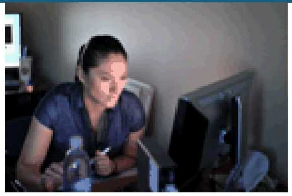
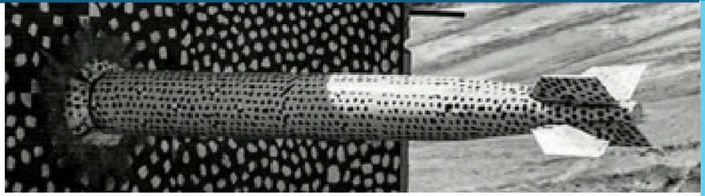
SAND2020-5943C

Developing Signatures-Based Safeguards for Enrichment Facilities





Presented By:

Nathan Shoman, Philip Honnold, Benjamin B. Cipiti, and James Garner

SAND-

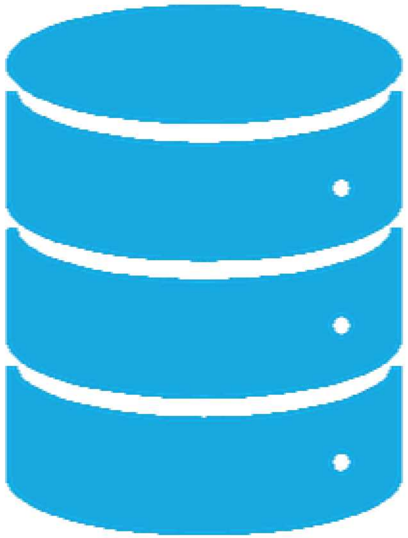
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

- Motivation
- Process Model
- Machine learning background
- Machine learning framework
- Current results
- Conclusions and future work

Current IAEA safeguards perform well, but could be improved through effectively leveraging existing data

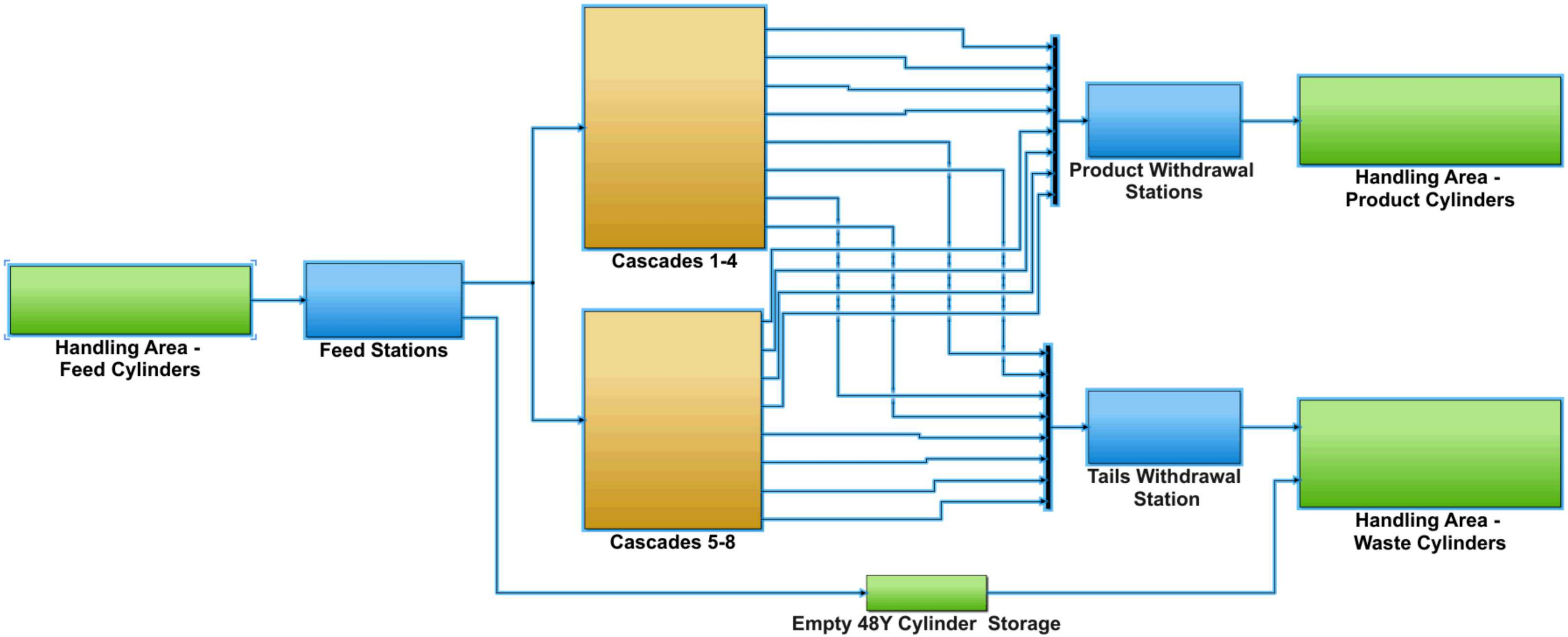
- Often enrichment plant operators make an accountancy declaration for an MBA and IAEA verifies the declaration
- Verification measures include auditing and independent measurements of the declared MBA
- Measurement technologies used for verification are frequently used to target a specific safeguards task
  - OLEM and environmental swipes to detect changes in enrichment
  - Accountancy scales to detect excess production
- Could leveraging multiple data streams improve the probability of detection for all safeguards tasks?
- Example use case: The use of temperature and pressure utilized by OLEM may also improve the probability of detection for excess production
- Machine learning could provide better probability of detection using existing safeguards data streams through improved data utilization
  - Success in other domains such cybersecurity



