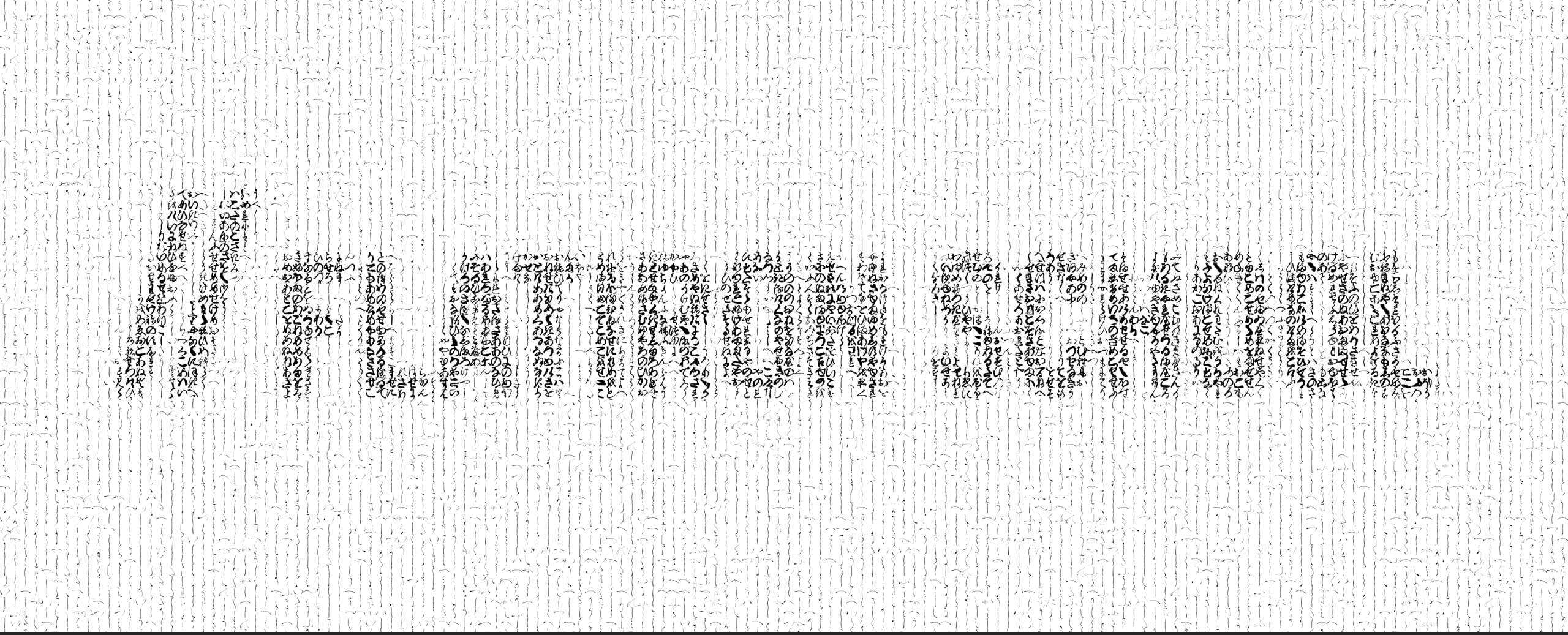
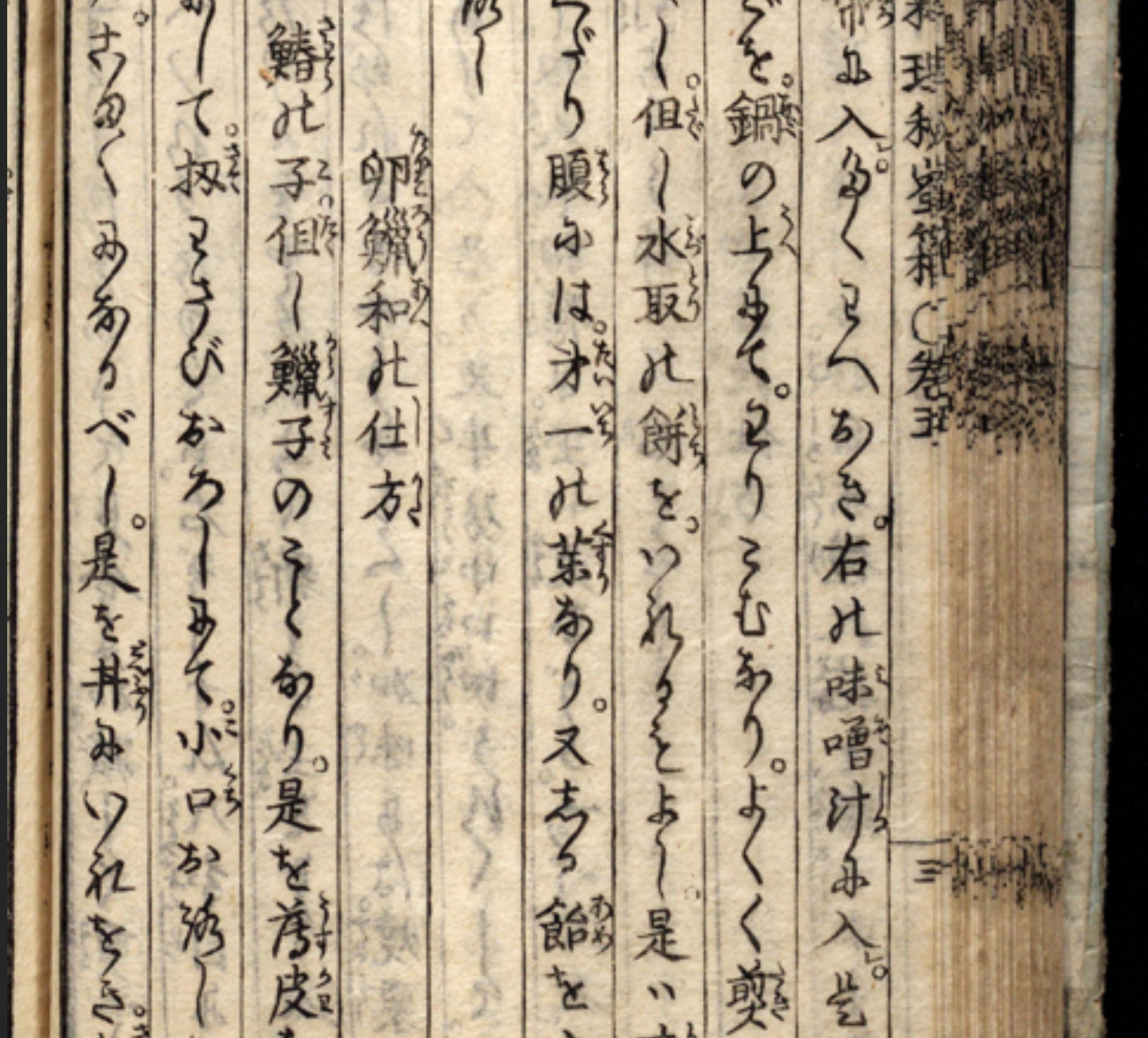


