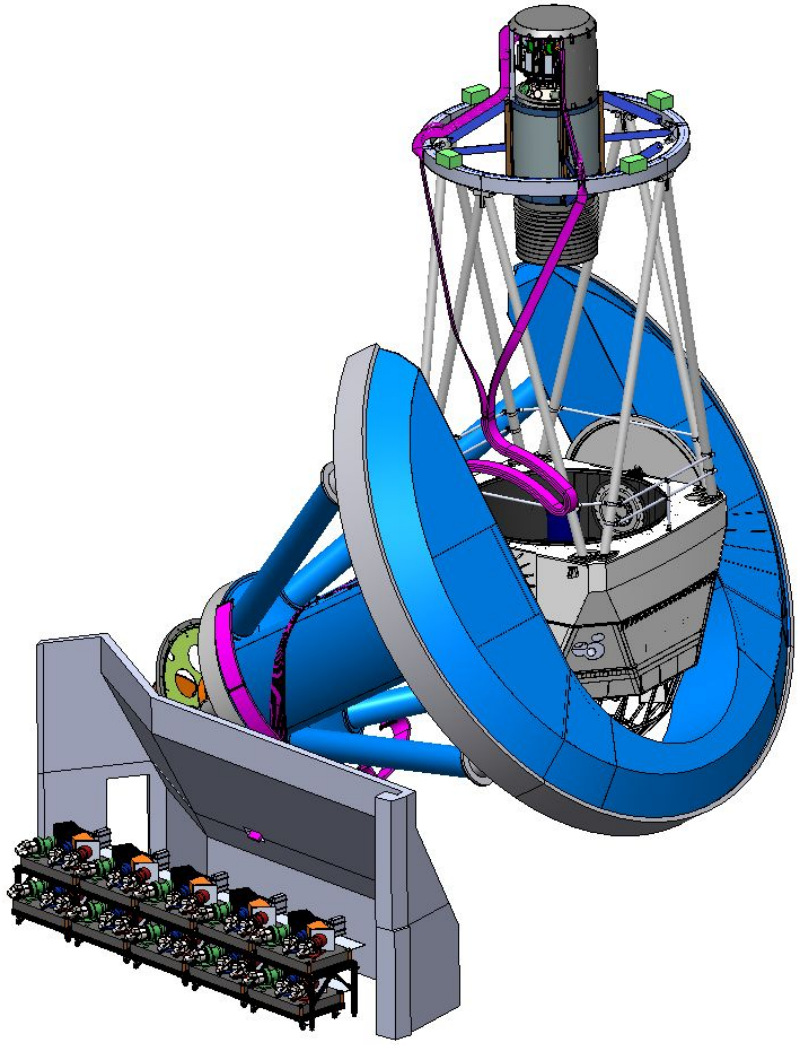


# The Future of QuasarNP (DESI Project 235)

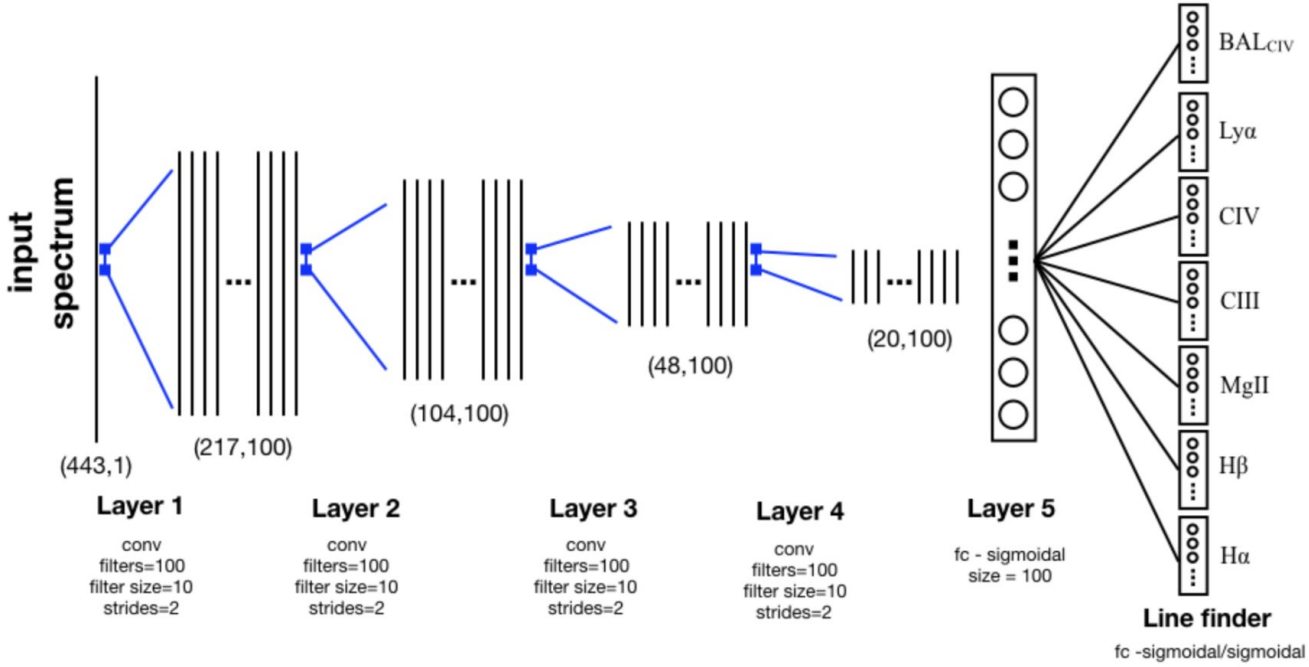
Dylan Green, UC Irvine

Collaboration Meeting

