NIPS 2017 SUMMARY (HEP PERSPECTIVE)

Savannah Thais,
Yale University
IML Meeting
01/26/2018

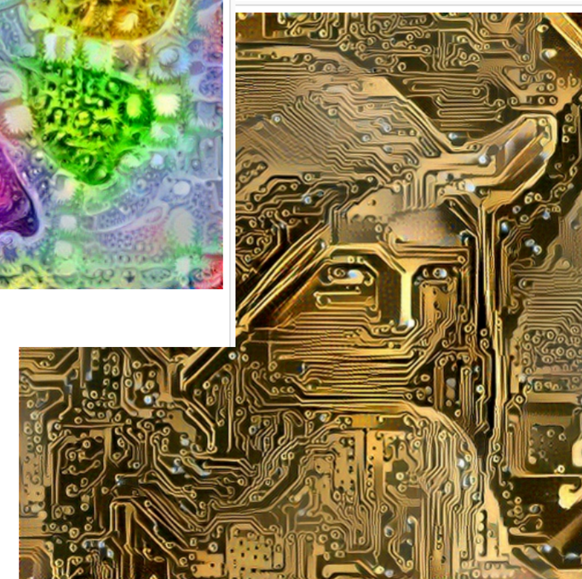
Outline

1. Introduction
2. Deep Learning for Physical Sciences Workshop
3. NIPS Conference
4. Other Interesting Workshops/Symposia

- 2017 Neural Information Processing Systems Conference
- Features main conference + tutorials + symposia + workshops
- ~8000 participants!
- Associated Women in Machine Learning (WiML) and Black in AI Conferences

- Cool Deep Art Contest!

By Raghavendra Adla



By Mehrdad

- ~800 accepted papers, 53 workshops, and 9 tutorials
- All accepted papers are available on website, and featured talks available on Facebook
- Best paper awards:
  - **Safe and Nested Subgame Solving for Imperfect-Information Games:** Noam Brown and Toumas Sandholm
  - **Variance-based Regularization with Convex Objectives:** Hong Namkoong and John C Duchi
  - **A Linear-Time Kernel Goodness-of-Fit Test:** Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, Arthur Gretton

Deep Learning for Physical Sciences

Workshop at the 31st Conference on Neural Information Processing Systems (NIPS)

December 8, 2017

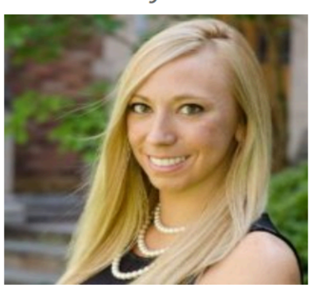
- NIPS Workshop targeting ML researchers interested in scientific applications and researchers in the physical sciences
- 30 accepted papers, 5 invited talks, and 6 contributed talks

