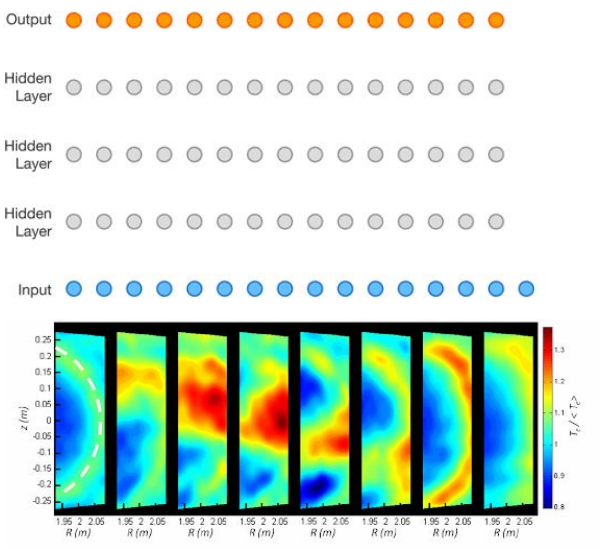
DIII-D disruption prediction using deep convolutional neural networks on raw imaging data



R.M. Churchill[1], the DIII-D team

7th Annual Theory and Simulation of Disruptions Workshop
Princeton, NJ Aug 5, 2019

Special thanks to:

- **DIII-D team generally, specifically Ben Tobias[1], Yilun Zhu[2], Neville Luhmann[2], Dave Schissel[3], Raffi Nazikian[1], Cristina Rea[4], Bob Granetz[4]**
- **PPPL colleagues: CS Chang[1], Bill Tang[1], Julian Kates-Harbeck[1,5], Ahmed Diallo[1], Ken Silber[1]**
- **Princeton University Research Computing[6]**

https://deepmind.com/blog/wavenet-generative-model-raw-audio/

1 PPPL
2 UC DAVIS UNIVERSITY OF CALIFORNIA
3 GENERAL ATOMICS
4 MIT PSFC
5 HARVARD UNIVERSITY
6 PRINCETON UNIVERSITY

Motivation: Automating classification of fusion plasma phenomena is complicated

- Fusion plasmas exhibit a range of physics over different time and spatial scales
- Fusion experimental diagnostics are disparate, and increasingly high time resolution
- How can we automate identification of important plasma phenomena, for example oncoming disruptions?

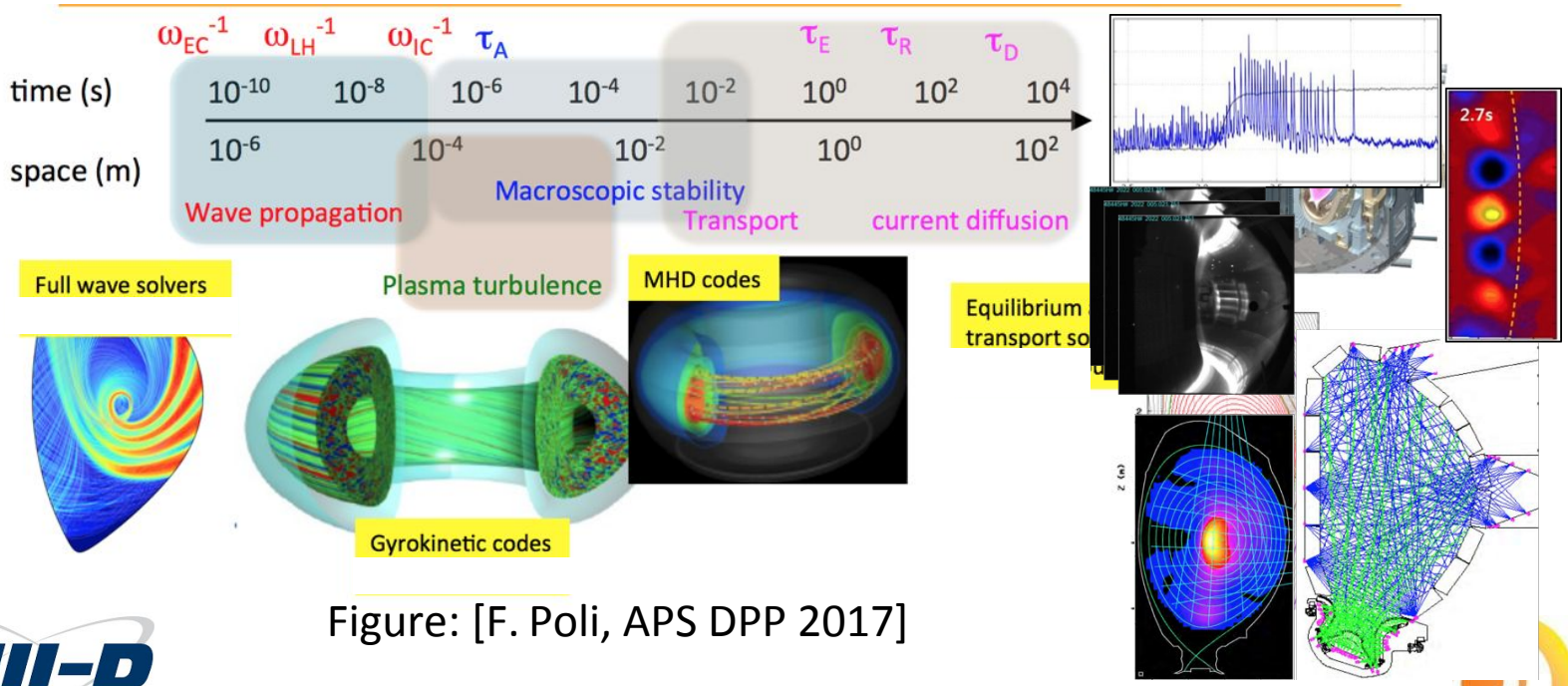


Figure: [F. Poli, APS DPP 2017]