# QuasarNet

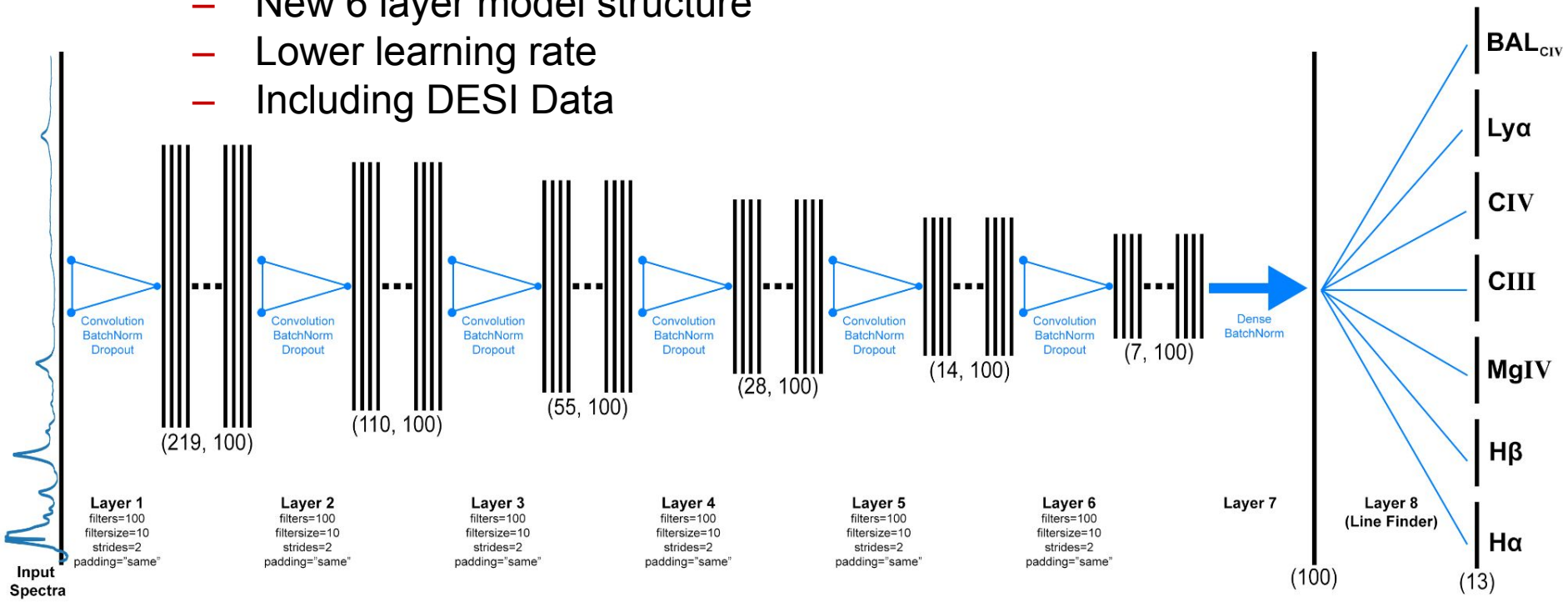
- QuasarNet is a neural network that identifies quasars directly from flux-calibrated spectra.
  - http://arxiv.org/abs/1808.09955, http://arxiv.org/abs/2007.10348