Japanese *Kuzushiji* Character Classification

STEPHEN LANIER

Problem Statement

Japanese characters, of which there are more than 4,000, are especially hard to recognize as *kuzushiji*, cursive–font characters. Can machine learning help us identify them?



Value

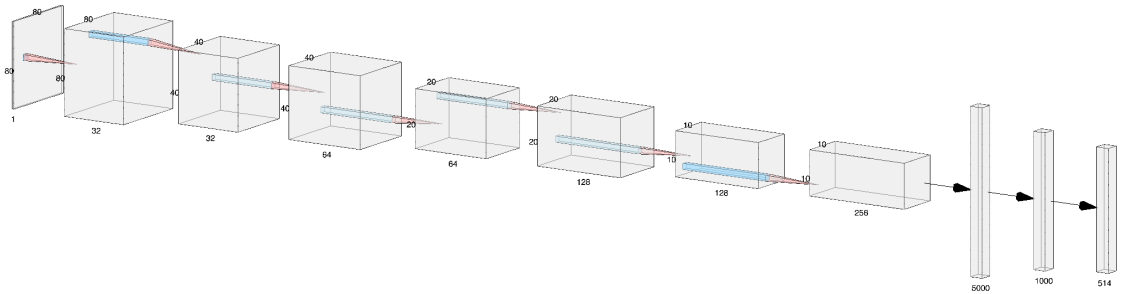
"Japan has millions of books and over a billion historical documents… [and yet] there are very few fluent readers of kuzushiji *today (only 0.01% of modern Japanese natives)."*