Development of an enrichment model with thermophysical feedback is challenging

- Tracking of discrete and continuous entities
- Thermophysical feedbacks
- Modeling realistic measurement systems
- Accurate representations of measurement errors
- Capture of normal facility variations

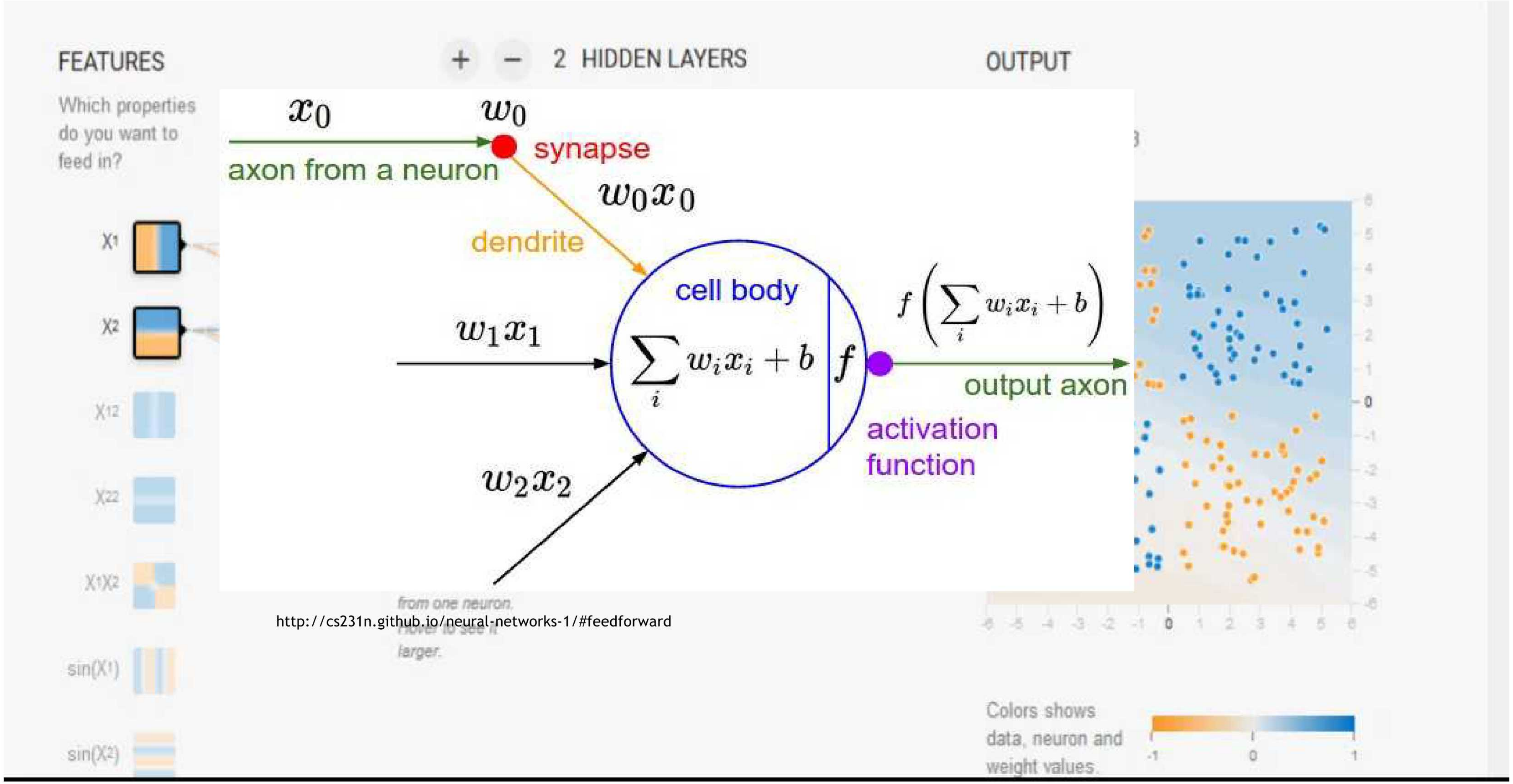
GCEP model is modeled in MATLAB Simulink which allows for easier conversion between discrete and continuous entities



