12.18.24

Multiple-Character Optical Character Recognition

Nathaniel Liu, Chandler Nelson, Jakub Piwowarczyk

Introduction

- Optical Character Recognition (OCR) attempts to process an image of some characters and match it to a predetermined list of symbols in an alphabet.

- Requires the translation of analog world to digital information

- The concept can be traced back to the late 1920's, early 1930's - updates in computational power made it possible nowadays

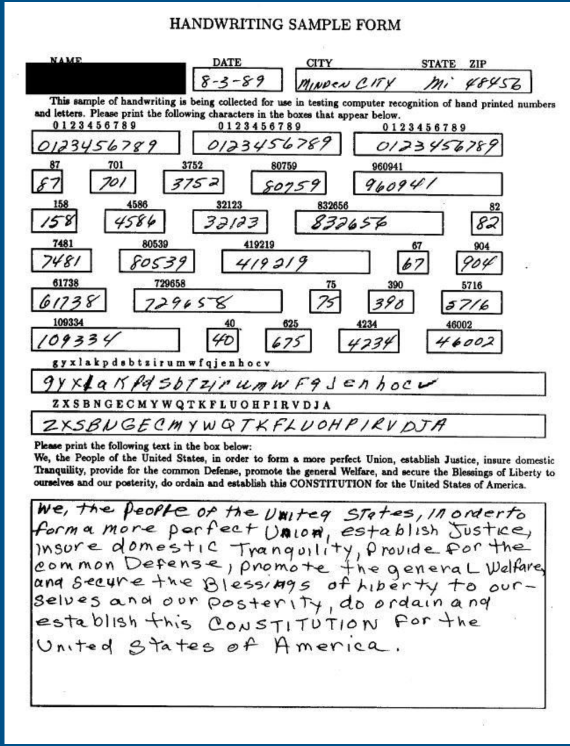
- OCR is used in many technologies nowadays, from scanned document transcription to on-the-fly translation

Our Problem/Topic

- Single Character OCR:
  - Given a certain number of possible states, classify to the most likely one
- Multiple Character OCR:
  - Since letters/numbers are normally in strings, attempt to read the string together

Dataset Overview

- Special Database 19 contains NIST's comprehensive corpus of training materials for handwritten document and character recognition.
- Key Features:
  - Uppercase letters, lowercase letters, and number
  - Sample forms collected from 3,600 writers.
  - A total of 810,000 character images, each annotated with ground truth classifications.



HANDWRITING SAMPLE FORM

NAME [redacted] DATE 8-3-89 CITY MINDEN CITY STATE MI ZIP 48456

This sample of handwriting is being collected for use in testing computer recognition of hand printed numbers and letters. Please print the following characters in the boxes that appear below.

0123456789 0123456789 0123456789

87 701 3752 80759 960941

158 4586 32123 832656 82

7481 80539 419219 67 904

61738 729658 75 390 5716

109334 40 625 4234 46002

g y x l a k p d s b t z i r u m w f q j e n h o c v

Z X S B N G E C M Y W Q T K F L U O H P I R V D J A

Please print the following text in the box below:
We, the People of the United States, in order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common Defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our posterity, do ordain and establish this CONSTITUTION for the United States of America.

Methodology: OCR Architecture (RF)

- Features Utilized:
  - Histogram of Oriented Gradients (HOGs)
    - 9 Directions
    - 8x8 pixels per cell
    - 2x2 cells per block
    - L2-Hys Norm
    - Gradients in a particular direction
  - Hu Moments
    - Invariant to shift and scale
- Random Forest
  - Trained with 100 Estimators
  - Split Criterion: Gini
  - Max features in single split: 8

Methodology: OCR Architecture (DL)

- Model Architecture
  - Input: 32×32 grayscale image.
    - Resized, converted to grayscale and normalized
  - CNN Feature Extractor:
    - Conv1 (1,32,32 → 32,32,32), kernel size = 3, ReLU.
    - Conv2 (32,32,32 → 64,32,32), kernel size = 3, ReLU, MaxPooling.
    - FC1 (64*16*16 –> 128).
    - FC2 (128 → 62).

Methodology: OCR Architecture (DL)

- Model Architecture
  - Input: 32×32 grayscale image.
    - Resized, converted to grayscale and normalized
  - CNN Feature Extractor:
    - Conv1 (1 → 32), ReLU.
    - Conv2 (32 → 64), ReLU, MaxPooling.
    - FC1 (64*16*16 –> 128).
    - FC2 (128 → 62).

