KuroNet and Motivating the Kaggle Competition

Pre-Modern Japanese Kuzushiji Character Recognition with Deep Learning

Alex Lamb*, Tarin Clanuwat*, Asanobu Kitamoto

What is the right task for Kuzushiji?

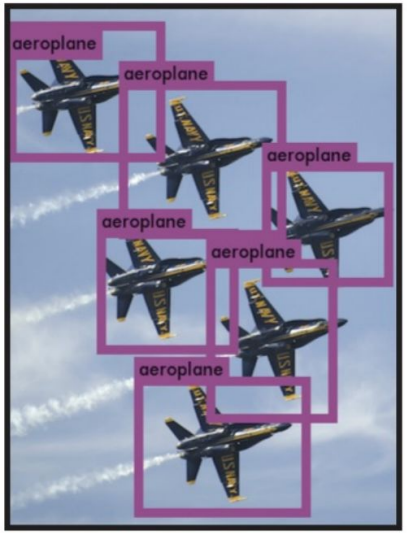
- **Stage 1 : predict character centers on each page**
  - We have the data, but slightly less than text sequence data

- Stage 2: predict a text sequence for each page
  - We have the data
  - How to extract the sequence for a given page?
  - Will it make the model's task too hard?

- Stage 3: predict translation to modern japanese
  - Where to get a dataset with ground truth translations?
  - The right translation is subjective and could be very hard to predict
  - High risk of destroying original meaning and poetry

Object Detection Algorithm

- Supervised task:
  - Image ➜ set of objects in the image
- Examples of object detection:



What model is good for Kuzushiji?

- Local Patches from a Page are hard to understand in isolation.

?

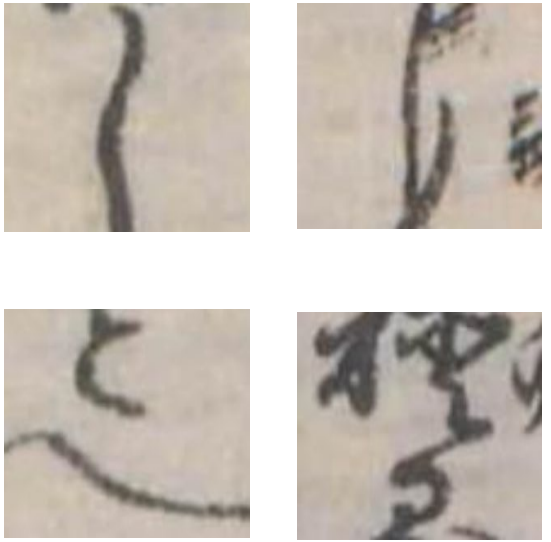
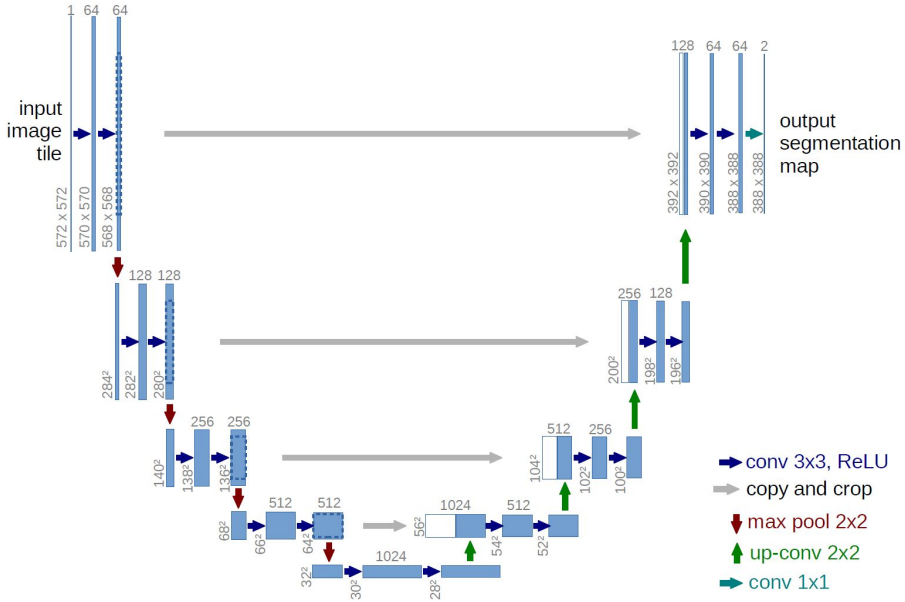


