Optical Character Recognition for the E-Log

Justin Rower, CCI 2020

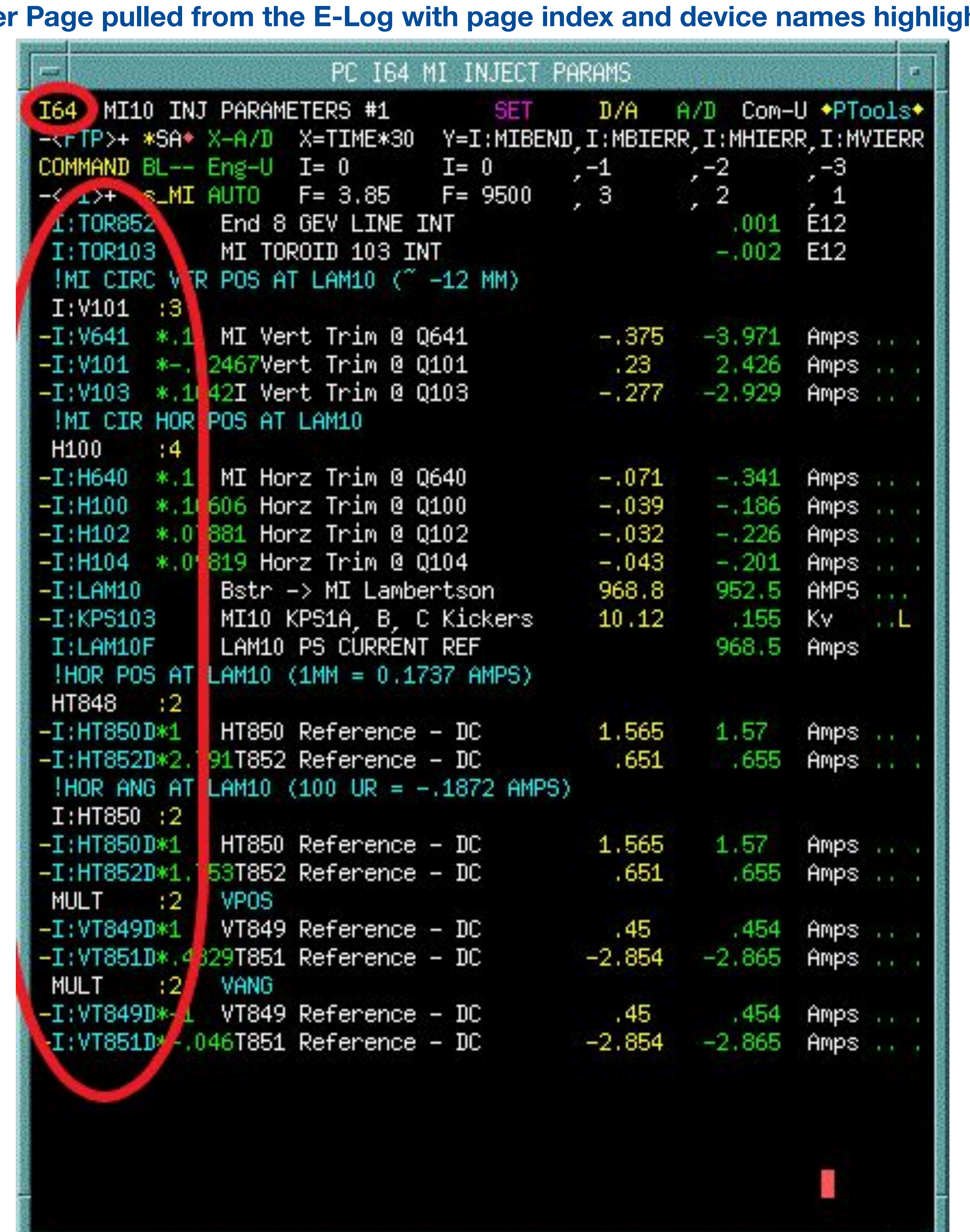
Abstract

The Main Control Room (MCR) E-Log has over 200k images that contain information which could be used to understand the history of the MCR. However, the images are not currently searchable. We set out to automatically extract searchable information from the images using a bidirectional recurrent neural network (RNN) fed with convolutional feature maps. Although limited accuracy was achieved, future directions of research should lead to better performance.