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

PPPL

Outline

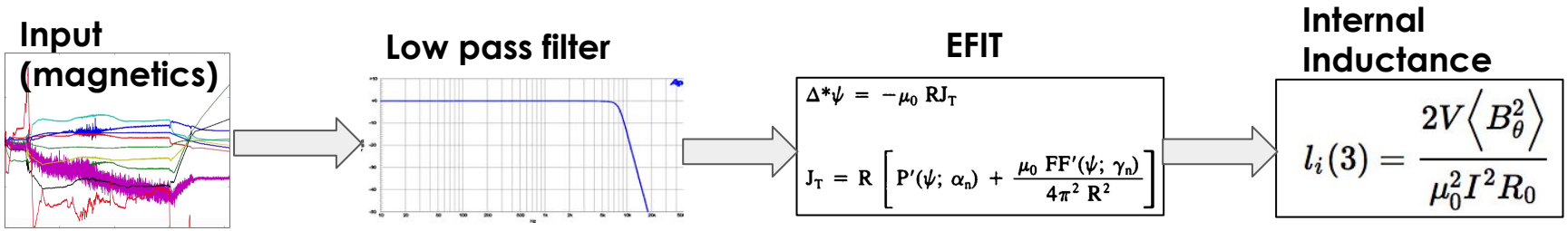
- Paradigm for deep learning
- Deep convolutional neural networks for long time series
- Initial results with ECEi for Disruption Prediction
- Future directions/Conclusions

Outline

- **Paradigm for deep learning**
- Deep convolutional neural networks for long time series
- Initial results with ECEi for Disruption Prediction
- Future directions/Conclusions

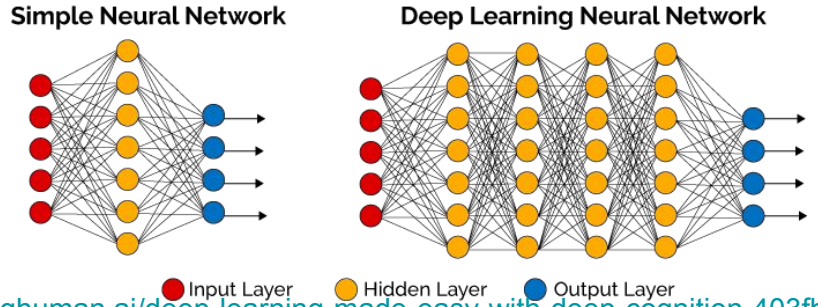
Neural networks can be thought of as series of filters whose weights are "learned" to accomplish a task

- Fusion experiment/simulation have a wide variety of data analysis pipelines, use prior knowledge to get result



$$\ell_i = f_n(\ldots(f_3(f_2(f_1(M))))\ldots)$$

- Neural networks (NN) have a number of layers of "weights" which can be viewed as filters (esp. Convolutional NN). But these filters are *taught* how to map given input through a complicated non-linear function to a given output.