- QuasarNet is designed to identify 6 Emission lines and 1 Broad Absorption Line (BAL)

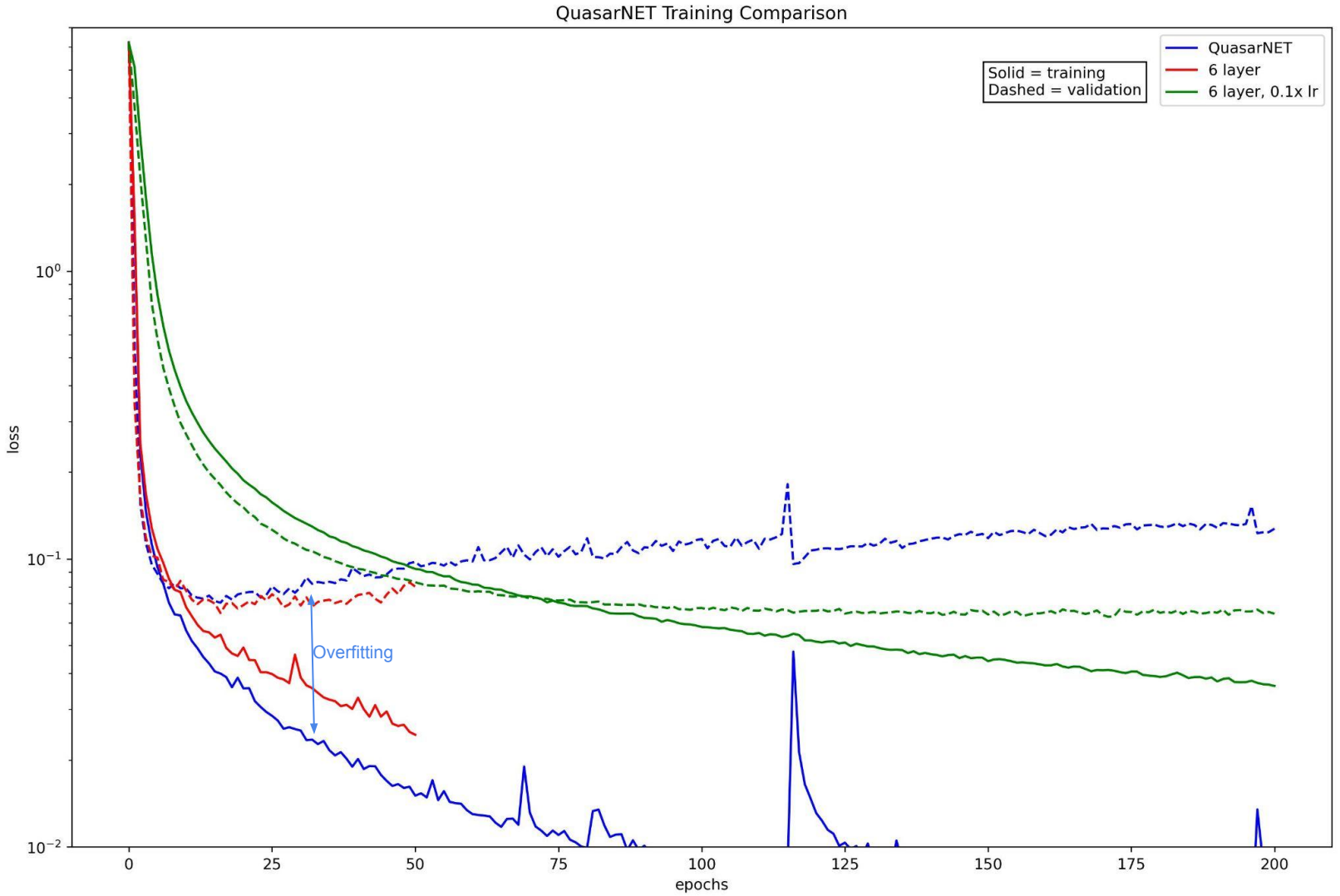


# QuasarNET/NP Today

- QuasarNP is a “pure” numpy implementation of QuasarNet that runs in a standard DESI environments (without tensorflow).
  - https://github.com/desihub/QuasarNP
- Current QuasarNET weights were trained by James Farr in 2020
- Validation QSO performance could be improved:
  - New 6 layer model structure
  - Lower learning rate
  - Including DESI Data