Atılım Güneş Baydin
University of Oxford

Prabhat
NERSC, Berkeley Lab

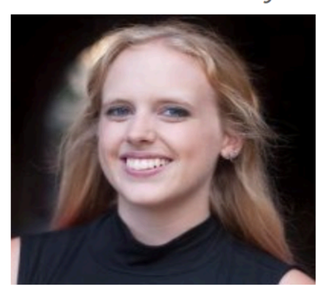
Kyle Cranmer
New York University

Frank Wood
University of Oxford



Michela Paganini
Yale University

Daniela Huppenkothen
New York University

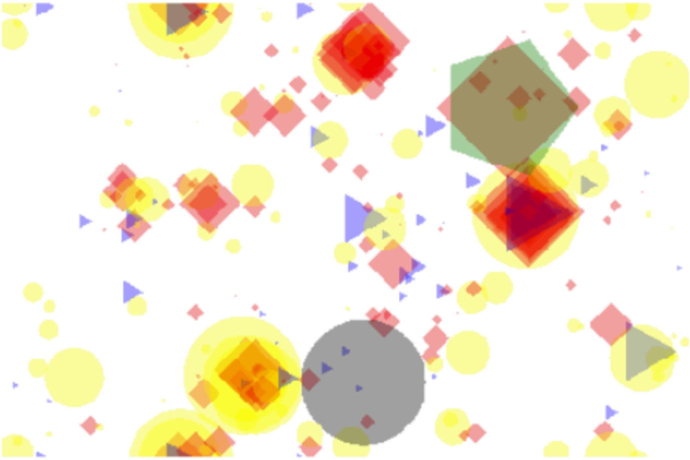


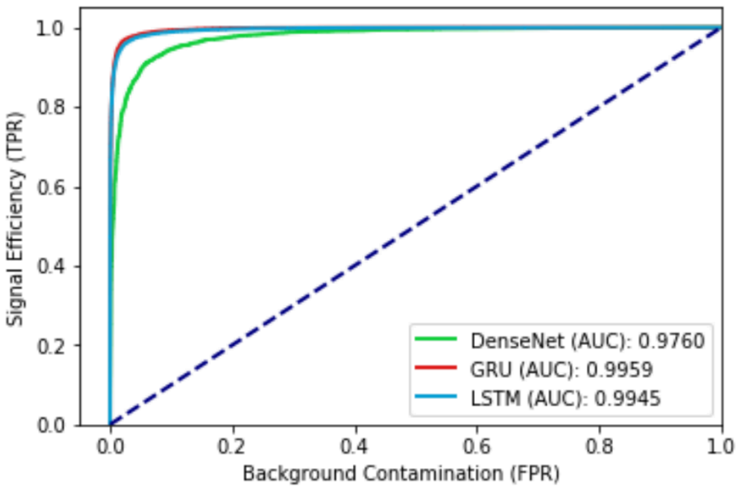
Savannah Thais
Yale University

Ruth Angus
Columbia University

DL4PS LHC Contributions: Classifiers

- Deep Topology Classifiers for a More Efficient Trigger at the LHC: Weitekamp et al
- Represent an event as a sequence of particle flow candidates or an abstract image for improved trigger performance
- Use an RNN to embed particle flow candidates then process with an LSTM or GRU, images are processed with a CNN
- Initial experiment to select ttbar events, currently dominated by W+jet and multijet events when using single lepton trigger



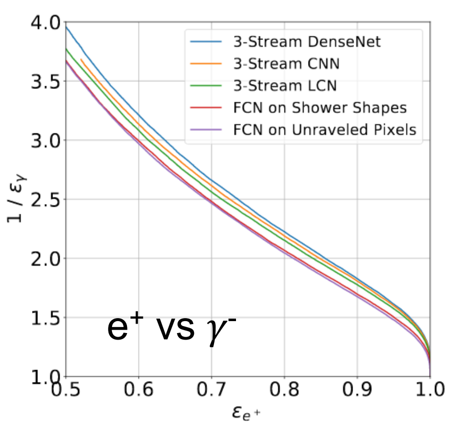


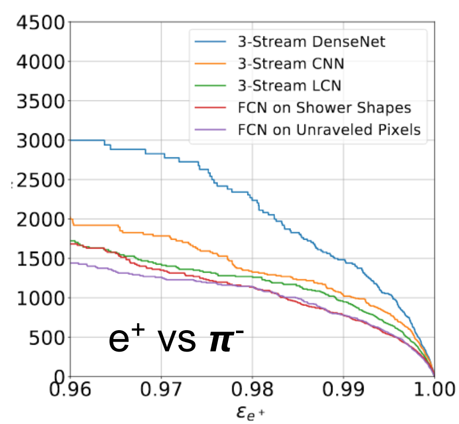
ML Classifiers have much better background rejection and are topology specific!

DL4PS LHC Contributions: Classifiers

Survey of ML Techniques for High Energy Electromagnetic Shower Classification: Paganini, de Oliveira, Nachman

- Proposes multi-stream DenseNet to classify particles from ATLAS calo cell info
- Every layer receives feature map of all previous layers, counters the vanishing gradient problem in CNNs

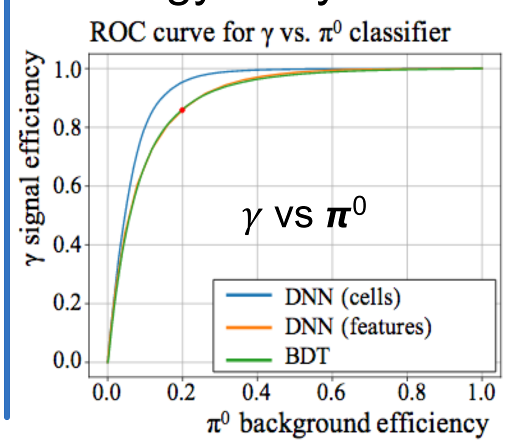


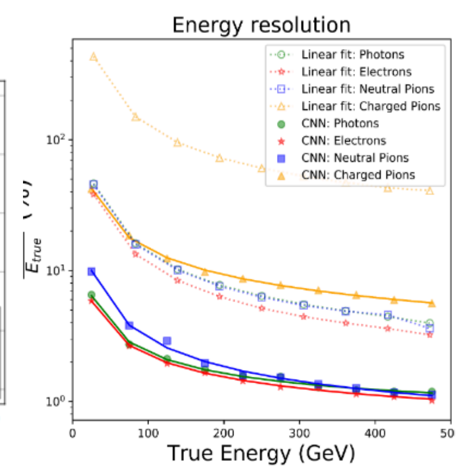


- DenseNet out performs other feature and cell based classifications

Calorimetry with DL: Particle Classification, Energy Regression, and Simulation for HEP: Carminati et al