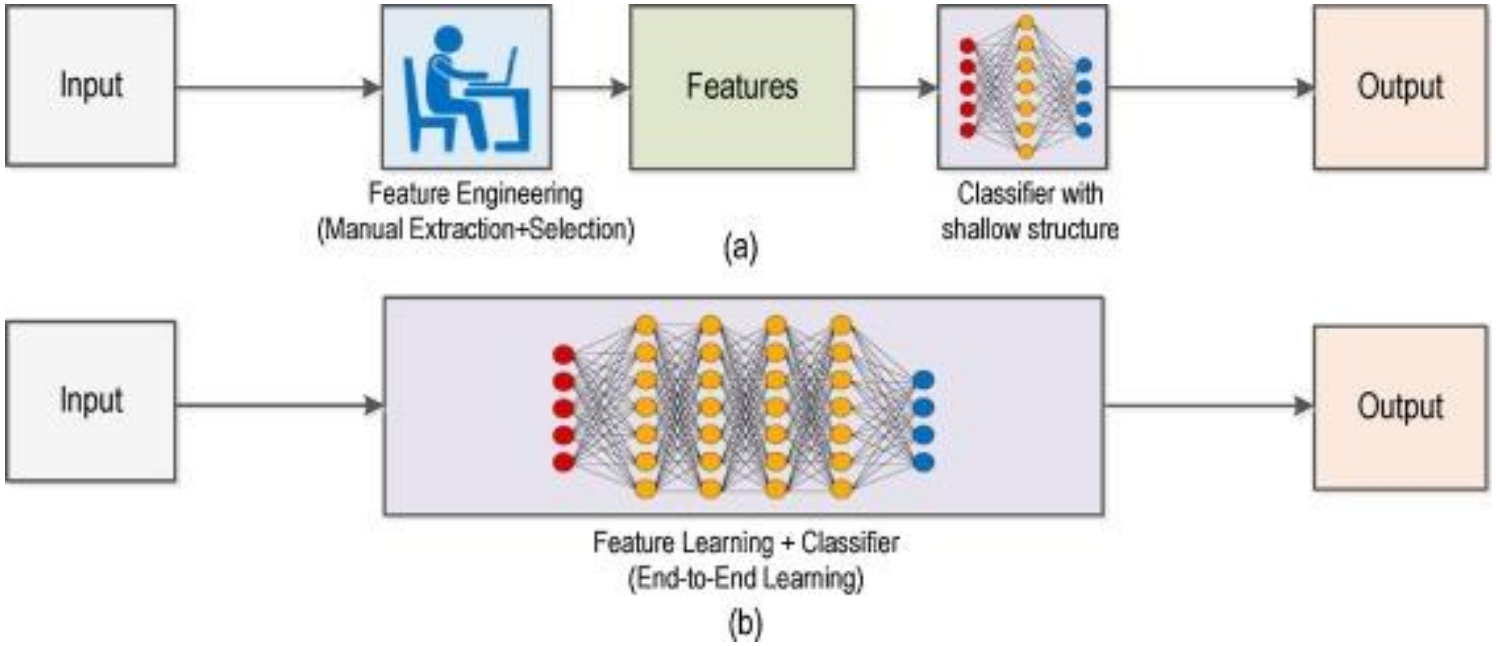


https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

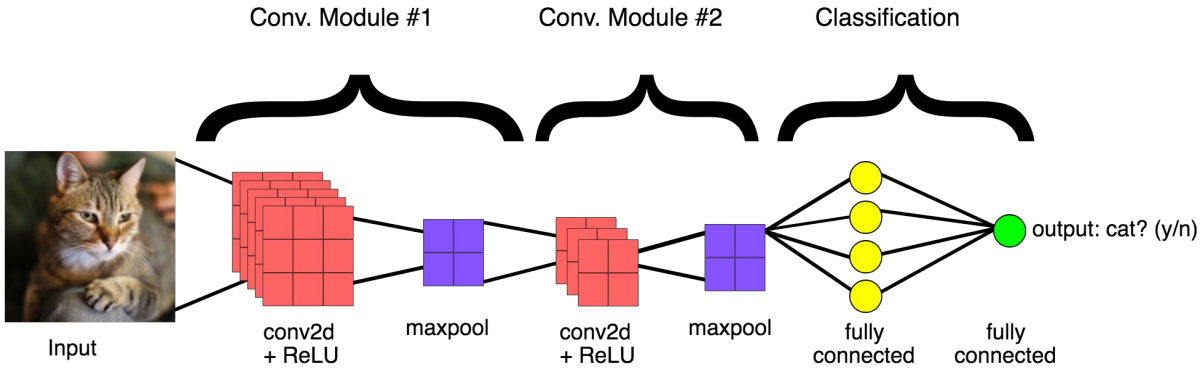
PPPL

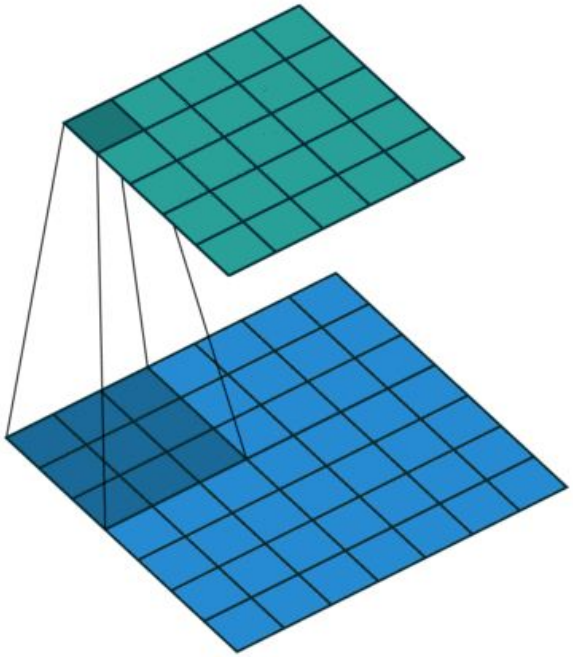
Deep learning enables end-to-end learning



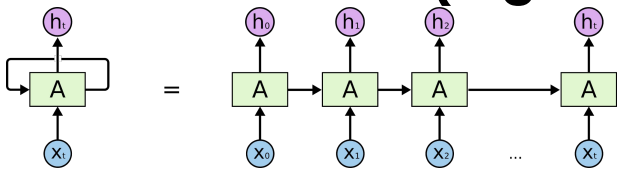
- Traditional machine learning focused on hand developed features (e.g. shape in an image) to train shallow NN or other ML algorithms
- Deep learning (multiple layer NN) enable end-to-end learning, where higher dimensional features (e.g. pixels in an image) are input directly to the NN

Convolutional neural networks (CNN)





- CNN's very successful in image classification, whereas Recurrent NN (e.g. LSTM) are often used for sequence classification (e.g. time series)



- But viewing NN as "filters", no reason CNN can't be applied to sequence machine learning also

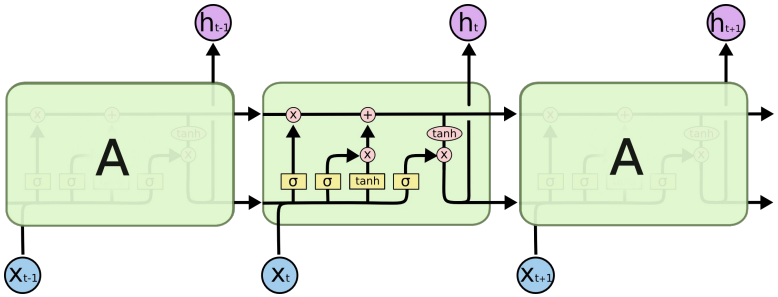
http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html
https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks

Outline

- Paradigm for deep learning
- **Deep convolutional neural networks for long time series**
- Initial results with ECEi for Disruption Prediction
- Future directions/Conclusions

Challenges for RNN/LSTM on long sequences

- Typical, popular sequence NN like LSTM *in principle* are sensitive to infinite sequence length, due to memory cell technique



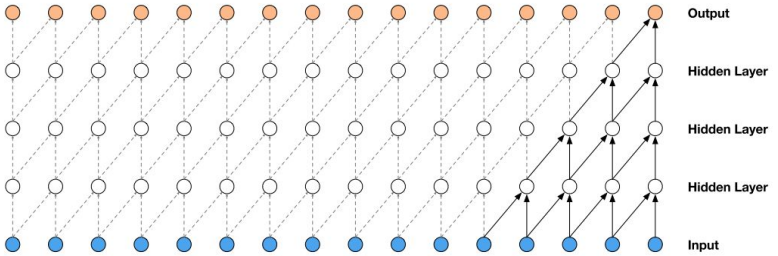
- However, in practice they tend to "forget" for phenomena with sequence length >1000 (approximate, depends on data)
- If characterising a sequence requires T_{long} seconds, and short-scale phenomena of time-scale T_{short} are important in the sequence, to use an LSTM requires $\frac{T_{long}}{T_{short}} \lesssim 1000$
- Various NN architectures enable learning on long sequences (CNN with dilated convolutions, attention, etc.)