Preliminary Categorization and E-Log

- Focused on parameter pages
  - Text heavy
- Large potential label space for device names
  - More than 36^{14} possibilities ($6.14*10^{21}$)
  - Very data heavy
- Page index label space is much smaller
  - 26*11*11*11 = 34,606 possible labels
  - Very high information per label

Parameter Page pulled from the E-Log with page index and device names highlighted



Tools and Architecture

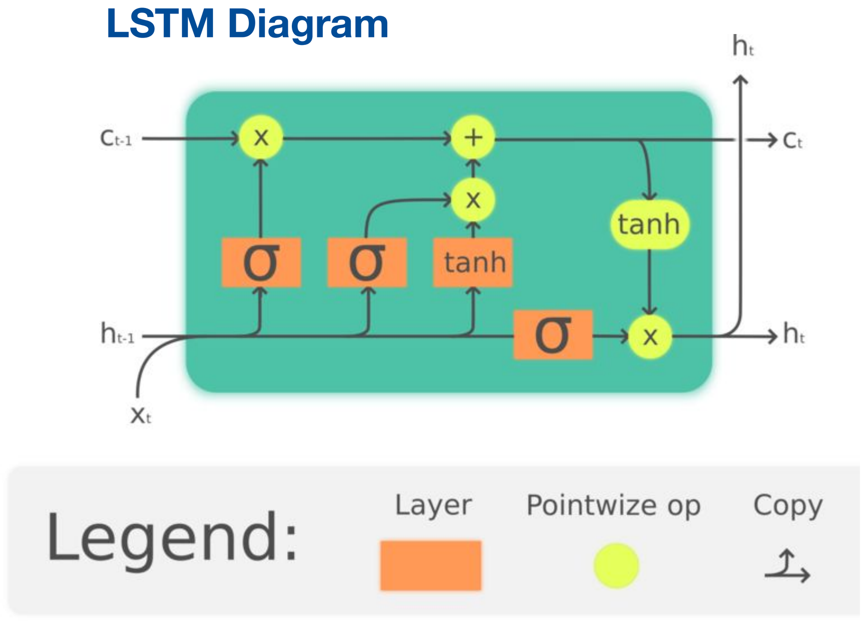
- Human-generated dataset of labeled pictures
- Images automatically cropped to highlight page index
- Trained in a Docker container on Google Cloud Platform
  - Nvidia Tesla V100 training GPU
- Git for data organization

Network Architecture

- Conv + MaxPool2d
- Conv + MaxPool2d
- Dropout
- Flatten
- LSTM ➡
- LSTM ⬅
- Dense, dim 128
- Dense, dim 4 ([A-Z], 3x[0-9, blank])