Detecting anomalies in multivariate time-series data is difficult!

- Few to no anomalies available for training, which implies unsupervised training
- Traditional methods for anomaly detection may not perform well due to temporal dependencies
- False positives increase with measurement error and noise

What are neural approaches, really?



http://cs231n.github.io/neural-networks-1/#feedforward

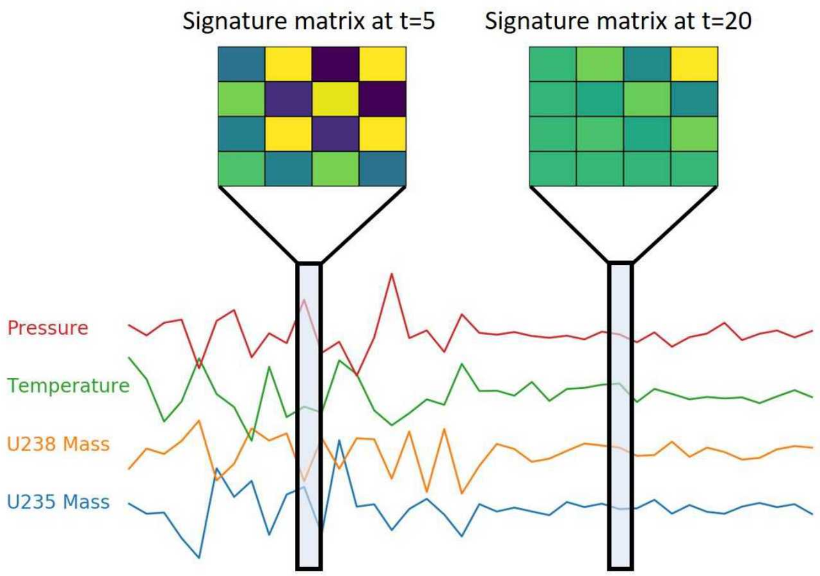
https://playground.tensorflow.org/

Developing the data format – signature matrix

- Correlations between different pairs of features can be used to characterize system status [Hallac 2017, Song 2018]

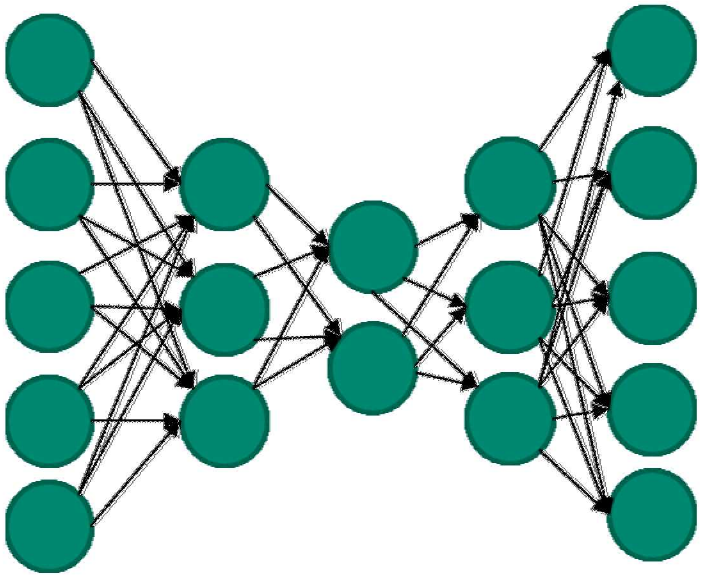
- $m_{ij}^{t} = \frac{\sum_{\delta=0}^{w} x_i^{t-\delta} x_j^{t-\delta}}{\kappa}$