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Dilated convolutions enable efficient training on long sequences

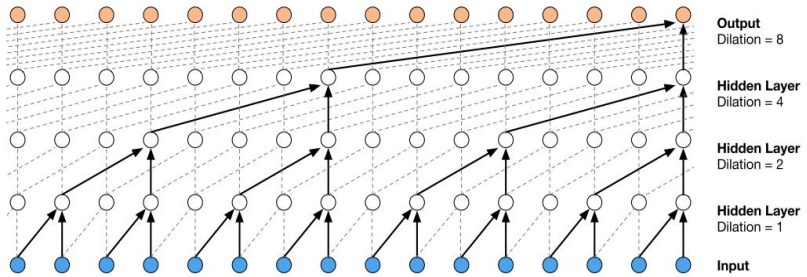
- **One difficulty using CNNs with causal filters is they require large filters or many layers to learn from long sequences**
  - Due to memory constraints, this becomes infeasible

Normal convolution



- **A seminal paper [*] showed using dilated convolutions (i.e. convolution w/ defined gaps) for time series modeling could increase the NN receptive field, reducing computational and memory requirements, and allowing training on long sequences**

Dilated convolution



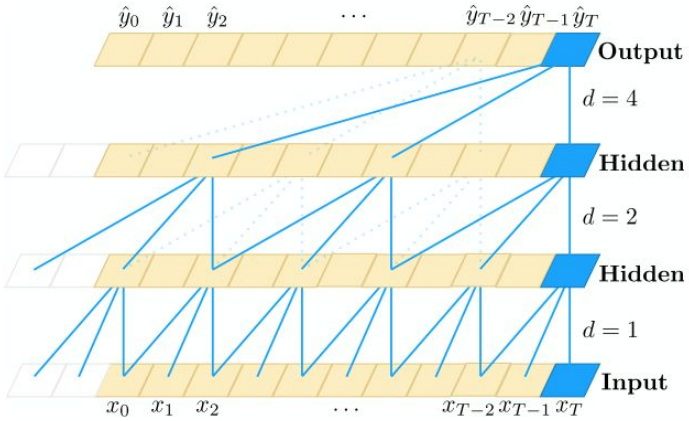
[* A. Van Den Oord, et. al., WaveNET: A Generative Model for Raw Audio, 2016]

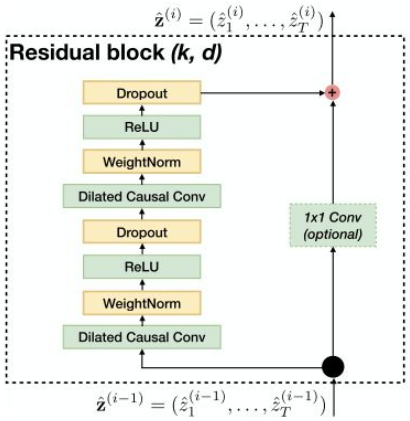
DIII-D NATIONAL FUSION FACILITY SAN DIEGO

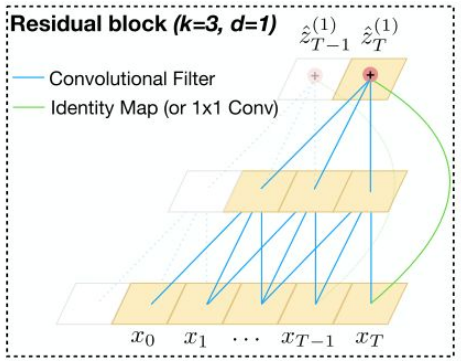
PPPL

Temporal Convolutional Networks

- **Temporal Convolutional Network (TCN) architecture [*] combines causal, dilated convolutions with additional modern NN improvements (residual connections, weight normalization)**
- **Several beneficial aspects compared to RNN's:**
  - Empirically TCN's exhibit longer memory (i.e. better for long sequences)
  - Non-sequential, allows parallelized training and inference
  - Require less GPU memory for training







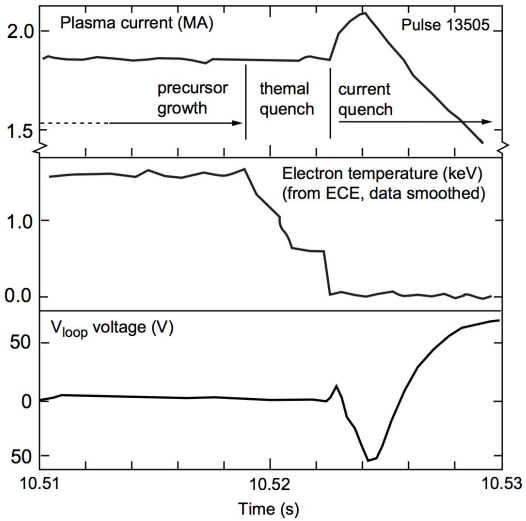
[* Bai, J.Z. Kolter, V. Koltun, http://arxiv.org/abs/1803.01271(2018)]

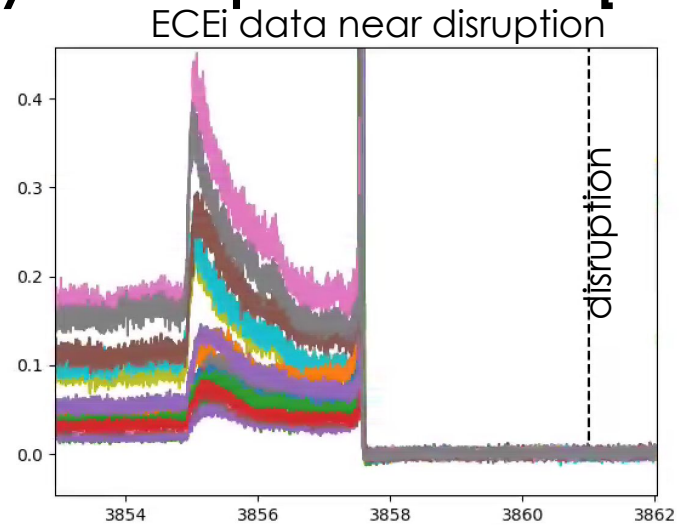
Outline

- Paradigm for deep learning
- Deep convolutional neural networks for long time series
- **Initial results with ECEi for Disruption Prediction**
- Future directions/Conclusions

Machine learning for Disruption Prediction

- **Predicting (and understanding?) disruptions is a key challenge for tokamak operation, a lot of ML research has been applied**
  [Vega *Fus. Eng.*, 2013 , Rea *FST* 2018, Kates-Harbeck *Nature* 2019, D. Ferreira arxiv 2018]
  - Most ML methods use processed 0-D signals (e.g. line averaged density, locked mode amplitude, internal inductance, etc.)
  - **Can we apply deep CNNs directly to diagnostic outputs for improved disruption prediction?**
- **Electron Cyclotron Emission imaging (ECEi) diagnostic has temporal & spatial sensitivity to disruption markers [Choi *NF* 2016]**