Single-Character OCR (DL)

Hyperparameters

- Batch Size: 128
- Number of Epochs: [10, 30, 50]
- Learning Rate: [1e-3, 1e-4, 1e-5]
- Optimizer: Adam
- Loss Function: Cross Entropy

Dataset

- 810,000 Samples
- Each samples is a 32 x 32
- 0.8 Train Ratio
- 0.1 Validation Ratio
- 0.1 Test Ratio

Methodology: Multiple-Character OCR Architecture

- Input: 32×256 grayscale image.
- CNN Feature Extractor:
  - Conv1 (1 → 64), BatchNorm, ReLU, MaxPooling.
  - Conv2 (64 → 128), BatchNorm, ReLU, MaxPooling.
- LSTM Sequence Model:
  - BiLSTM with 256hidden units and 2 layers.
- Fully Connected Layer:
  - Maps LSTM outputs to 62 output classes
- CTC Loss:
  - Aligns predicted sequences with ground truth labels.

Multiple-Character OCR

Hyperparameters

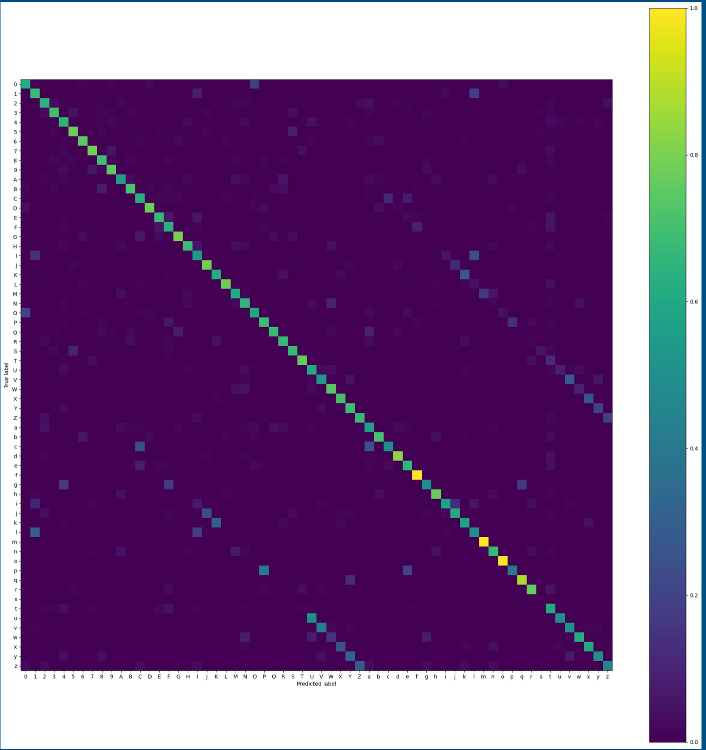
- Batch Size: 64
- Number of Epochs: 15
- Optimizer: Adam
- Learning Rate: 0.001
- Loss Function: CTC Loss

Dataset

- 50000 generated samples
- Each samples is a 32 x 256
- 0.8 Train Ratio
- 0.1 Validation Ratio
- 0.1 Test Ratio