# QuasarNET Today

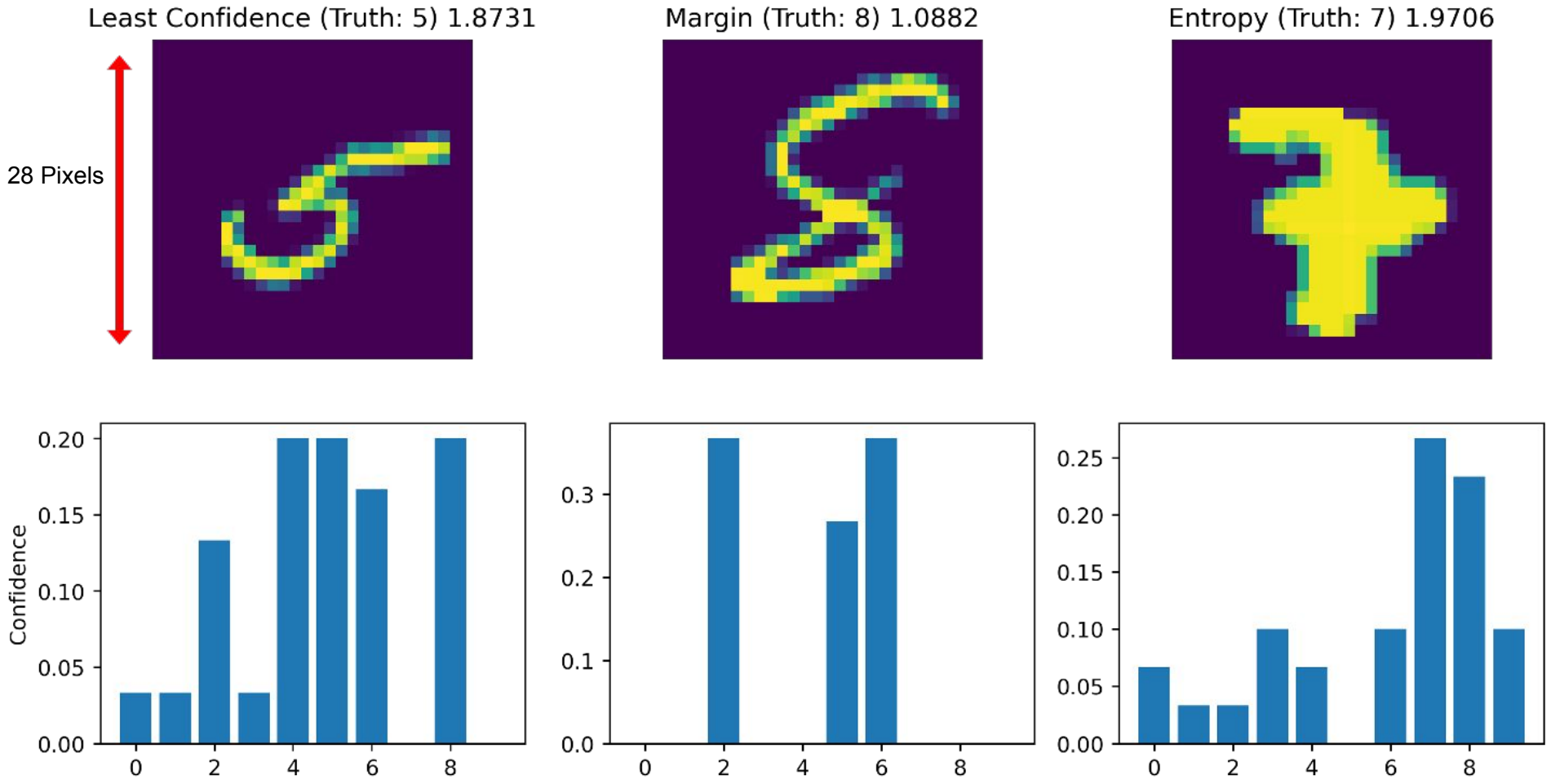


# What's Next?

- Active Learning: Algorithm to iteratively improve neural network (NN) performance:
  - (a) Use an ensemble of classifiers to quantify uncertainty
  - (b) Determine unlabeled data which is "confusing"
    - (i) Confirm that confusing data isn't an outlier in data-space