- Multiple signature matrices can be formed to detect anomalies of different lengths
- Shorter anomalies can be detected with shorter windows, but will perform poorly on longer, more subtle anomalies
- Anomaly detection can now be formulated as pattern detection



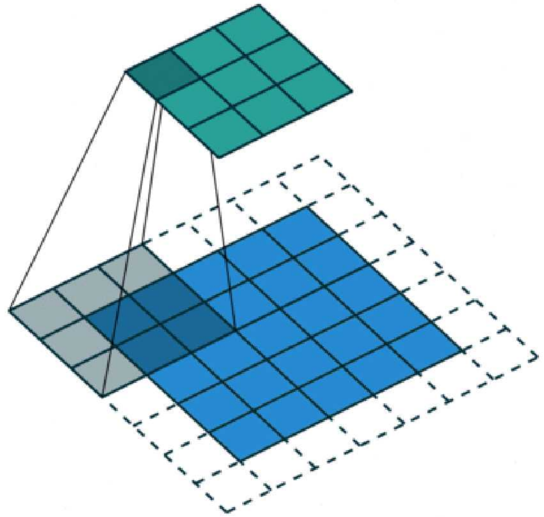
Determining the network structure - autoencoder

- Autoencoder is an unsupervised machine learning approach
  - Similar conceptually to principal component analysis
- Attempts to use a compressed representation of the input to reconstruct an output
- Compression forces the network to learn important features and underlying data structure
- Autoencoder should have a hard time recognizing patterns that are anomalous

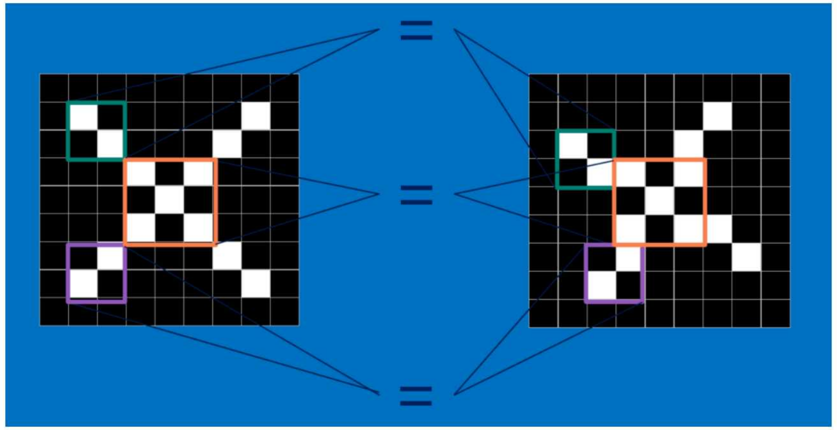


Choosing the layers – convolution

- Convolutional neural networks (CNNs) excel at a variety of image tasks
- Operate on images in small patches
- Attempts to find patterns within training data set
- More robust than literal comparison
  - Translations / rotations would fail in literal case
- Much fewer parameters than fully connected network
  - Images can be large even when scaled down