–Kuzushiji Recognition, Kaggle

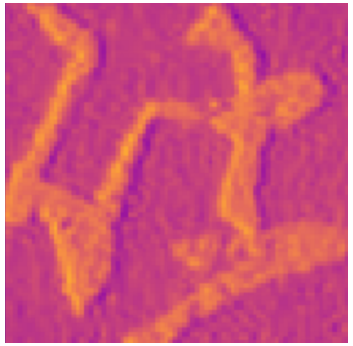
Methodology

Prep

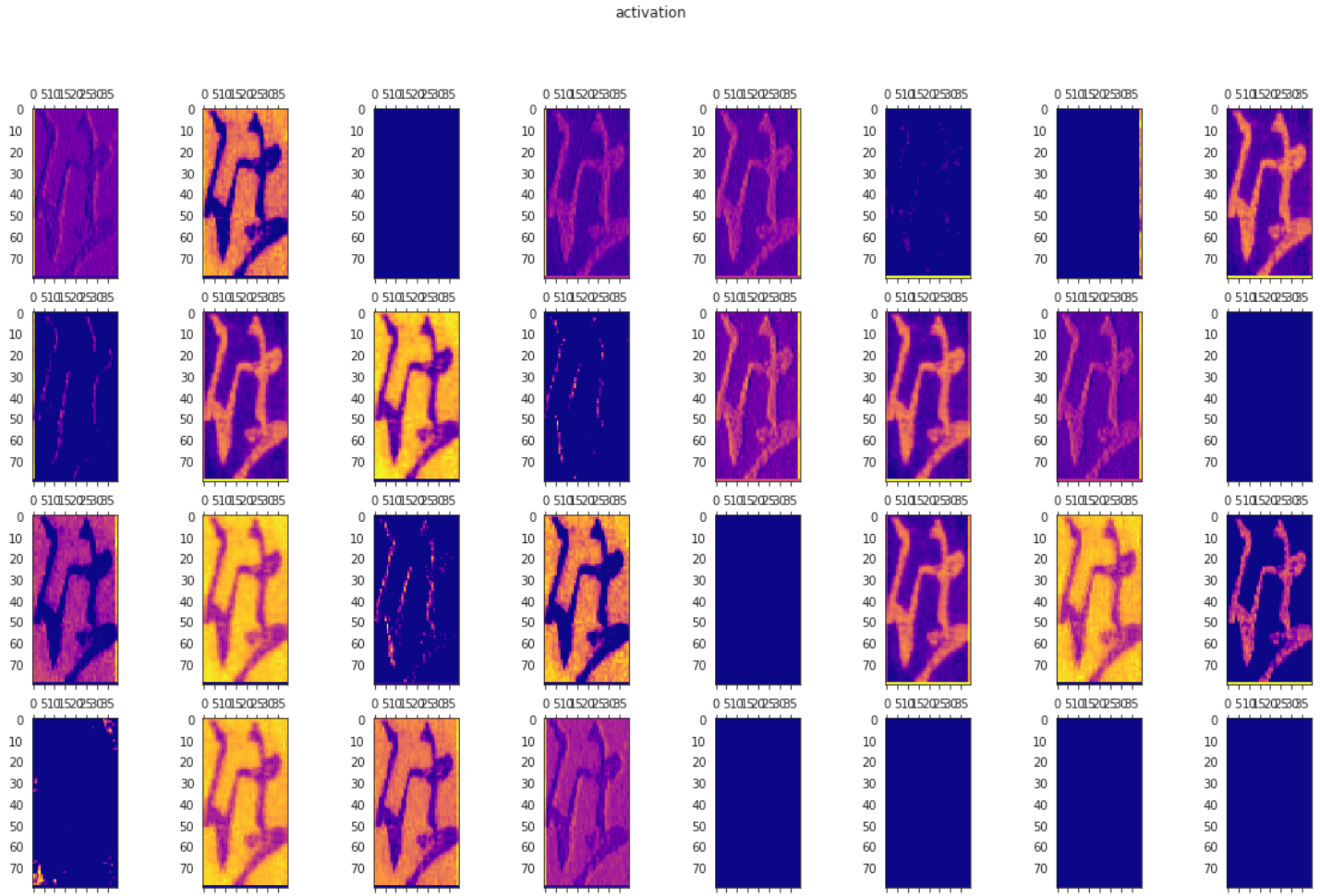
Train



Interpret



Results



Results

Test Accuracy				
Baseline	Xception	ResNet152V2	InceptionResNet	FCN
0.95	0.85	0.80	--	0.65

```
火 (55):        ['人', '大', '左', 'rare', 'rare', 'rare']
為 (58):        ['ゐ', 'ゐ', '湯', 'rare', 'rare', 'rare']
無 (31):        ['む', 'む', '共', '知', 'rare', 'rare', 'rare']
然 (28):        ['rare', 'rare', 'rare', 'rare']
焼 (52):        ['rare', 'rare', 'rare', 'rare']
煎 (37):        ['直', 'rare', 'rare', 'rare', 'rare']
煮 (96):        ['者', '黄', 'rare', 'rare', 'rare']
爰 (15):        ['rare', 'rare', 'rare']
父 (27):        ['と', 'rare']
牛 (26):        ['め', '半']
物 (243):       ['ま', 'ゐ', '類', 'rare', 'rare', 'rare']
玉 (71):        ['に', 'の', 'ば', 'む', 'も', '王', '給', '給']
王 (25):        ['つ', 'に', 'に', 'に', 'ヱ', '玉', '玉', '玉']
```

Results

Conclusions

Model is able to classify characters with high accuracy (approaching human-level ability)

High success at this level means possibly can incorporate more characters

Simplest model currently has best performance, may change with more experimentation

Future Work

- More work on the FCN
- A dashboard for easier interaction
- Showcase predictions on image (possibly with semantic segmentation)

ACKNOWLEDGEMENTS

Thank you!

Special thanks to Jacob Eli Thomas, Victor Geislinger, and Jeff Herman, my instructors at Flatiron School, for their encouragement, instruction, and guidance.