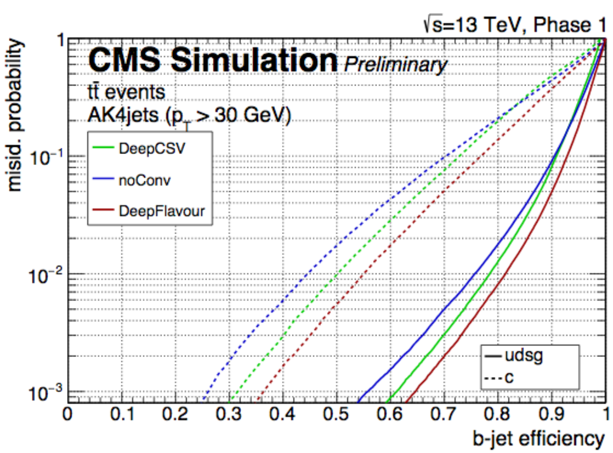
- Explores DL applications for a variety of ATLAS calo based software problems
- Classification DNN using flattened cell information from ecal and hcal
- Energy reconstruction CNN (separate convolutions for ecal and hcal)
- Basic GAN for ecal generating 3D energy arrays

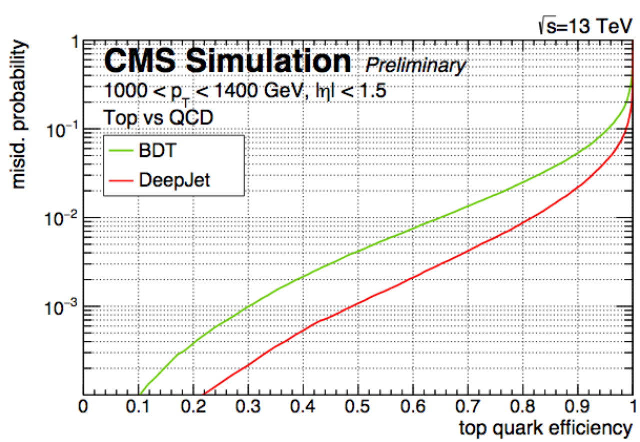


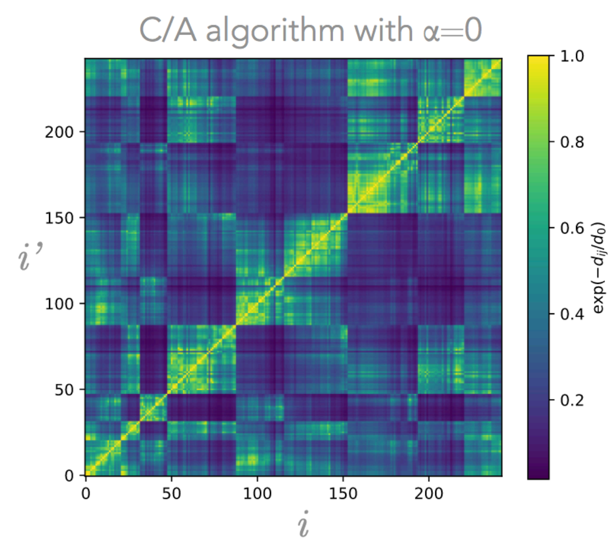


DL4PHS LHC Contributions: Jets

- Tips and Tricks for Training GANs with Physics Constraints: de Oliveira, Paganini, Nachman
  - Nice summary of common issues when using GANs for HEP applications: sparsity, attribute conditioning, training procedures, etc
- DeepJet: Generic Physics Object Based Multiclass Classification for LHC Experiments: Markus Stoye et al
  - CMS CNN jet classifier using particle candidate features, adaptable for wide jets
- Neural Message Passing for Jet Physics: Isaac Henrion et al
  - Graph embedding of jets: node are particles and connecting weights are learned
  - Outperforms previously studied RNN embedding of jets