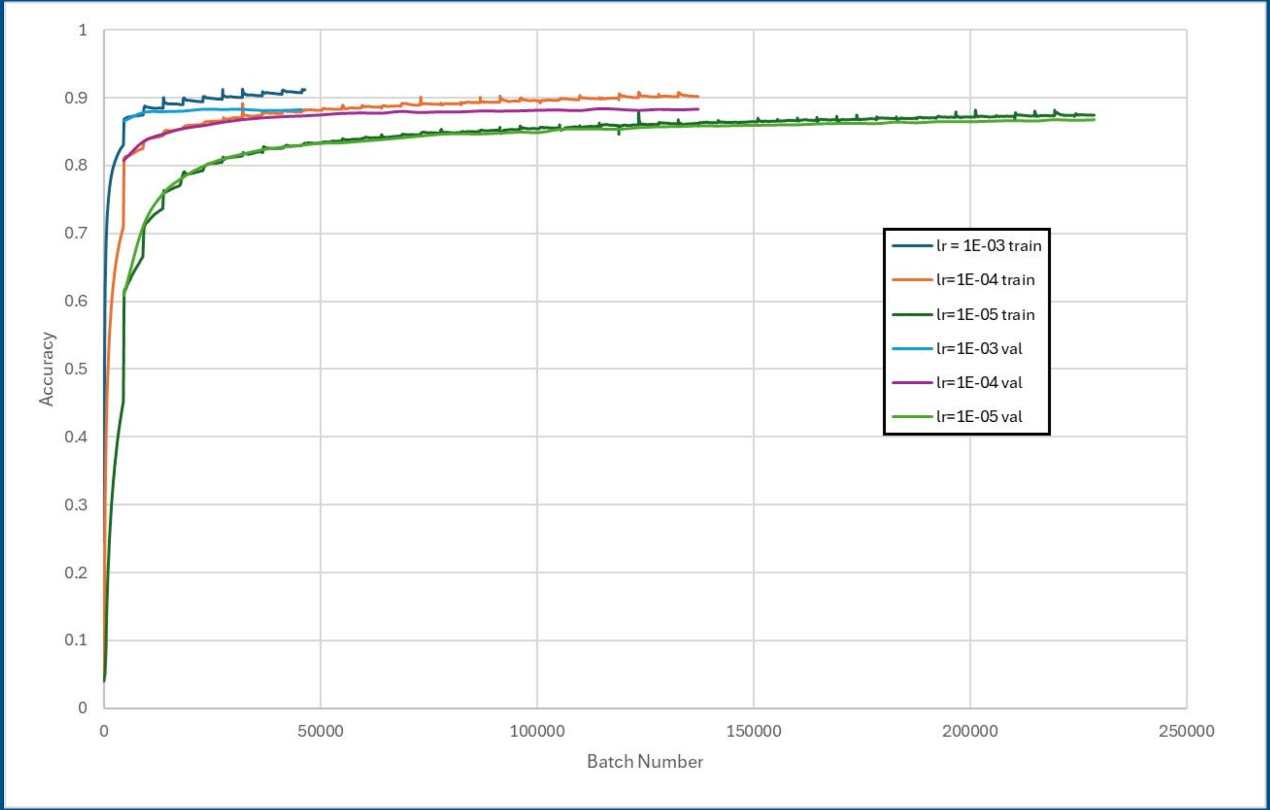
Results: Single-Character OCR (RF)

- Testing Accuracy of 68.2%
- Considerable confusion between some upper and lower case letters
- Could not seem to get lower case “s” at all



Confusion Matrix of Random Forest Classifier

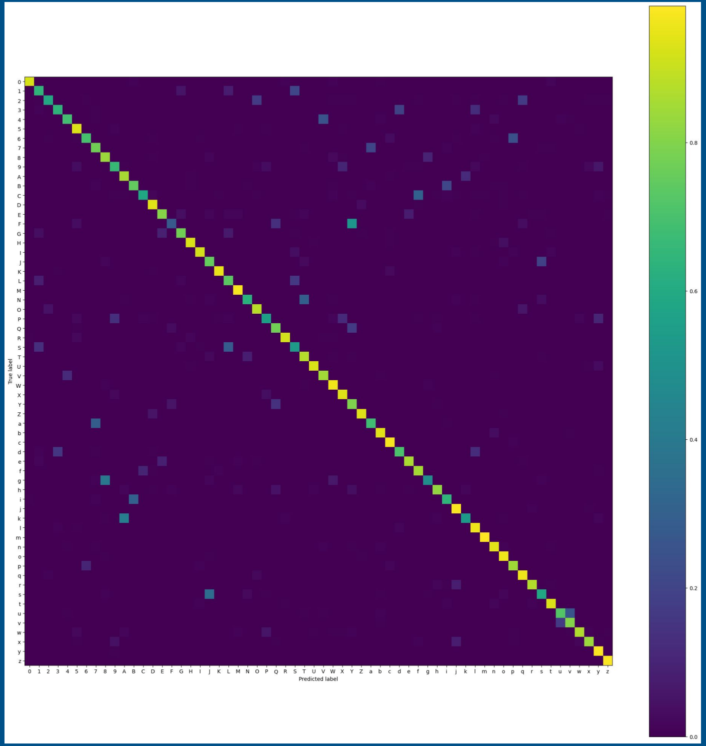
Results: Single-Character OCR (DL)