  - (c) Label this data with ground truth values
  - (d) Include newly labeled data in training and repeat.
- Practical Example: Modified National Institute of Standards and Technology (MNIST) Handwritten Digits

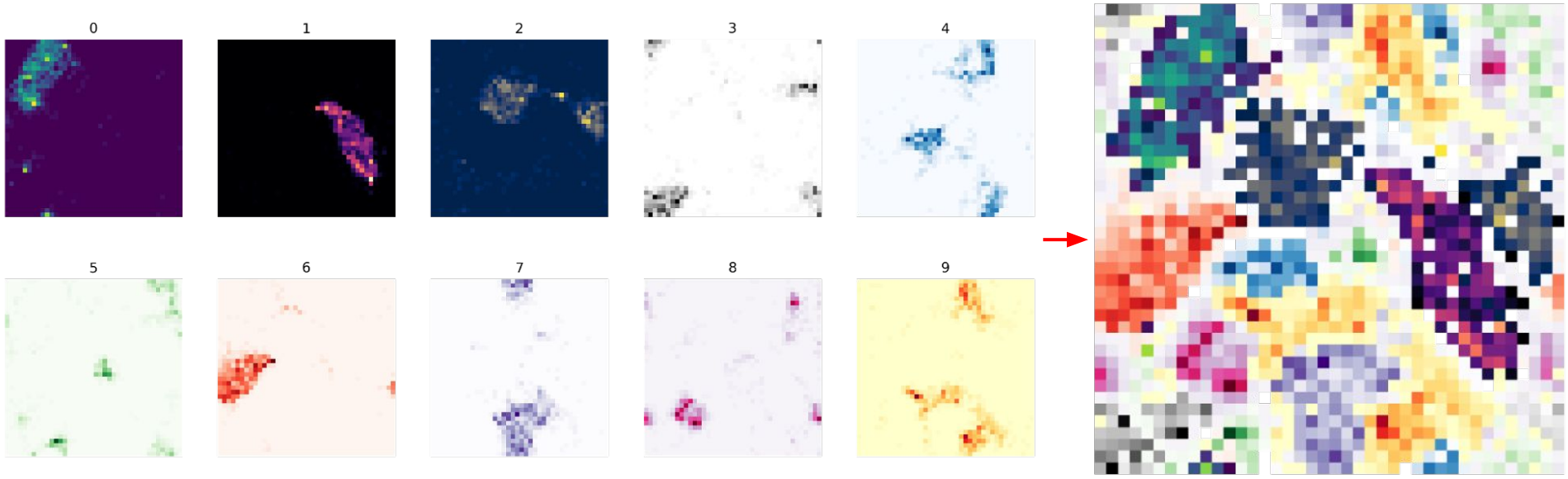
# Active Learning (Example)