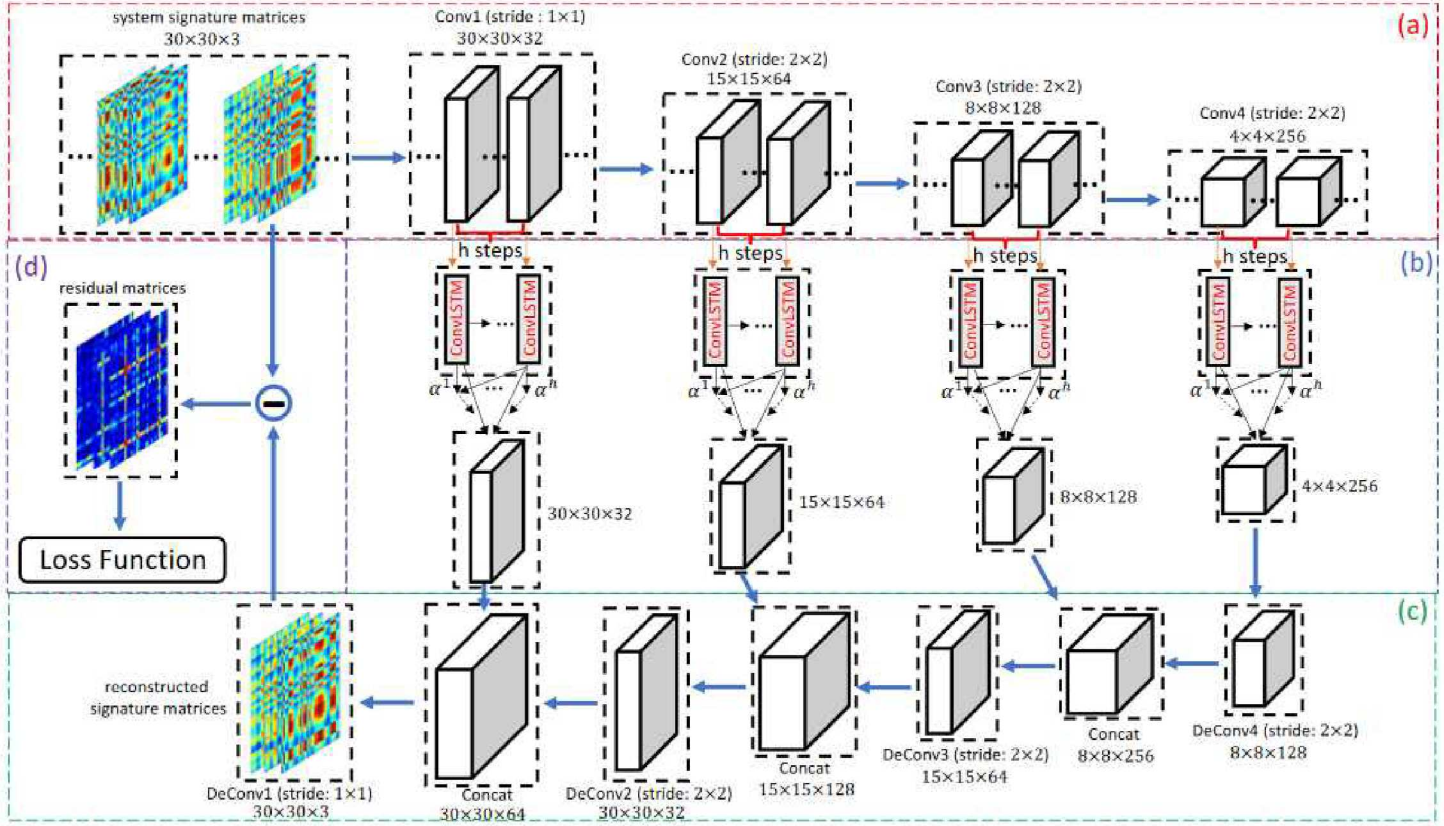
https://github.com/vdumoulin/conv_arithmetic



https://brohrer.github.io/how_convolutional_neural_networks_work.html

Pulling it together – Multi-Scale Recurrent Encoder Decoder (MSCRED)