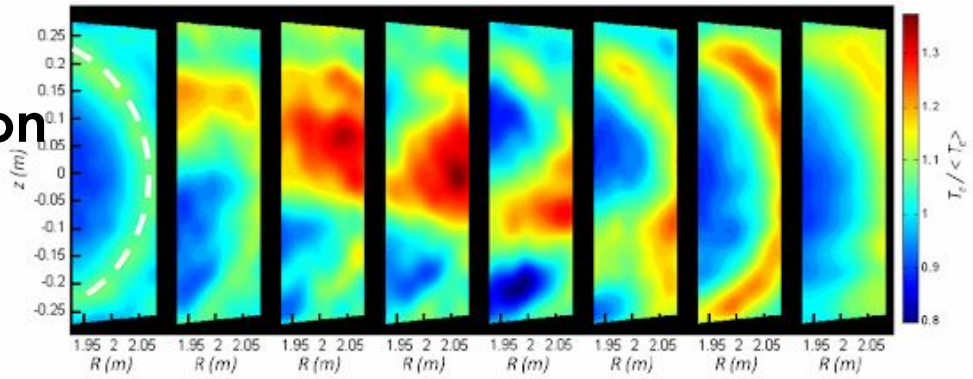




ITER Physics Basis, Chapter 3 Nucl. Fusion 39 (1999) 2251–2389.

DIII-D Electron Cyclotron Emission Imaging (ECEi)

- **ECEi characteristics:**
  - Measures electron temperature, Te
  - Time resolution (1 MHz) enabling measurement of δTe on turbulent timescales
  - Digitizer sufficient to measure entire DIII-D discharge (~O(5s))
  - 20 x 8 channels for spatial resolution
  - Some limitations due to signal cutoff above certain densities
- **Sensitive to a number of plasma phenomena, e.g.**
  - Sawteeth
  - Tearing modes
  - ELM's
- **Due to high temporal resolution (long time sequences), and spatial resolution, ECEi is a good candidate for applying end-to-end TCN**

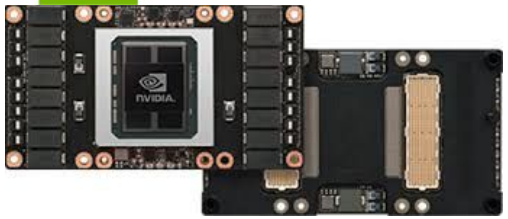


[B. Tobias et al., RSI (2010)]

https://sites.google.com/view/mmwave/research/advanced-mmw-imaging/ecei-on-diii-d

Dataset and computation

- **Database of ~3000 shots (~50/50 non-disruptive/disruptive) with good ECEi data created from the Omfit DISRUPTIONS module shot list [E. Kolemen, et. al.]**
  - "Good" data defined as all channels have SNR>3, avoid discharges where 2nd harmonic ECE cutoff
- **ECEi data (~10 TB) transferred to Princeton TigerGPU cluster for distributed training (320 nVidia P100 GPU's, 4 GPU's per compute node)**



globus

OMFIT

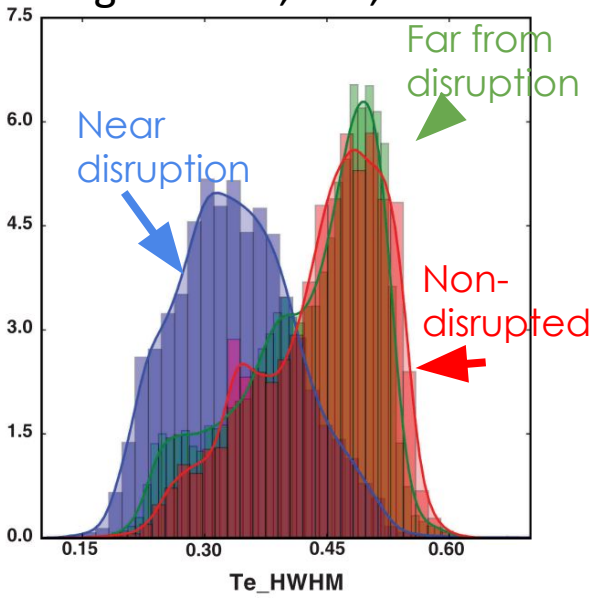
DIII-D NATIONAL FUSION FACILITY SAN DIEGO

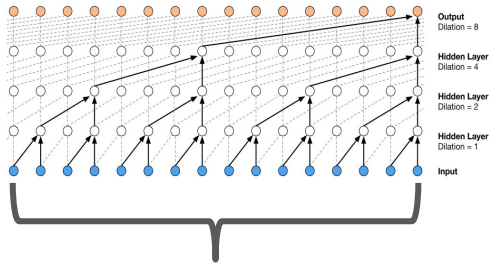
PPPL

Setup for training neural network

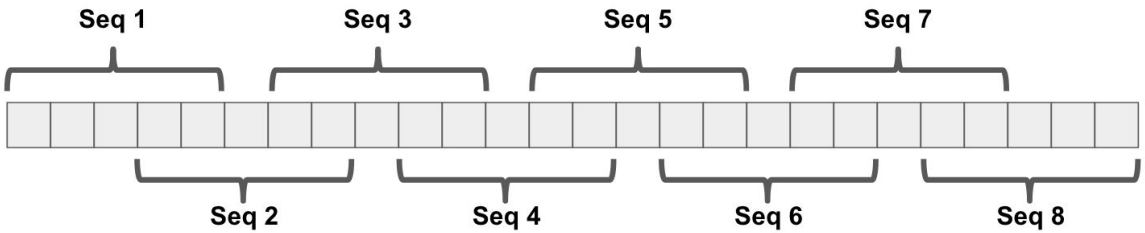
- **Each time point is labeled as "disruptive" or "non". For a disruptive shot, all time points 300ms or closer to disruption are labelled "disruptive"**
  - Times before 350ms have similar distribution to non-disruptive discharges [Rea *FST* 2018]
- **Binary classification problem (disruptive/non-disruptive time slice)**
- **Overlapping subsequences of length >> receptive field are created, length mainly set by GPU memory constraints**

Figure: Rea, *FST*, 2018





Receptive field
(# inputs needed to make 1 prediction)