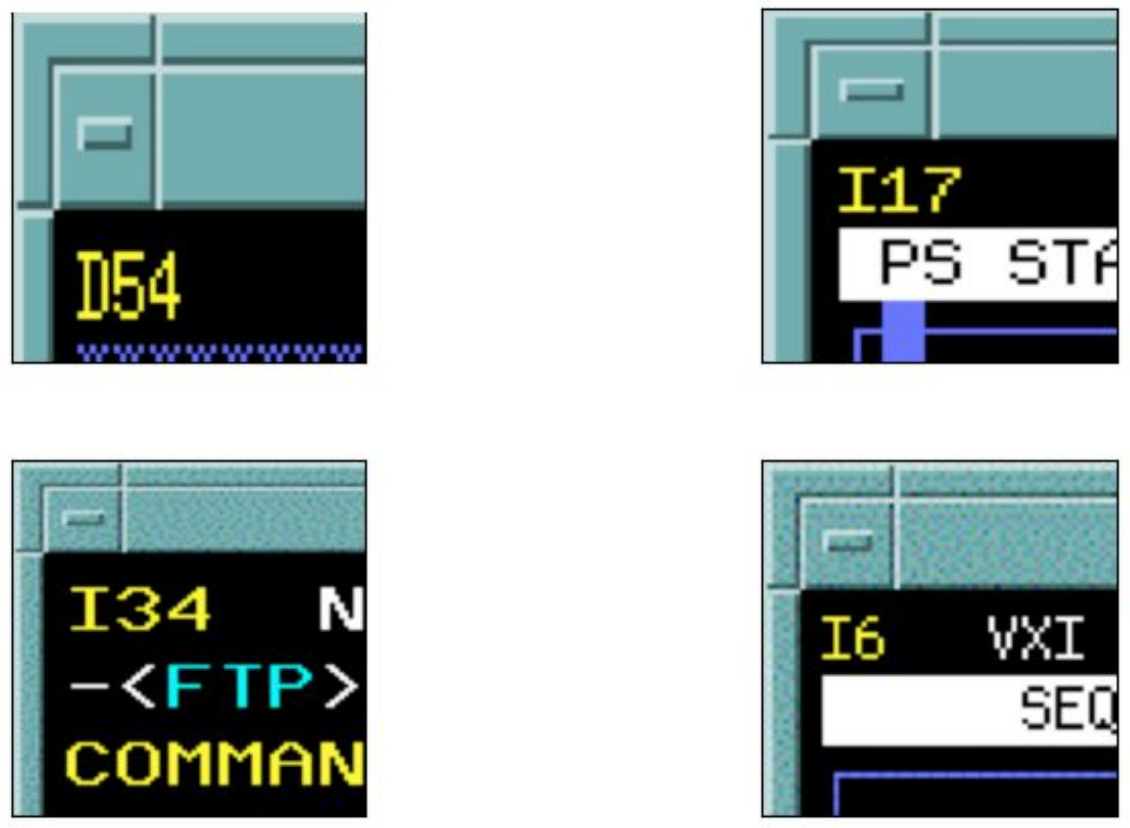
Loss Function: binary_crossentropy

LSTM Diagram



https://commons.wikimedia.org/wiki/File:The_LSTM_cell.png
https://creativecommons.org/licenses/by/4.0/legalcode