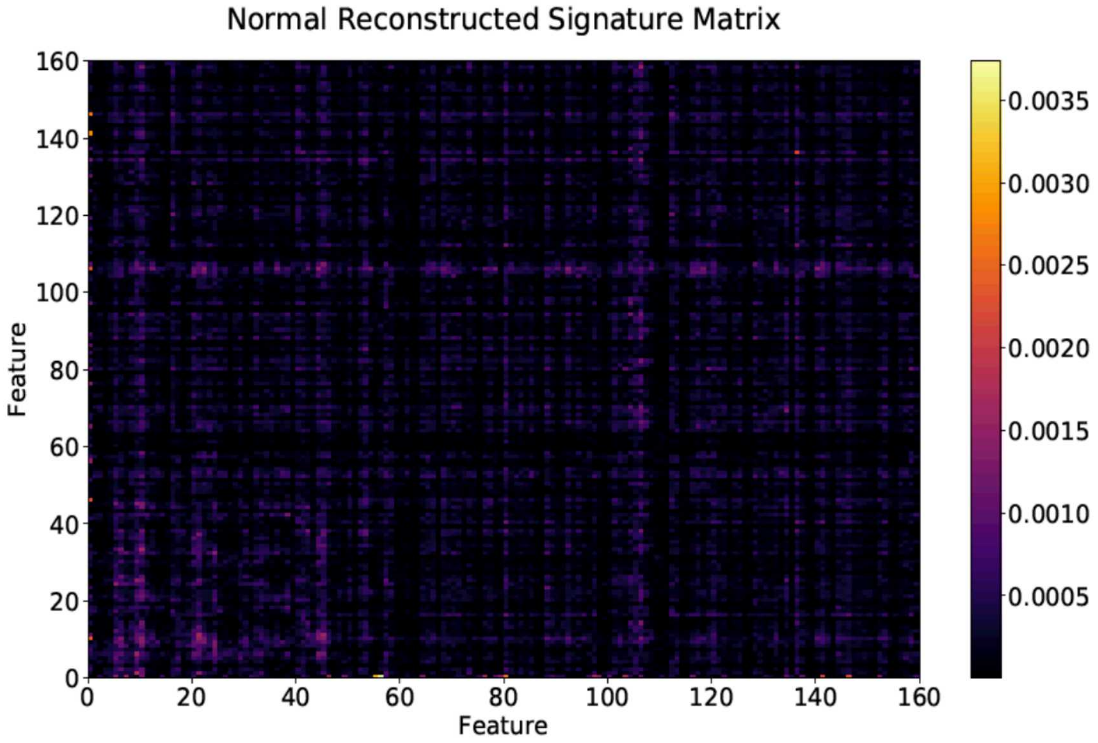
- Multi-Scale Recurrent Encoder Decoder (MSCRED) was developed by Zhang, et al.
- Designed to detect anomalies, provide the location, and provide information on the intensity
- State-of-the-art performance on benchmarks
- Anomalies determined by quantity of pair-wise correlations poorly reconstructed



Zhang, et al. “A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data”, *AAAI 2019*.