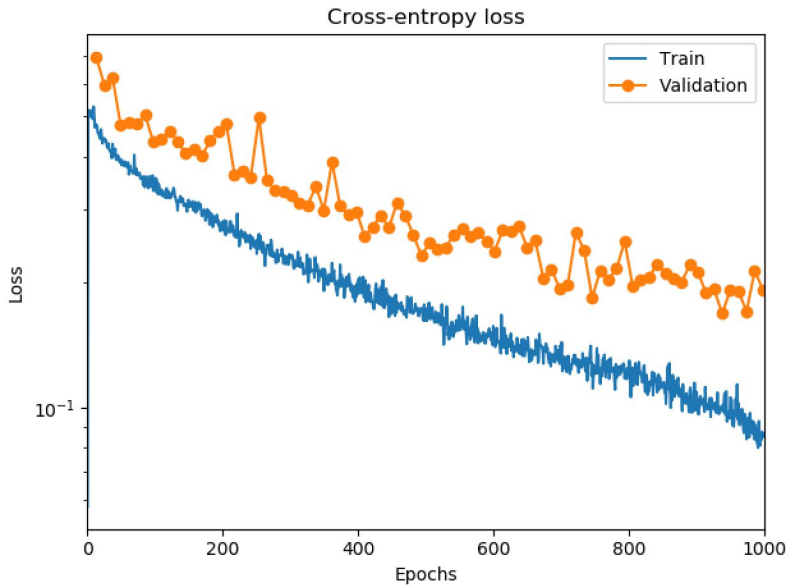


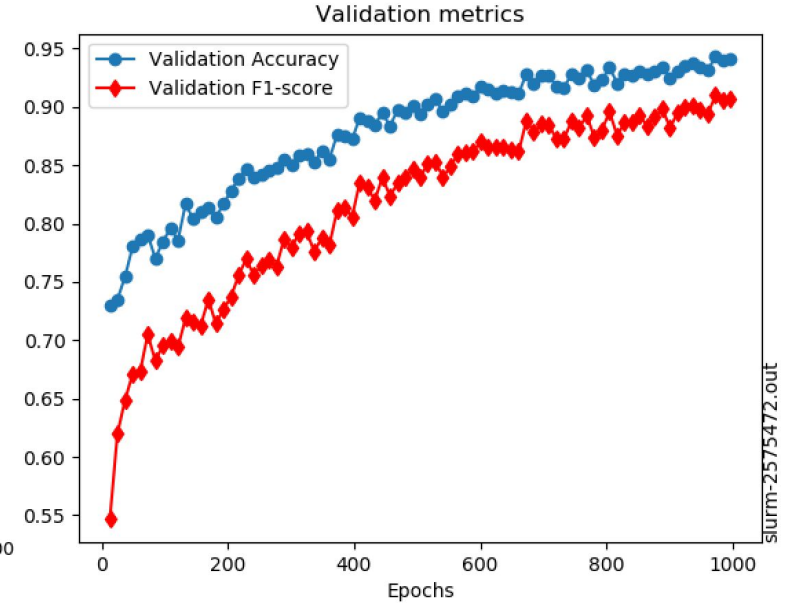
DIII-D NATIONAL FUSION FACILITY SAN DIEGO

PPPL

Current, initial results using ECEi only

- **Training on the subset of data, the loss does continually decrease, suggesting the network has the capacity necessary to capture and model disruptions with the ECEi data**
- **F1-score is ~91%, accuracy ~94%, on individual time slices.**
  - Additional regularization and/or training with larger dataset can help improve.
- **Run on 16 GPU's for 2 days.**





$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

$$= \frac{2}{\frac{TP+FP}{TP} + \frac{TP+FN}{TP}}$$

TP: True positives

TN: True negatives

FP: False positives

FN: False negatives

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

PPPL

Outline

- Paradigm for deep learning
- Deep convolutional neural networks for long time series
- Initial results with ECEi for Disruption Prediction
- **Future directions/Conclusions**

Future Possibilities

- **Deep CNN architectures (e.g. TCN) can be applied to many fusion sequence diagnostics, e.g. magnetics, bolometry, synchrotron radiation, etc.**
  - Tying together multiple diagnostics in a single or multiple neural networks can give enhanced possibilities, and sensitivity to the various types of disruptions
  - Can be used to create "automated logbook", identify various phenomena. Especially important for longer pulses + more diagnostics + higher time resolution diagnostics

- **Transfer learning can be explored for quickly re-training CNN on a different machine with few examples (use simulations for more examples, corrected by experiment? See Humbird 2018 NIF work, and Bill Tang/Ge Dong ongoing work for MFE)**

Future Possibilities (cont.)

- **Interpretability techniques offer the possibility of "opening the black box", identifying *why* the neural network makes a disruption prediction**

- **Physics informed learning: incorporating theoretical or simulation physics based conservation constraints [Raissi arxiv 2017]**
  - Adding in constraints to the loss function to punish unphysical solutions, or incorporating prior knowledge into the neural network structure

Interpretability of neural networks

- Techniques exist to determine which parts of input were most important for a prediction (e.g. which pixels important for self-driving car algorithm, Bojarski arxiv 2017)

- This offers the promise of identifying the root cause of the predicted disruption, and may be able to give empirical feedback to theory/simulation for scenarios to further explore and explain

Conclusions

- **Deep convolutional neural networks offer the promise of identifying multi-scale plasma phenomena, using end-to-end learning to work with diagnostic output directly**
  - TCN architecture with causal, dilated convolutions allows predictions to be sensitive to longer time sequences while maintaining computational efficiency

- **Initial work training a TCN on reduced ECEi datasets yields promising results for the ability of the TCN architecture to train a disruption predictor based on the ECEi data alone.**

- **Future work will compare the benefit of using full time resolution, including other diagnostics such as magnetics, tackling the issues related to transferring to new machines, and exploring interpretable algorithms to identify root cause of disruption predictions**

References

1. https://deepmind.com/blog/wavenet-generative-model-raw-audio/
2. F. Poli, APS DPP (2017)
3. https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351
4. http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html
5. https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks
6. A. Van Den Oord, et. al., https://arxiv.org/pdf/1609.03499.pdf (2016),
7. S. Bai, et. al., http://arxiv.org/abs/1803.01271 (2018).
8. ITER Physics Basis, Chapter 3, Nucl. Fusion 39 (1999) 2251–2389.
9. J. Vega, et. al., Fusion Eng. Des. 88 (2013) 1228–1231.
10. C. Rea, et. al., Fusion Sci. Technol. (2018) 1–12.
11. Kates-Harbeck, et. al., submitted (2018)
12. D. Ferreira, http://arxiv.org/abs/1811.00333, (2018)
13. M.J. Choi, et. al., Nucl. Fusion 56 (2016) 066013.
14. https://sites.google.com/view/mmwave/research/advanced-mmw-imaging/ecei-on-diii-d
15. B. Tobias, et. al., Rev. Sci. Instrum. 81 (2010) 10D928.