- Ensemble of 30 Classifiers on MNIST:



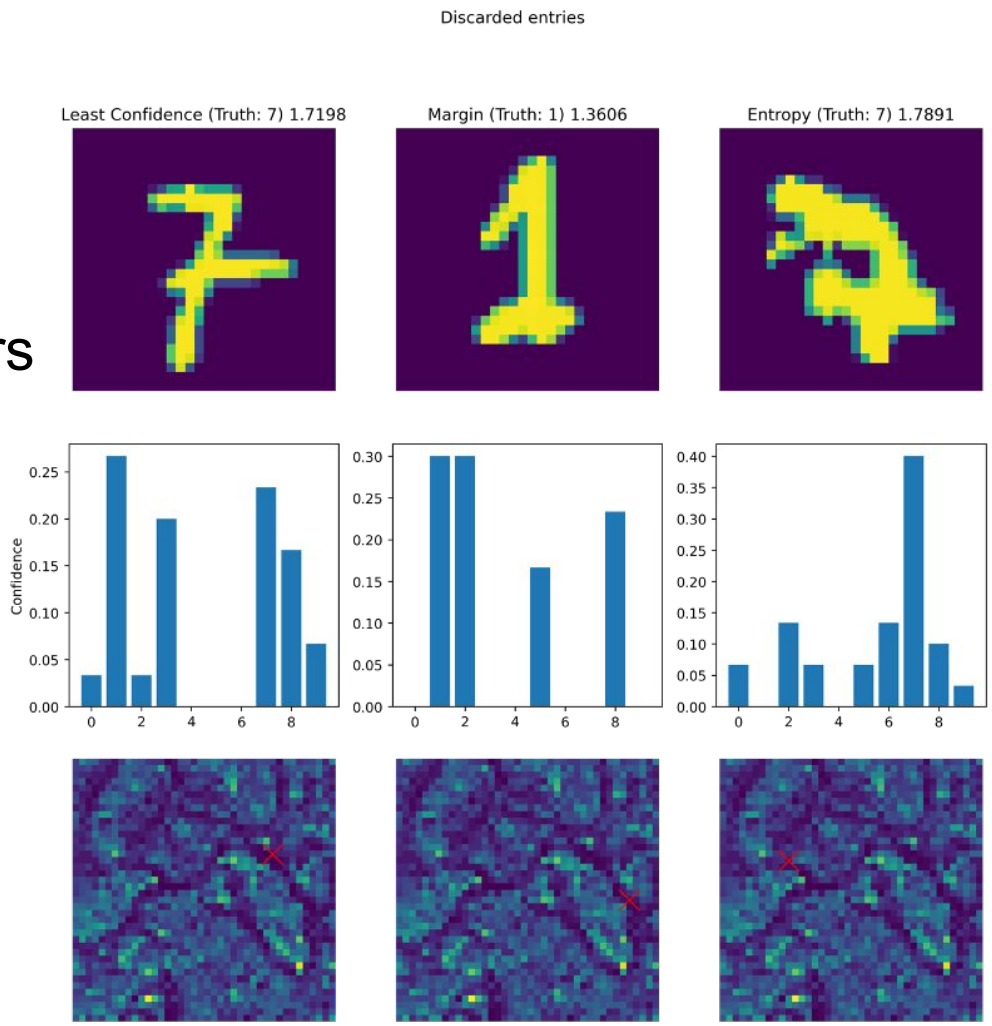
# Self Learning Maps (SOMs)