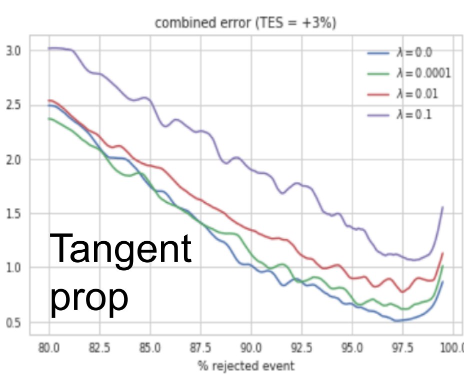


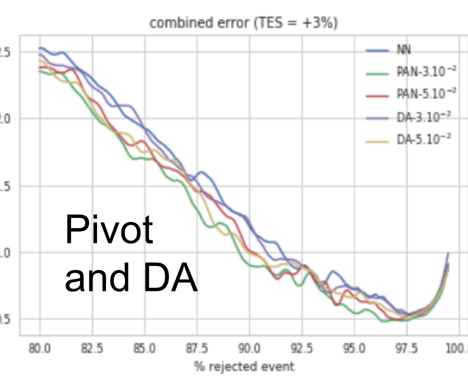


DL4PS LHC Contributions

Adversarial Learning to Eliminate Systematic Errors: a Case Study in HEP: Estrade et al

- Attempts to leverage ML to reduce uncertainties in a cross-section measurement
- Data augmentation, pivot adversarial network, and tangent propagation

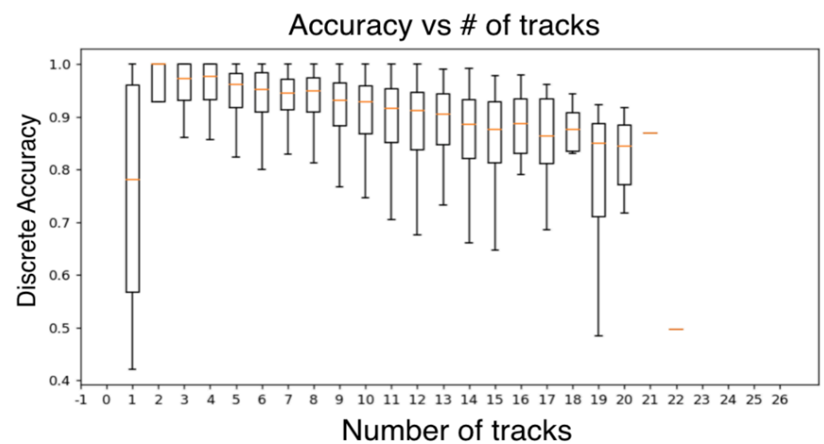




- DA and Pivot outperform baseline, tangent prop does not
- Indicates a knowledge free assumption is ok for HEP applications

Particle Track Reconstruction with Deep Learning: Farrel et al

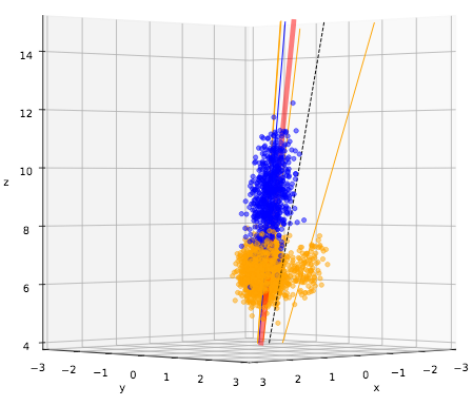
- Imaged based approach to track reconstruction: RNN with individual layers of detector and CNN with image of full detector
- Point based ML: RNN to predict the spacepoint in the next layer and RNN to track class probability assignment vector for each spacepoint