Training and Validation Accuracy over 50 Epochs

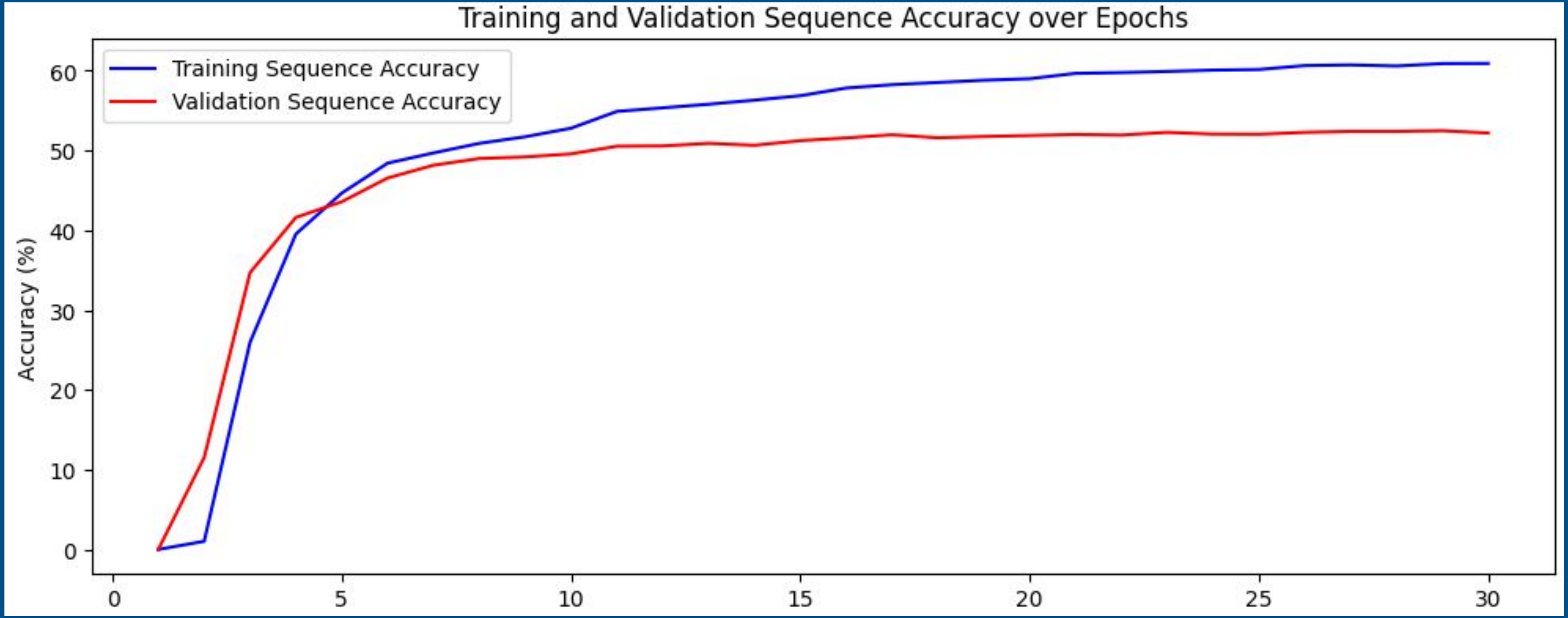
Results: Single-Character OCR (DL)

- Testing Accuracy of 88.2%
- Slight confusion with characters like “8” and “g”