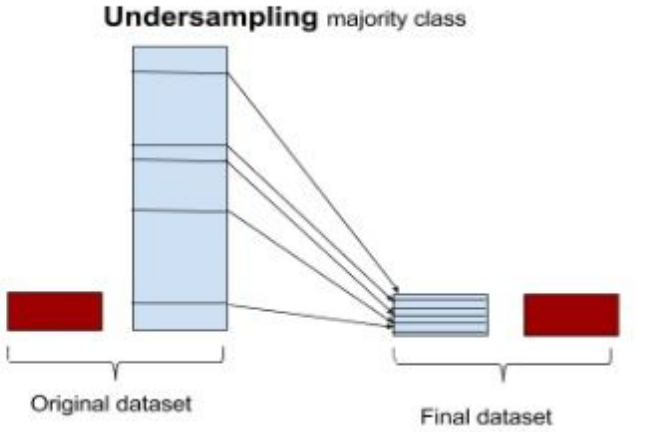
DIII-D NATIONAL FUSION FACILITY SAN DIEGO

PPPL

END PRESENTATION

Setup for training neural network

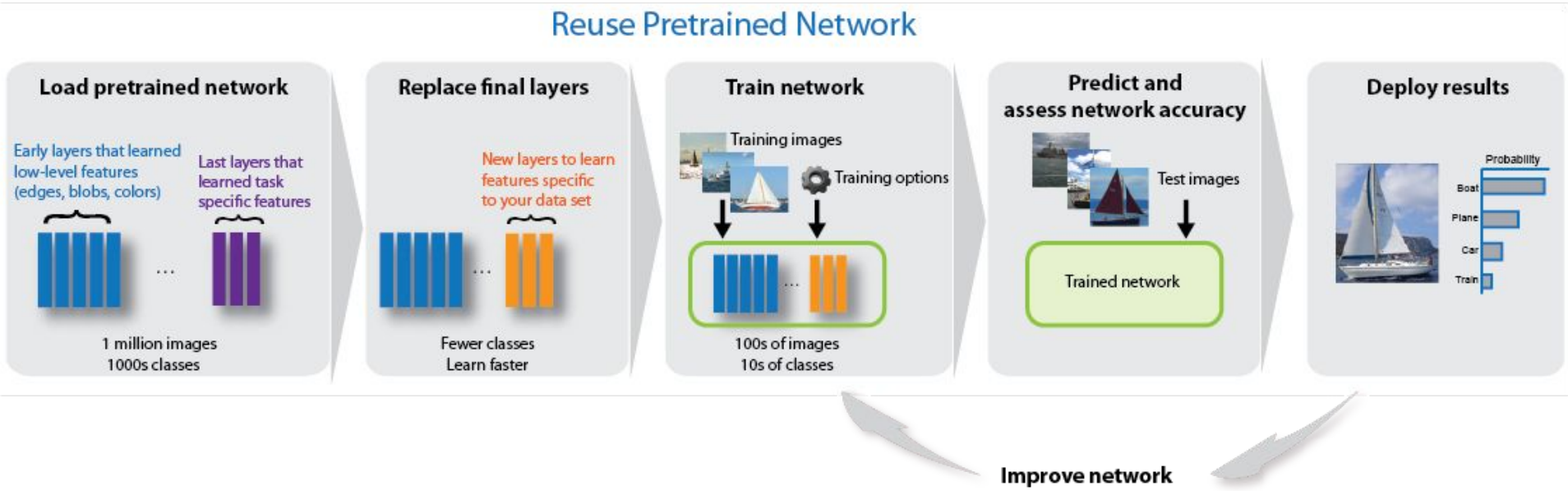
- **Starting with smaller subsets of data, working up.**
- **Data setup**
  - Downsampled to 100 kHz
  - Sequences broken up into subsequences of 78,125 (781ms)
  - Undersampled subsequence training dataset so that 50/50 split in non-disruptive/disruptive subsequences (natural class imbalance ~5% disruptive subsequences).
  - Weighted loss function for 50/50 balancing of time slices classes
  - Full 20 x 8 channels used (but no 2D convolutions)
  - Data normalized with z-normalization (y - mean(y))/std(y)
- **TCN setup:**
  - Receptive field ~30,000 i.e. 300ms (each time slice prediction based on receptive field)
  - 4 layers, dilation 10, kernel size 15, hidden nodes 80 per layer



DIII-D NATIONAL FUSION FACILITY SAN DIEGO

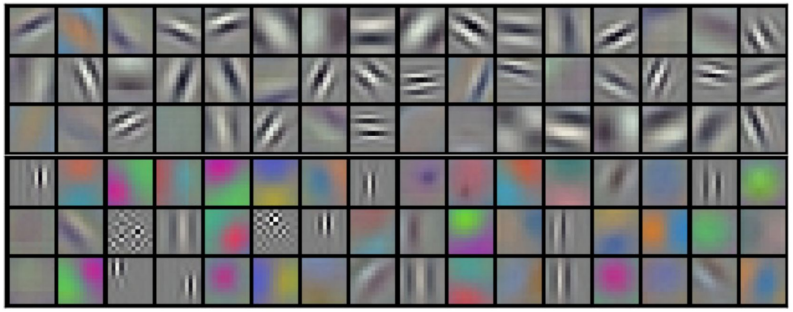
PPPL

Transfer learning



https://www.mathworks.com/help/deeplearning/examples/transfer-learning-using-alexnet.html;jsessionid=b71bf04f90335a09f702b9feb5e7