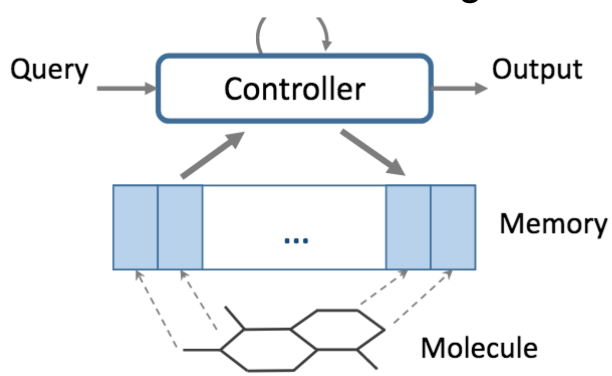


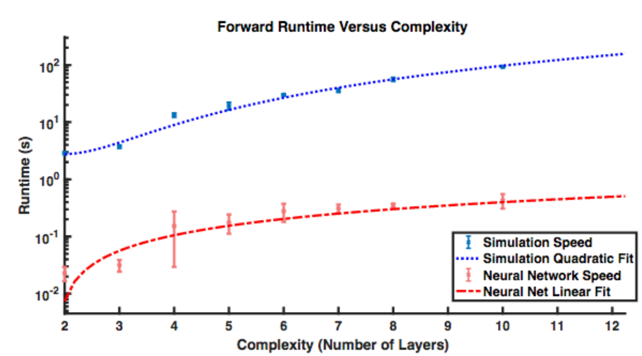
- Spacepoint RNNs more robust to PU

DL4PS: Simulation and Modeling

- Improvements to Inference Compilation for Probabilistic Programming in Large-Scale Scientific Simulators: Casado et al
  - Implements a probabilistic inference package that interfaces with scientific simulators
  - Successfully interfaces with tau decays produced in SHERPA
- Graph Memory Networks for Molecular Activity Prediction: Pham et al
  - Models molecular behavior in target environment as a standard RNN interacting with a matrix RNN (external memory)
- Nanophotonic Particle Simulation and Inverse Design Using Artificial Neural Networks: Peurifoy et al
  - Uses NN to produce a range of spectrum measurements of light scattered off a dielectric: avoids having to solve Maxwell's Equations for multilayer objects
  - Can also run the network 'backwards' to design materials for desired spectrum

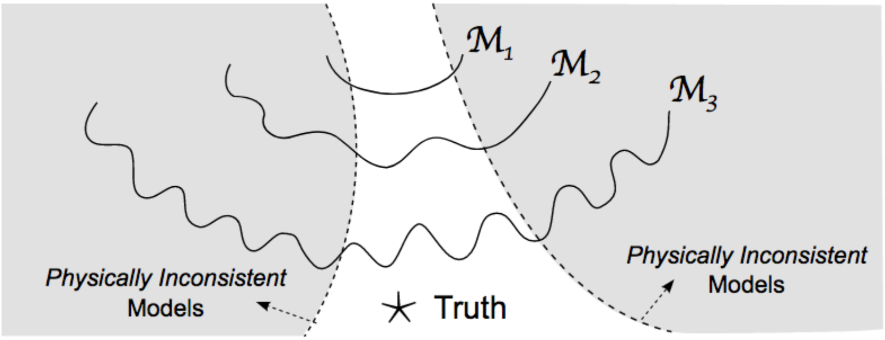


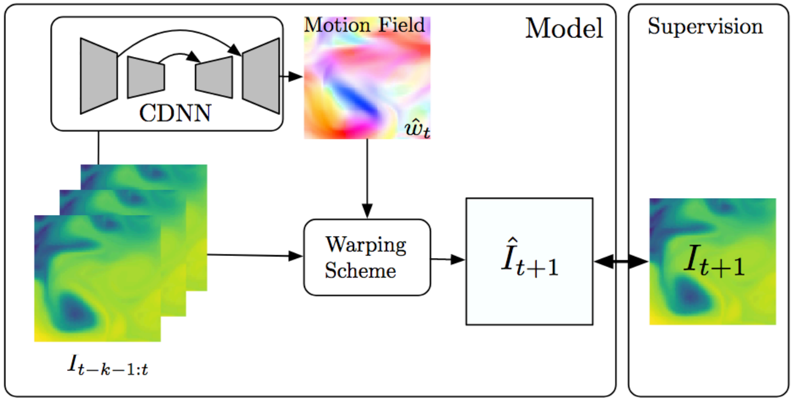




DL4PS: Physics Influenced ML

- How Can Physics Inform Deep Learning Methods in Scientific Problems: Recent Progress and Future Prospects: Karpatne et al
  - Develops a 'Physics Guided NN' that forces NN to be consistent with known physics (useful for developing a model from limited data)
  - Utilizes network design, pre-training, and post-training processing and pruning
  - Successfully applied to lake temperature modeling
- Towards a Hybrid Approach to Physical Process Modeling: Bezenac et al
  - Models complex physics processes underlying ocean currents using CDNN and warping to ultimately predict ocean temperature
  - Allows the algorithm to learn the physics and produces an interpretable latent space





