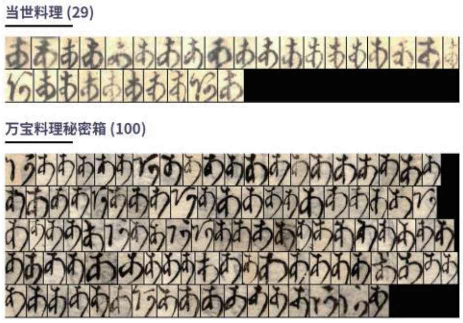


Fig1. Screenshot of original Training dataset「あ」
http://codh.rois.ac.jp/char-shape/unicode/U+3042/

2.2 Encoder: CNN + BLSTM

An original method for extracting the sequential features form images is to use convolutional neural network, however the native approach does not make use of the spatial dependencies between the features. So, we use the encoder that combines CNN and BLSTM to get the feature vectors from input images.

As shown in Fig2(a), the encoder first uses the convolutional layers to process the images to get robust and high-level features of images. In this process, two dimensional image converts to one dimensional feature map. A two-layer Bidirectional Long-short term memory (BLSTM) network [3] is applied after convolutional neural network to enlarge the range the feature sequences of input images.

2.3 Decoder: LSTM with Attention Mechanism

The decoder of model is LSTM with Attention mechanism. Connectionist temporal classification (CTC) uses a scoring function proposed in 2006[8], which deals with sequence problems where the timing is variable. However, CTC is limited in character recognition in phrases because it does not consider the dependencies between labels. Different from CTC method, attention based LSTM model can predict current label relying on

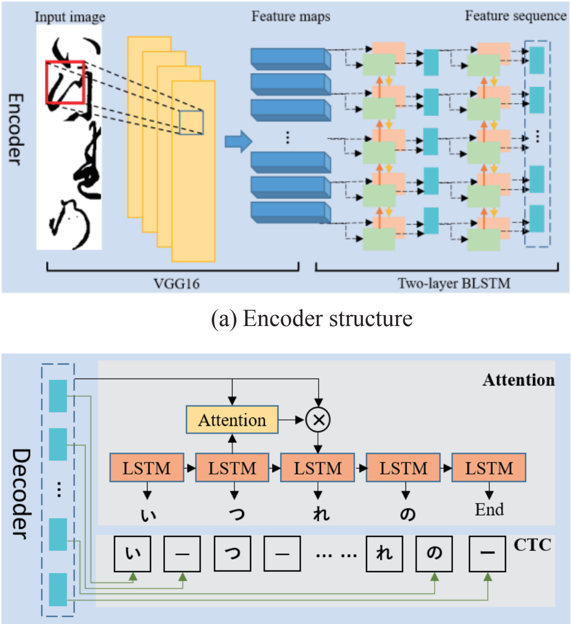
Contact: Akihiko Konagaya, The school of computing, Tokyo institute of Technology, kona@c.titech.ac.jp

the results of previous labels [7]. So, we adopted attention based model as decoder. In order to compare the performance of CTC and attention models, our decoder has the two models in its structure as shown in Fig2(b).

3. Experiments

3.1 Implementation

The structure of CNN in encoder is the convolutional layers of VGG16, which has 5 blocks. The memory units of each layer of BLSTM is 256. CTC based model and attention based model use the same encoder. To fit the input size of CTC model, there is a fully connected layer between BLSTM and CTC. The max step of attention model is set as 11 so that kuzushi-ji's recognition model can recognize up to 10 characters in a phrase. The parameters of CNN use pre-trained weights of VGG16, which are open on GitHub [8].



(a) Encoder structure

(b) Decoder structure of Attention and CTC

Fig2. Encoder and decoder structures of continuous kuzushi-ji recognition model

3.2 Prediction Accuracy on Test Dataset

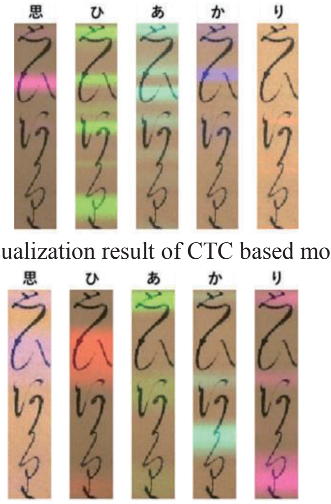
We perform two sets of experiments: one on kana dataset and the other on kana-kanji dataset. Table 1 shows the prediction results of different models. It indicates that the performance of attention model is much better than the one of CTC model. In the both models, the performance on kanji-kana dataset is not so good as kana dataset mainly due to the lack of sufficient number of kanji images in the original dataset.

Table1. The accuracy of prediction

Model	Kana dataset	Kanji-kana dataset
CTC	60.14%	48.16%
Attention	78.92%	59.80%

3.3 Comparison of CTC and Attention Model Results

Fig3 shows the difference of focusing points between two models when recognizing the phrases image "思ひあかり." Both model can give correct prediction, however, compared with CTC based model, attention based model can figure out more precise character positions on the image.



(a) Visualization result of CTC based model

(b) Visualization result of Attention based model

Fig3. Visualization result between two models: the highlight colors mean the positions where neural network focuses on when predicting a target character.

4. Conclusion

This paper proposes a deep learning method to recognize continuous kuzushi-ji phrases using the images of the tale of Genji. Compared with previous model, the proposed model with attention mechanism gets high accuracy of prediction.

Our future task is to improve the performance of recognizing kanji-kana mixed images by enhancing the decoding capability.

Acknowledgements

I would like to express my gratitude to the members of the Konagaya laboratory and Genji Pictorial DB Study Group for valuable comments and discussions in promoting this research.

References

[1] https://sites.google.com/view/alcon2017prmu/コンテスト結果

[2] Simonyan, Karen, and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science, 2014.

[3] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. ICASSP, 2013.

[4] Xu, Kelvin, et al. Show, attend and tell: Neural image caption generation with visual attention. International conference on machine learning. 2015.

[5] Selvaraju, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. ICCV, 2017.

[6] http://codh.rois.ac.jp/index.html.en

[7] Graves, Alex, et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. Proceedings of the 23rd international conference on Machine learning. ACM, 2006.

[8] Dabbish, Laura, et al. Social coding in GitHub: transparency and collaboration in an open software repository. ACM, 2012.