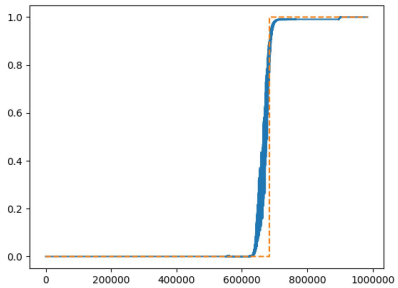
1st layers of CNN often exhibit basic filter characersitics, e.g. edge or color filters



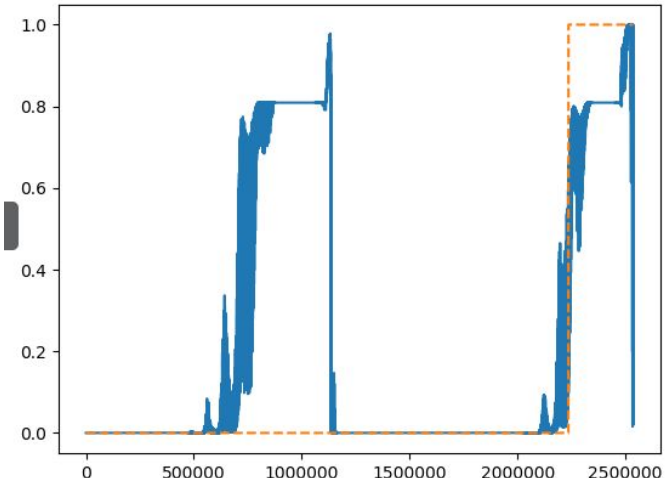
Alex Krizhevsky, et. al., NIPS 2012

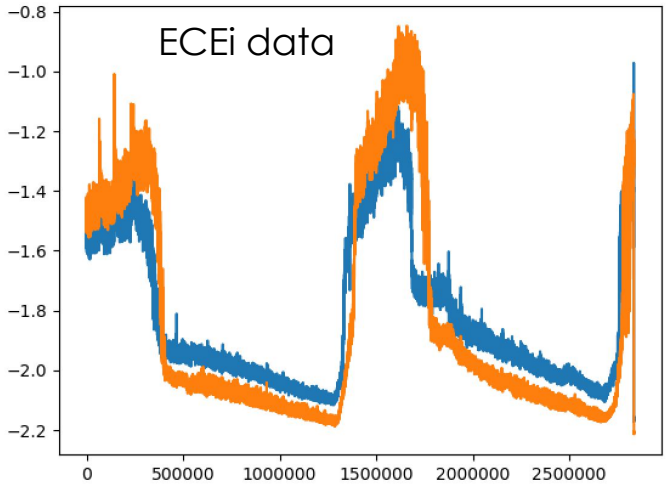
- **Shot where disruption alarm triggered >> 300ms before disruption due to very similar behavior in that time region to just before the disruption (drop in Te, followed by recovery)**

Target and prediction for disruptive shot



Target/prediction



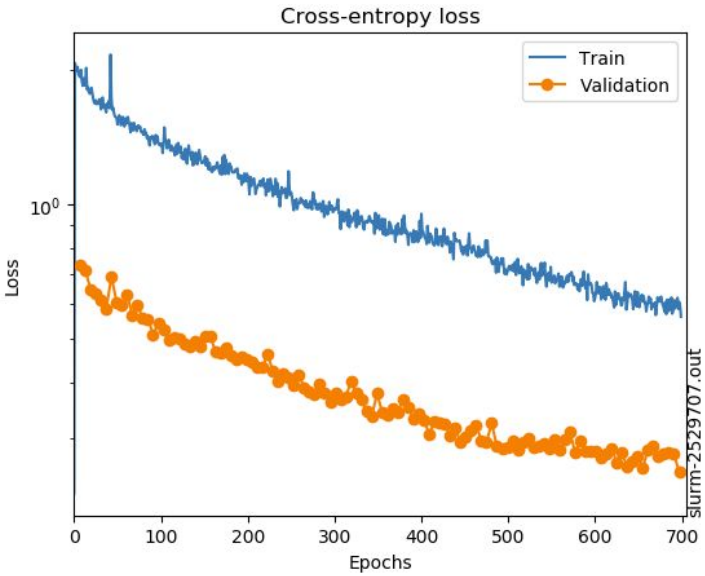


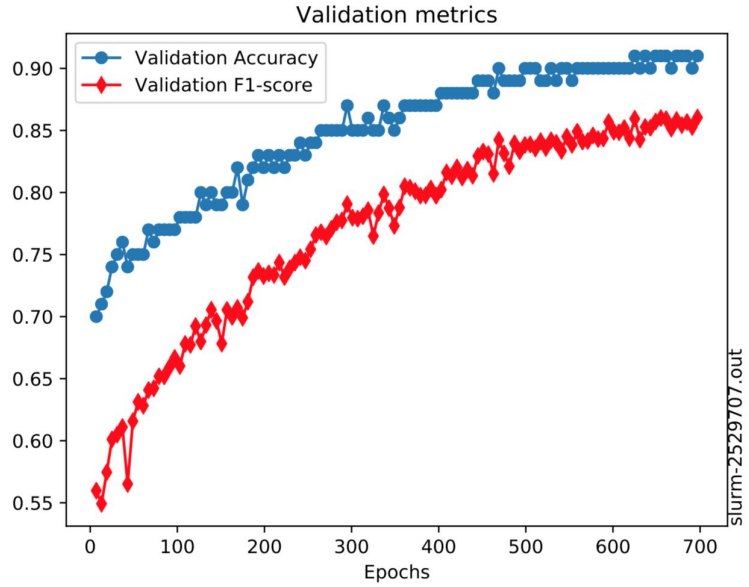
DIII-D NATIONAL FUSION FACILITY SAN DIEGO

PPPL

(previous results)

- **F1-score is ~86%, accuracy ~91%, on individual time slices.**
  - Additional regularization and/or training with larger dataset can help improve.
- **Run on 16 GPU's for 2 days.**





$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2}{\frac{TP+FP}{TP} + \frac{TP+FN}{TP}}$$

TP: True positives

TN: True negatives

FP: False positives

FN: False negatives

DIII-D NATIONAL FUSION FACILITY SAN DIEGO

PPPL