DL4PS Conclusions

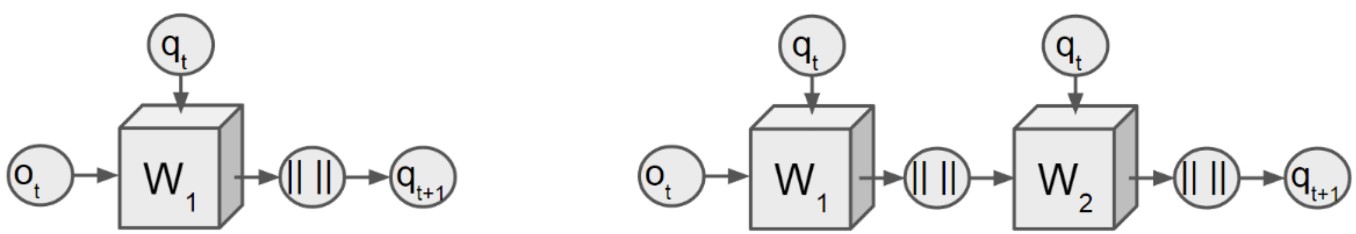
- Good attendance from both scientists and ML experts
- HEP applications were very well represented
  - Nice way to get an overview of current projects and see what areas are uncovered
- Developments in other physical sciences may be easier to adapt to our needs than general ML research
- Seems that a lot of LHC papers use toy datasets…good to see what is feasible but ultimately we want to actually use these techniques in our experiments
- HEP applications tend to focus on classification problems, interesting work in other fields on improving simulations and modeling

NIPS Conference: My Reading List

- What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision: Kendall and Gal
- A Unified Approach to Interpreting Model Predictions: Lundberg and Lee
- Estimating Accuracy from Unlabeled Data: A Probabilistic Logic Approach: Platanios et al
- Spherical Convolutions and Their Application in Molecular Modeling: Boomsma and Frellsen
- Batch Renormalization: Towards Reducing Minibatch Dependence in Batch Normalized Models: Ioffe
- Deep Sets: Zaheer et al
- Learning ReLUs via Gradient Descent: Soltanolkotabi
- Learning Populations of Parameters: Tian et al
- Pixels to Graphs by Associative Embedding: Newell and Dong
- An Empirical Bayes Approach to Optimizing Machine Learning Algorithms: McInerney
- A Dirichlet Mixture Model of Hawkes Processes for Event Cluster Sequencing: Xu and Zha
- Learning to Pivot with Adversarial Networks: Louppe et al
- PixelGAN Autoencoders: Makhzani and Frey
- AdaGAN: Boosting Generative Models: Tolstikhin et al

Predictive Recurrent Neural Networks

- Combine insight from RNNs and Predictive State Representations (PSRs) to produce accurate predictions in dynamical systems (paper)
- Predictive State Recurrent Neural Networks: Downey et al
  - Basic building block is a 3-mode tensor that computes bilinear combination of two input vectors:
  - Tensor W and bias vector b are the $q_{t+1} = \dfrac{W \times_2 o_t \times_3 q_t + b}{\|W \times_2 o_t \times_3 q_t + b\|_2}$ s
  - Tensor contraction of integrates information from q_T and o_T multiplicatively, which acts as a gating mechanism
  - Create multi-layer PSRNN by chaining on the observation



(a) Single Layer PSRNN (b) Multilayer PSRNN

Predictive RNNs

- Initialize the network using a PSR algorithm 2 Stage Regression: $q_{t+1} = (W \times_3 q_t)\,(Z \times_3 q_t)^{-1} \times_2 o_t.$
  - W corresponds to calculating joint distribution, Z corresponds to normalizing it to the conditional distribution
  - Modify 2SR for PSRNN by replacing Z with two term normalization, recover the functional form of single-layer PSRNN
  - Extend to multi-layer by using q_{t+1} predictions as observation inputs to following layer
- Then use Back Propagation Through Time to further refine
- Can also factorize to reduce number of learnable parameters by decomposing W: $W = \sum_{i=1}^{n} a_i \otimes b_i \otimes c_i$

$$\begin{aligned} q_{t+1} &= W \times_2 o_t \times_3 q_t + b \\ &= (A \otimes B \otimes C) \times_2 o_t \times_3 q_t + b \\ &= A^T \left(B o_t \odot C q_t\right) + b \end{aligned}$$