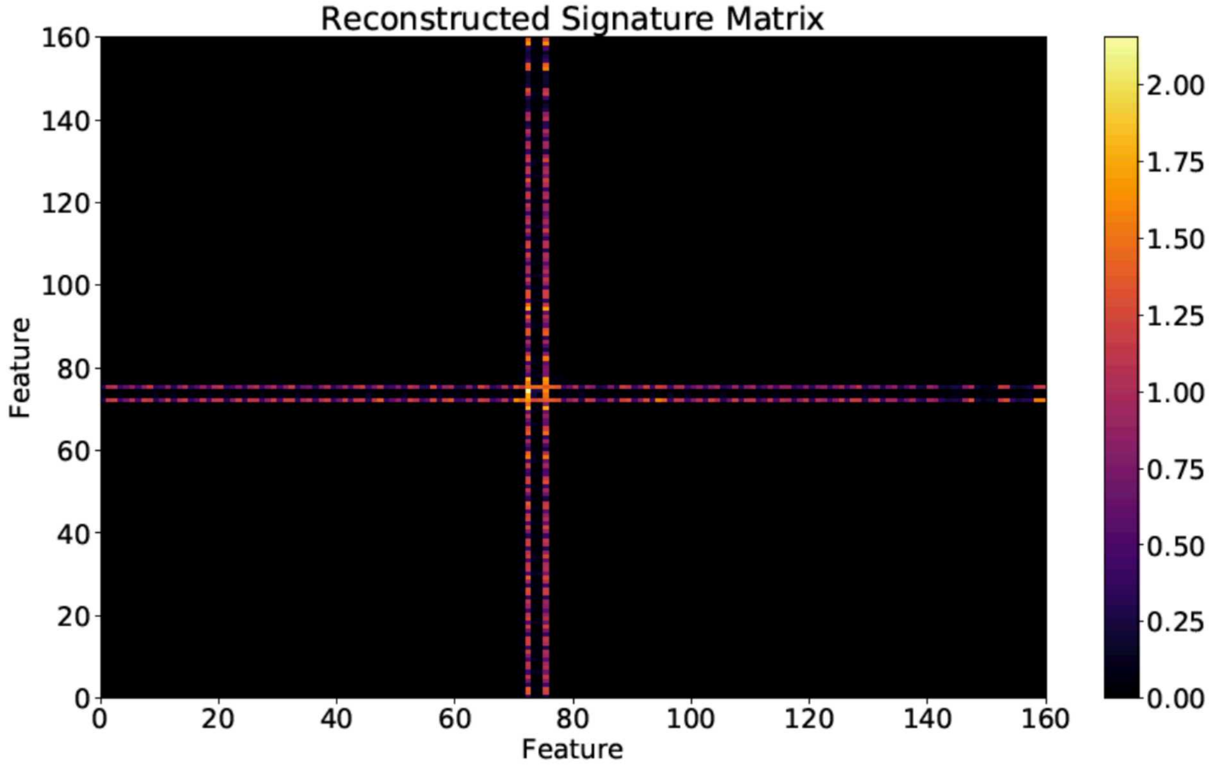
Conditions for initial benchmarking

- GCEP process model has approximately 300 different outputs
  - 160 used as 140 were linearly dependent
  - Final signature matrix size 160x160
- 1% systematic and random error applied to all measurements
- Relies solely on thermophysical measurements outside cascade hall (temperature and pressure)
  - Cylinder properties from accountancy scale to be added in future work



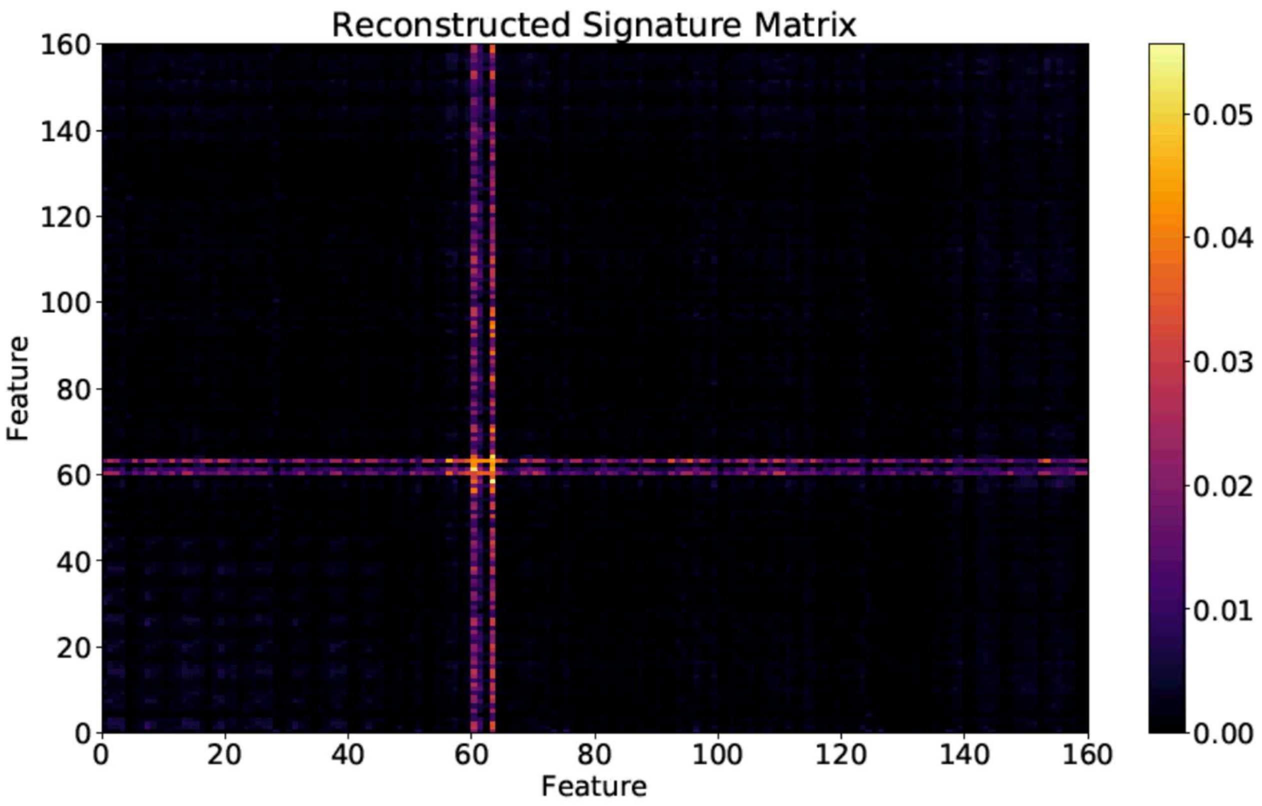
Evaluation case 1 – abrupt material loss