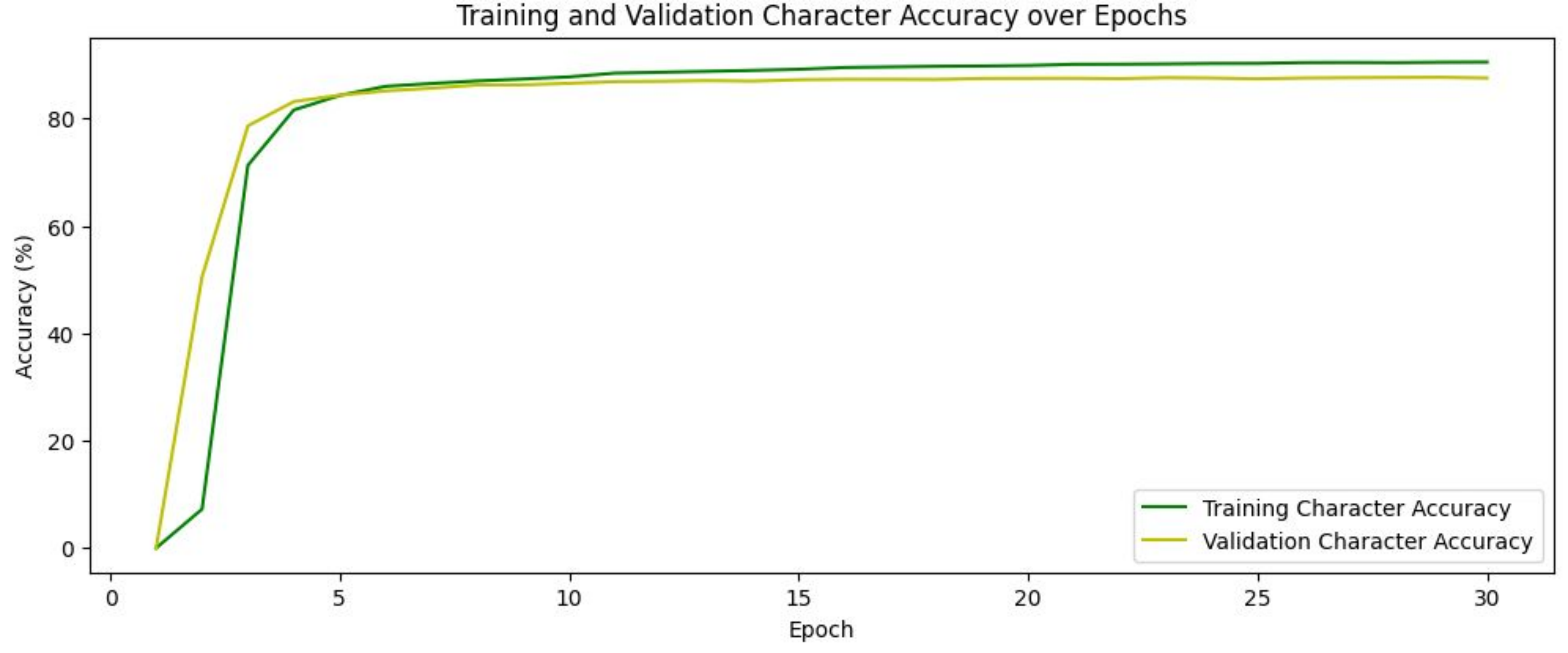
Confusion Matrix of LR=1e-4 Network

Results: Multiple-Character OCR



Training and Validation Sequence Accuracy over 30 Epochs

Results: Multiple-Character OCR



Training and Validation Character Accuracy over 30 Epochs

Additional Results: Multiple-Character OCR

On the test set, the model scored 87.76% character accuracy, but only a sequence accuracy of 52.88%.

Implications/Interpretation

- Single-Character OCR
  - Random Forest model does not perform as well as Deep-Learned model
    - Likely due to scale invariance in Hu Moments
  - Deep-Learned model performed quite well, additional data augmentation and context in word may push performance further
- Multiple-Character OCR
  - Model is able to learn character-level predictions and achieves relatively strong performance for multi-character sequences.
  - The lower sequence accuracy suggests challenges in modeling dependencies across longer sequences.

Challenges

- Limited computing resources.
- Dataset lacked characters in sequences, had to generate synthetic dataset for multiple character OCR.
- Difficulties getting Multiple-Character OCR to achieve higher sequence accuracy.

Future Work

- Investigate dataset augmentation and more intelligent loss functions
- Next steps would be to train the Multiple Character OCR on data more representative of how people write in real life
- Explore transformer-based architectures
- Try with more complex alphabets

Sources

1. Shunji Mori, Ching Y. Suen, and Kazuhiko Yamamoto, "Historical review of OCR research and development," *Proceedings of the IEEE*, vol. 80, no. 7, pp. 1029-1057, July 1992. Available: https://doi.org/10.1109/5.156468.
2. Lawrence O'Gorman, "The document spectrum for page layout analysis," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 2422, pp. 6-16, 1995. Available: https://doi.org/10.1117/12.373511.
3. National Institute of Standards and Technology, "NIST Special Database 19," [Online]. Available: http://doi.org/10.18434/T4H01C.
4. M. Marjani, M. Mahdianpari, and F. Mohammadimanesh, "CNN-BiLSTM: A Novel Deep Learning Model for Near-Real-Time Daily Wildfire Spread Prediction," *Remote Sensing*, vol. 16, no. 8, p. 1467, 2024. Available: https://doi.org/10.3390/rs16081467.
5. A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks," *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, USA, pp. 369-376, June 2006. Available: https://www.cs.toronto.edu/~graves/icml_2006.pdf.

Thank You