Image Patches

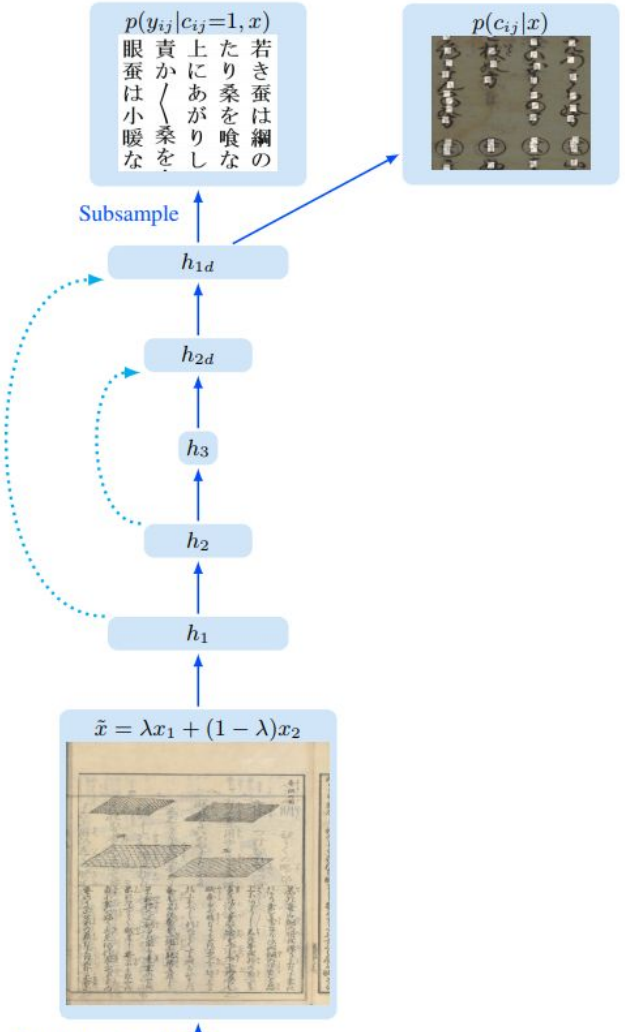
U-Net Architecture

- A special type of neural network which combines small-scale and high-level information



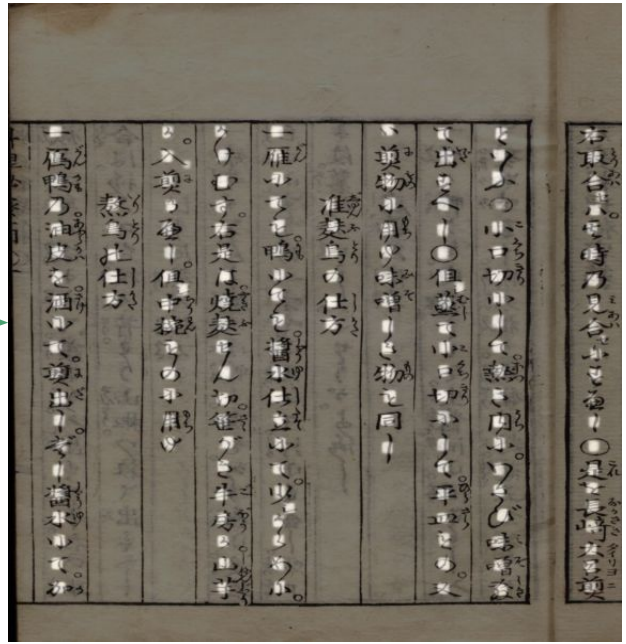
KuroNet

1. A Residual U-Net takes a page image as input
2. Produces a feature at each pixel position
3. At every position a binary classifier for characters
4. At positions with characters, predict the character