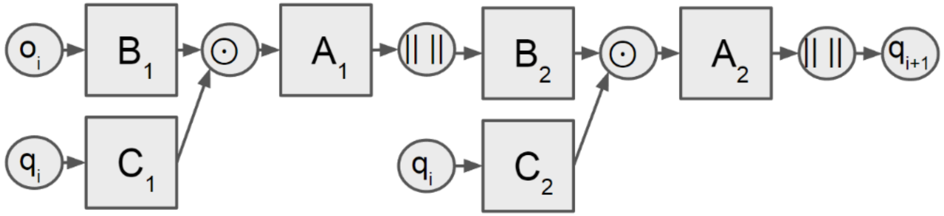
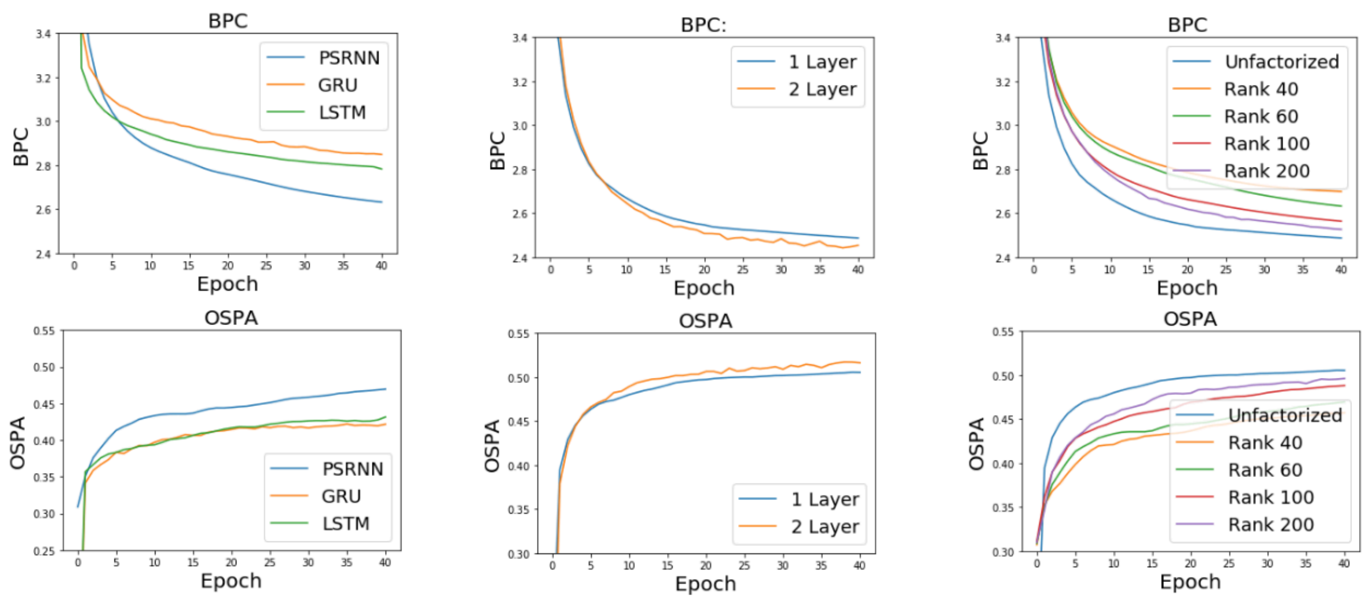


Figure 2: Factorized PSRNN Architecture

Predictive RNNs

- Tested on Penn Tree Bank, OpenAI Robot Swimmer Data, Human Motion Capture Data, and Handwriting Data
- Compared LSTMs, GRUs, RNNs, and PSRNNs



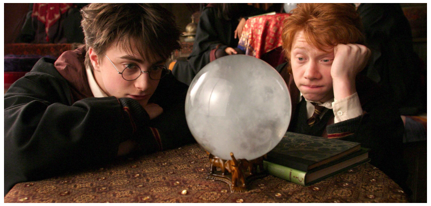
(a) BPC and OSPA on PTB. All models have the same number of states and approximately the same number of parameters.

(b) Comparison between 1- and 2-layer PSRNNs on PTB.

(c) Cross-entropy and prediction accuracy on Penn Tree Bank for PSRNNs and factorized PSRNNs of various rank.

