Recognition of Japanese Historical Hand-Written Characters Based on Object Detection Methods

Yiping Tang, Kohei Hatano, Eiji Takimoto Kyushu university

What is Kuzushiji?



Definition(*):

Kuzushiji is written with handwritten characters in Japanese historical literature.

• Difficulty in recognition:

(i) characters are often connected without explicit spaces

(ii) Characters are often simplified or abbreviated.

Segmentation is not easy

 Kuzushiji character of 「あ,a」 http://wwwap.hi.utokyo.ac.jp/ships/shipscontroller

https://www.nijl.ac.jp/pages/event/seminar/2015/old_books_tex



Recognition of Single Kuzushiji character

- Single Kuzushiji characters can be recognized with high accuracy by deep learning.
- [hayasaka+ 16]48 kinds of kuzushiji hiragana...70-80%
- [kitamoto16] 10 most frequent characters in CODH dataset ...96-97%
- PRMU2017 contest, 46 kinds of single kuzushiji...97.2%
- [Clanuwat+ 18] Kuzushiji-49 dataset... 97.33%

The background of Kuzushiji recognition --how to segmentation

• [Nguyen+ 17]



- 1. Find bounding boxes by multiple fixed size sliding windows.
- 2. Extract and process features using CNN, RNN.
- 3. Use CTC(Connectionist Temporal Classification) to derive the result.

Problem:

- 1. The predicted boxes in result will be some fixed size, and cannot fit the shape of character.
- 2. There will be lots of bounding boxes that only circle the part of character but seem as a full character.

The background of Kuzushiji recognition --how to segmentation

- [Kitamoto+ 19]
- 1. Create an annotation dataset for pixel units for learning
- 2. Train by U-net network
- 3. Predict the label of each pixel in full book page







Problem:

1. Need annotation data of

each pixel

- 1. Hard to train
- 2. Take up lots of memory

Our approach(1): Segmentation and recognition at the same time based on object detection method



Input:

digital image

• Output:

pair of label and bounding box

for each object



• Problem: How can we obtain learning data with segmentation information?

Kuzushiji segmentation dataset[Tang+18]



Character segmentation information







segmentation dataset

Character segmentation information, Use for learn

- Base on CODH dataset and PRMU contest dataset, have segmentation information and label information of image of each character.
- 77953 three-letter images and 12582 multi-letter images
- Removal of difficult data or erroneous data by double check by manual operation

Proposed method 1

- - get bounding box and label confidence information simultaneously
- Apply object detection[Redmon+ 18] to recognition of Kuzushiji



{bounding box1,label confidence 1}
{bounding box2,label confidence 2}
{bounding box3,label confidence 3}
{bounding box4,label confidence 5}
{bounding box6,label confidence 6}
{bounding box1,label confidence 7}
{bounding box2,label confidence 2}
{bounding box3,label confidence 3}
{bounding box4,label confidence 5}
{bounding box5,label confidence 5}
{bounding box5,label confidence 5}
{bounding box6,label confidence 6}
{bounding box7,label confidence 7}

• The darknet53 model is used as backbone network.

Aggregation method of Yolov3



--non maximum suppression (NMS)

- 1. Set the label confidence threshold and the overlap threshold.
- 2. Find the highest score box without repeating
- 3. Two proposals are considered to be in the same cluster when their IoU(Intersection over Union) is larger than the overlap threshold, only keep the one with the highest score in the cluster.
- 4. Loop 2, 3 until there are no new box can be find

Problem:

- 1. Unable to guarantee the number of output characters
- 2. Bad handling overlay problem of characters



Proposed method 1

-aggregation method



{coordinate/label info...} {coordinate/label info...}

- 1. Record the center of each box.
- 2. Assume the number of clusters of Kuzushiji characters as K.
- 3. A box with a maximum label confidence of character in each cluster regarded as the representative.

Advantage: Since a plausible box is selected for each character cluster, recognition is rarely discarded or passed.

Evaluation criteria for bounding boxes

Given the sequence of predicted bounding boxes $(\hat{x}, \hat{y}, \hat{w}, \hat{h})$ and ground truth bounding boxes (x, y, w, h), the consistency rate (CR) of the predicted sequence of boxes is defined as for formula.



CR only focuses on differences of bounding boxes in the vertical direction, which is sufficient for our purpose.

• Training

70,000 images(three characters) from dataset[Tang+18] for training.

evaluation

Other 7,000 images(three characters) from dataset[Tang+18] for testing.

Results

		LER	0.07-Y-CR	0.05-Y-CR
1	Histogram+Resnet53	36.6%	63.46%	55.96%
2	[Nguyen+ 17]	12.88%		
3	YOLOv3	6.65%	75.8%	66.9%
$\check{4}$	FGDM-a	5.24%	81%	71.3%
(5)	FGDM-b	2.5%	83.18%	73.3%

④FGDM-a is denoted as the result of FGDM with the same learning rate of YOLOv3 and ⑤FGDM-b is the one with decreasing learning rate by multiplying 0.1 in every 40000 rounds.

Future work

Recognition for Kuzushiji images of more than three characters(Lv3).



(Use original YOLOv3)





• Thanks