Example of Character Center Prediction

Ground Truth Character Centers

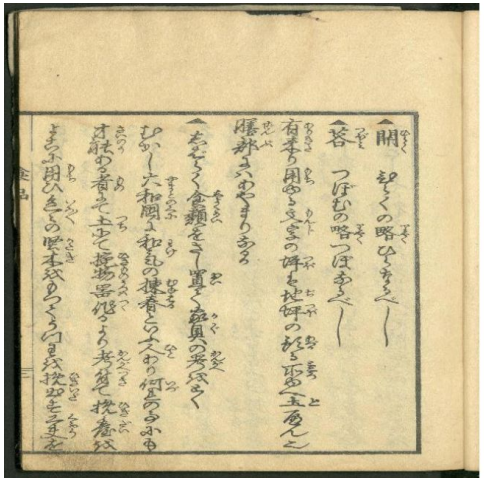


Predicted Character Centers

Mixup Regularizer

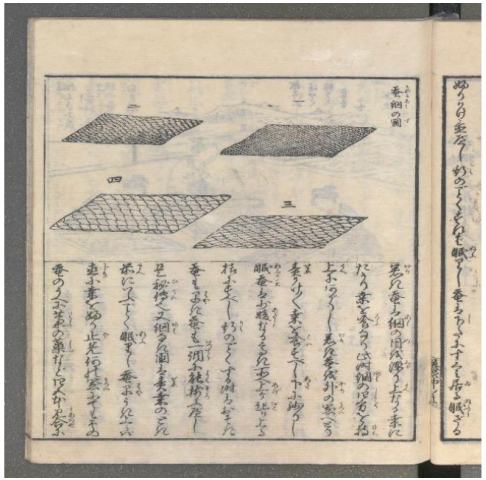
- Train on linear interpolations of input images.
- Mimics the “bleed through” of the adjacent page - trains model to ignore it.

30%

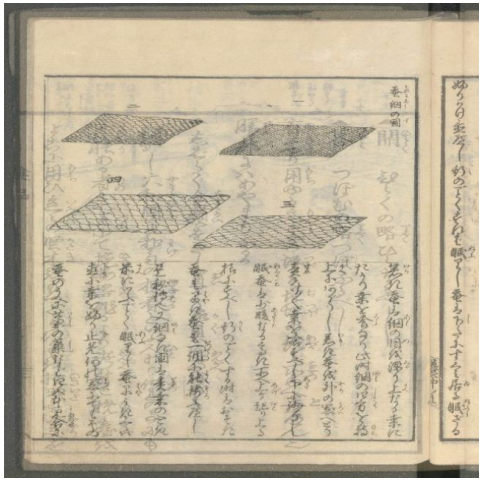


+

70%



=



F1-score Evaluation Metric

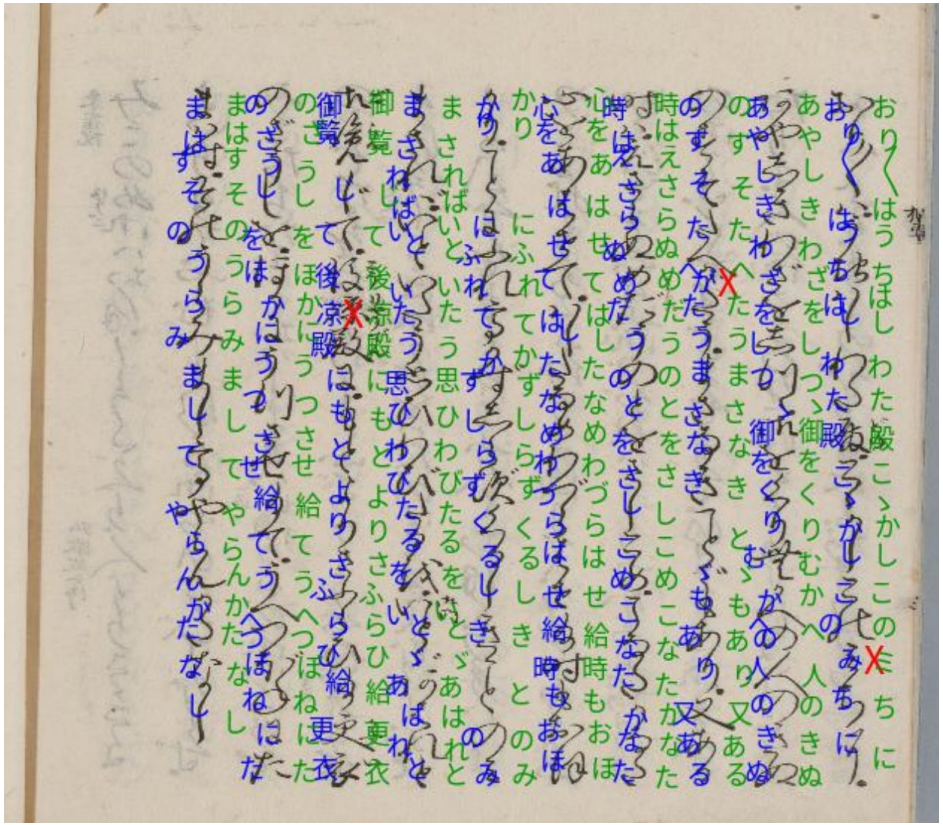
- What kinds of mistakes can our model make?