Cropped images of example parameter pages showing index

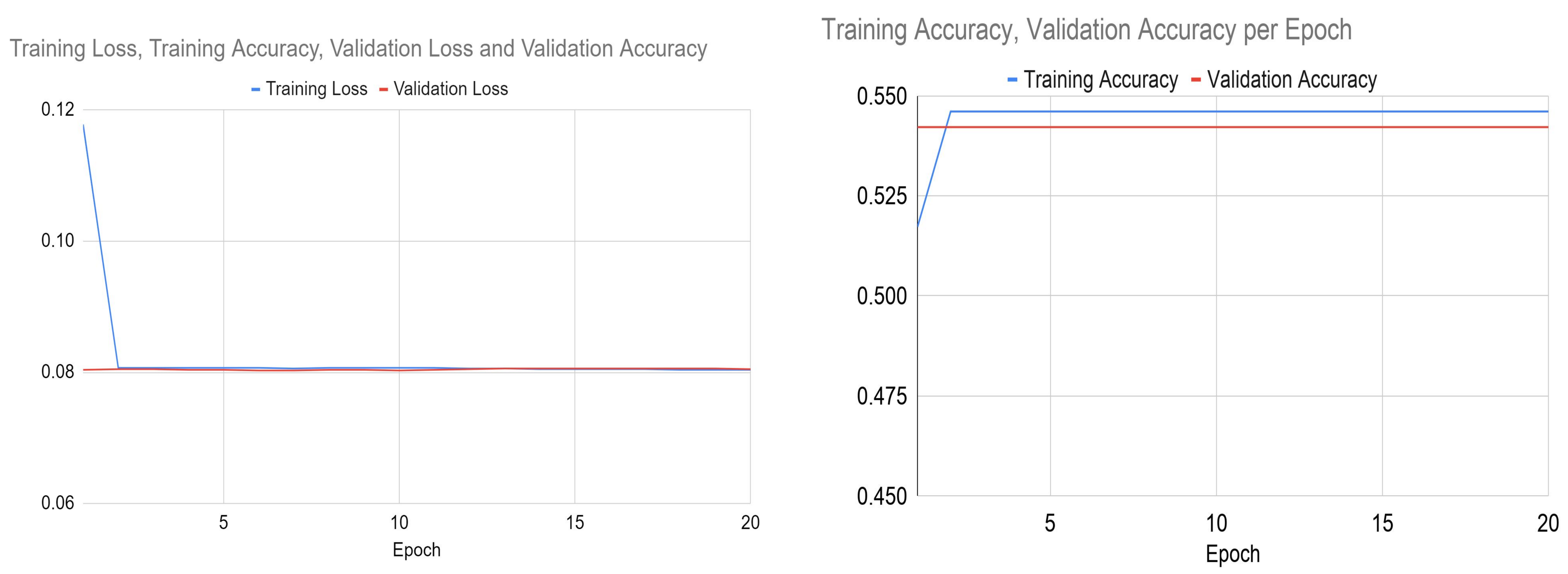


- Final Dataset: 3506 Labeled Images
- Total Alphabetical Characters Found: 15/26

Results

- Most recent training appears to guess the same label for every image
- Accuracy and loss level out quickly, but overfitting is avoided
  - Validation scores are comparable to training scores
- Restricted labels from 26 character to 15 characters found in the data improved the loss (0.1867 → 0.0804) and accuracy (0.2323 → 0.5461)

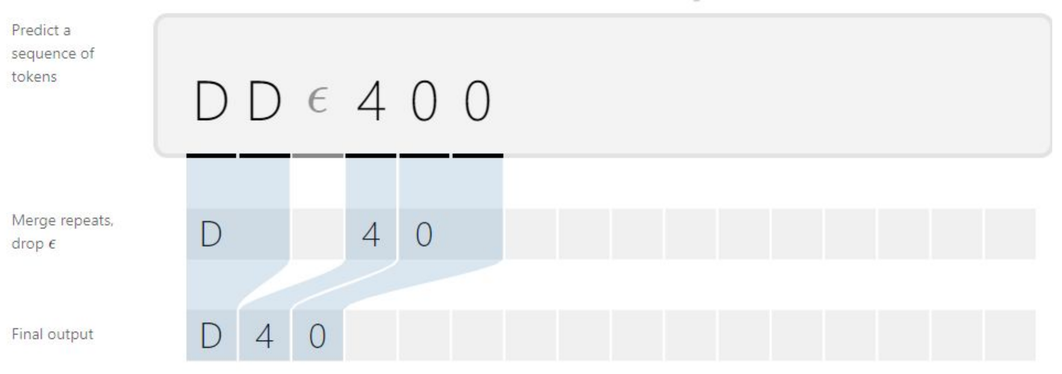
Loss and accuracy during training (Restricted character set)



Future Steps

- Implement connectionist temporal classification (CTC) loss to help guide network training
  - Requires implementing CTC Encoding/Decoding
  - Forgiving of repeated characters in output
- Multi-Headed Attention(Transformer)

CTC Decoding - Duplicated characters are combined, ε represents null character



https://distill.pub/2017/ctc/
https://creativecommons.org/licenses/by/4.0/legalcode

Acknowledgements

This work was supported in part by the U.S. Department of Energy, Office of Workforce Development for Teachers and Scientists. I would like to thank my mentors Jason St. John and Kyle Hazelwood, as well as my team members Giovani Leone and Terence Njekeu. Lastly I would like the thank Fermilab for hosting and the CCI program.