- Unsupervised machine learning method of dimensionality reduction
- Similar data is closer together



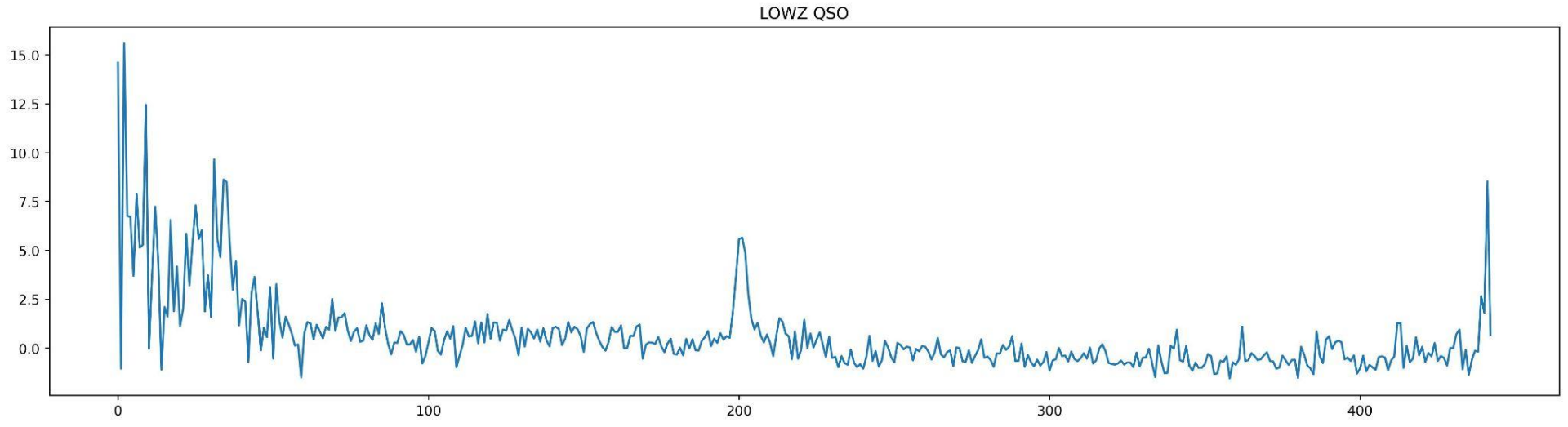
# How do we use SOMs?

- Use ensemble to determine confusing entries
- Use SOM to determine if confusing entries are outliers or representative
- If confusing entries are outliers, discard them as choices



# Active Learning for QuasarNET

- Algorithm to iteratively improve QNET performance:
  - (a) Use an ensemble of QNET classifiers to quantify uncertainty
  - (b) Determine (new) spectra which are “confusing”
    - (i) Use SOM to determine if confusing data is an outlier
  - (c) Visually Inspect and label this data with ground truth values
  - (d) Include newly labeled Spectra in training and repeat.



# Conclusion & Future Plans

- New version of QuasarNET is already potentially better than old
- DESI specific improvements can be gained by including DESI data in training
- Further improvements to be had with active learning
- Two phases of VI (200 spectra total, in “minibatches” of 50):
  - 100 spectra Early-Mid July
  - 100 spectra Mid July-Early August
  - Extra Credit: 100 additional spectra ~Mid August
- Currently looking for VI “experts” to help out!
- Questions?