- Abrupt material loss for a single location within the GCEP
- Brightly colored bands indicate high reconstruction error
- Error for anomalous band is 1-2 compared to normal reconstruction error of ~0.003.
- Feature corresponds to location of material loss



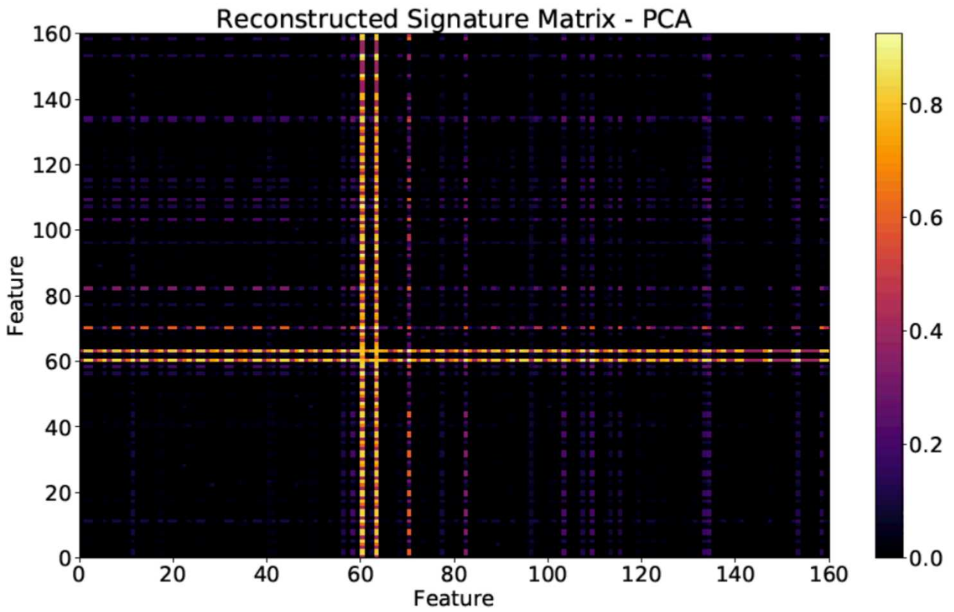
Evaluation case 1 – protracted material loss

- Protracted material loss at single location in GCEP
- Brightly colored bands indicate high reconstruction error
- Error for anomalous band is 0.05 compared to normal reconstruction error of ~0.003 and abrupt of 1-2.
- Note that magnitude of reconstruction error is correlated with magnitude of anomaly



Inconsistent results warrant future work

- Signature matrices inconsistent when detecting anomalies
- Used principal component analysis in similar way as autoencoder
- Results suggest insufficient compression in autoencoder inner layers
  - Future work will improve inner autoencoder layers
- Although PCA has increased results in some cases, it cannot provide information on anomaly magnitude
  - PCA also provides poor normal reconstruction which would inevitably hurt probability of detection
  - PCA limited to linear mapping vs CNN's non-linear mapping



Conclusions and Future Work

- Initial results are promising, but need improvement
- Demonstrated ability to detect, locate, and diagnose anomalies in GCEP dataset
- Considerable future work remains
  - Formalizing probability of detection
  - Increasing MSCRED compression
  - Examining other anomalous scenarios

Acknowledgements

- The authors would also like to acknowledge the National Nuclear Security Administration’s Office of Nuclear Safeguards (NA-241) for funding this work
- There are various open source packages that made this work possible including scikit-learn and Tensorflow.
- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.