Additional paper: Wang et al

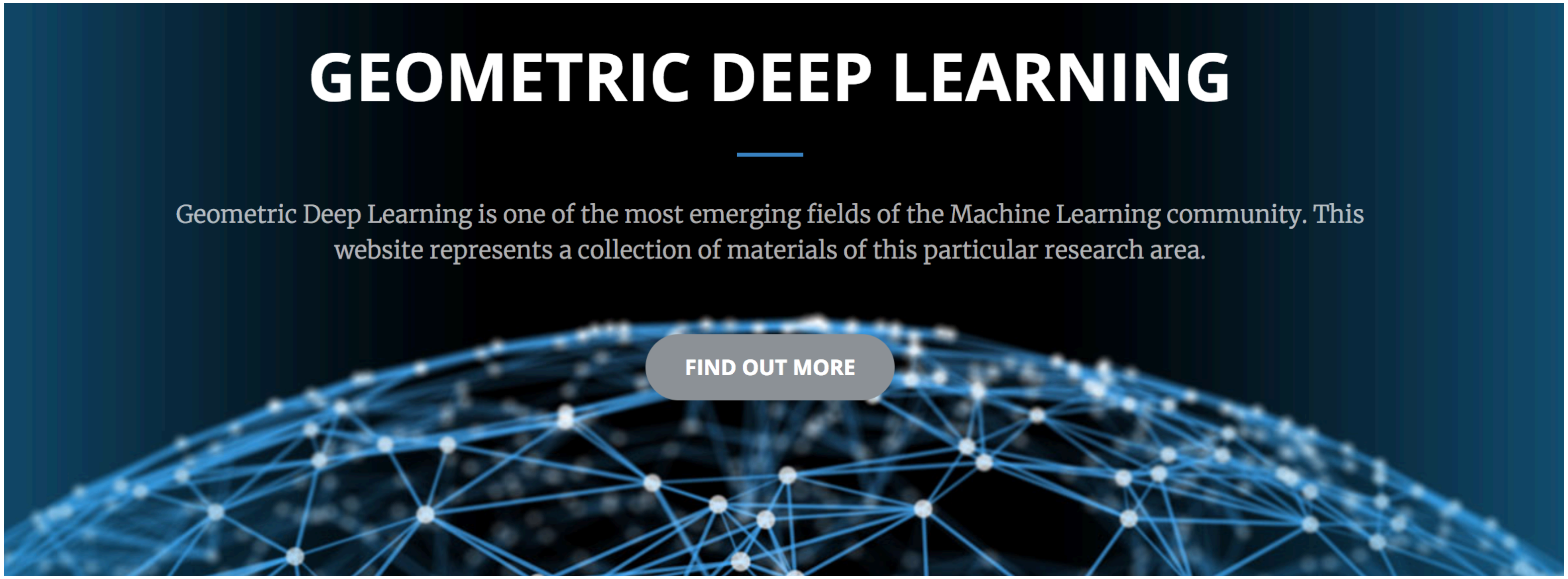
Interpretable ML Symposium



- Relevant to HEP/general scientific applications, as well as AI ethics
- 35 accepted papers, 6 invited talks, 2 panels (website)
- Introduced Explainable Machine Learning Challenge and hosted debate on 'Interpretability is Necessary in Machine Learning'
- Interpretable Discovery in Large Image Datasets: Wagstaff
  - Combines anomaly detection with CNNs to create interpretable explanations of new events
- Debugging the ML Pipeline: Zhu
  - Guidelines for partially automating the tuning process of building an ML model
- Interpretable Deep Learning Applied to Plant Stress Phenotyping: Ghosal et al
  - Uses gradient-weighted class activation mapping to discover which parts of image NN is learning are most relevant to classification
- Learning and Visualizing Localized Geometric Features Using 3D CNN: Ghadi et al
  - 3D Application of gradient-weighted class activation mapping

Machine Learning for Molecules and Materials Workshop

- 22 accepted papers, 14 invited talks sessions, and panel (some are on website)
- Sessions on ML in Chemistry, Kernel Learning with Structured Data, and DL Approaches
- Quantum Machine Learning: von Lilienfeld
  - Exploits underlying chemical redundancies in different molecules
- Neural Network Quantum States: Carleo
  - Reinforcement learning to describe unitary time evolution of interacting quantum systems
- Distilling Expensive Simulations with NNs: Vinyals et al
  - Compresses the knowledge of multiple classifiers into one algorithm
- Learning Hard Quantum Distributions with Variational Autoencoders: Roccheto et al
  - Encodes probability distributions associated with states from different classes



- Applies DL to non-Euclidean geometries: graph and manifold structured data (relevant to physics, bio, etc)
- Nice tutorial at NIPS: introduces the field, describes existing solutions and applications, discusses future directions
  - Including ATLAS shout out!
- Many more resources (papers, tutorials, workshops) on website

Machine Learning in Computational Biology Workshop

- Unfortunately papers and talks are currently not on website
- But some interesting invited talks, could probably dig up papers:
  - A new sparse PCA algorithm with guaranteed asymptotic properties and applications in methylation data: Eran Halperin
  - Denoising scRNA-seq Data Using Deep Count Autoencoders: Gocken Eraslan

Conclusions

- More HEP people attending and submitting to NIPS! This is very important to ensure the robustness of our ML applications and to accurately represent our problems to ML community
- Likely to have annual ML for Physics Workshop
- Much work from main conference and other workshops and symposia can be applicable to our work with clever thinking:
  - A lot of work is being done in interpretability techniques!
  - New boosting, normalization, convolution designs, etc can be implemented in our algorithm design
  - More techniques can be incorporated in our simulation chain to reduce computation time and mis-modeling
- Come to NIPS next year (and register early)!