	True Character is X	True Character is Y	No Character Present
Predict X	True Positive	False Positive	False Positive
Predict Y	False Positive	True Positive	False Positive
Predict Nothing	False Negative	False Negative	Not counted

- False positives hurt precision, False negatives hurt recall

- F1-score is a special (harmonic) averaging of precision and recall.
  - Intuitively both precision and recall need to be high for F1-score to be high.

Results



On kaggle dataset:
F1-score of 90.2

Green Predicted

Blue Ground Truth

Red Checkmark for Errors

What did we learn from KuroNet?

- An approach based on predicting bounding boxes while looking at a whole page can work well.
- On most documents it is possible to get fairly high accuracy.
- Predicting where characters are can be just as challenging as predicting a character’s identity.
- In short: framing the task as *object detection* is reasonable.

Want to learn about or use KuroNet?

- We released KuroNet API online that anyone can use!
  http://codh.rois.ac.jp/kuronet/
- We published a paper at ICDAR 2019 on KuroNet:
  - https://arxiv.org/abs/1910.09433

Kaggle Competition

- KuroNet achieved reasonable results but didnt solve the task, motivating us to open a competition hosted on kaggle

- Ran for 3 months:
  - 2652 submissions
  - 338 competitors
  - 293 teams
  - 15k prize pool (split across top 5)

Kaggle Solutions

- Top F1-score of 0.95.
- 11 teams achieved a better F1-score than KuroNet.
- 50 teams scored above 0.80 (fairly usable systems)

#	△pub	Team Name	Notebook	Team Members	Score	Entries	Last
1	—	**tascj**			0.950	13	25d
2	—	**Konstantin Lopuhin**			0.950	60	23d
3	—	**Kenji**			0.944	161	23d
4	▲1	**YoudaoOCR**			0.942	49	23d
5	▼1	**See--**			0.940	42	25d

Brief Overview of Top-5 Solutions

- Tascj used a Cascade-RCNN and a small ensemble
- Konstantin Lopuhin used a Faster-RCNN to segment and a classifier stage
- Kenji used a Faster-RCNN to segment and a classifier stage
- Kolman segmented into text columns and used an LSTM with CTC to recognize characters
- See-- used a modified CenterNet

What did we learn from Kaggle?

- Some existing object detection algorithms work well
  - Faster R-CNN and Cascade R-CNN produced excellent results without any Kuzushiji-specific techniques.

- At the same time, other techniques struggled.
  - YOLO performed quite badly despite substantial effort.
  - “CenterNet” performed well but required more effort and domain-specific tuning to get working.

- Several leading approaches had models that performed detection and classification jointly.

Questions?

- Contacts:
  - Alex Lamb: lambalex@iro.umontreal.ca
  - Tarin Clanuwat: tarin@nii.ac.jp
  - Kitamoto Asanobu: kitamoto@nii.ac.jp

- Feel free to send us an email if you have any questions.