

AI Seminar

General information: Meetings now usually take place on **Fridays at 1pm** on Zoom, but please check the schedule.

Please contact Finn O'Shea (foshea@slac.stanford.edu), Auralee Edelen (edelen@slac.stanford.edu) or Kazuhiro Terao (kerao@slac.stanford.edu) if you are interested in giving a talk or want to join the mailing list!

Also check out the main AI/ML web page: <https://ml.slac.stanford.edu/>

- Upcoming Seminars
 - Fast AI at the edge for particle physics
 - The Learnt Geometry of Collider Events
- Past Seminars
 - Reconstructing the Subhalo Mass Function from Strong Gravitational Lensing using Simulation-Based Inference
 - "All the Lenses": Toward Large-Scale Hierarchical Inference of the Hubble Constant Using Bayesian Deep Learning
 - Black-box optimisation with Local Generative Surrogates and its application in the SHiP experiment
 - Vector Symbolic Architectures for Autonomous Science
 - Bayesian Techniques for Accelerator Characterization and Control
 - Computational Imaging: Reconciling Physical and Learned Models
 - Deep Learning for Anomaly Detection
 - Machine-Learning for Modeling Complex Materials and Media
 - Autonomous analysis of synchrotron X-ray experiments with applications to metal nanoparticle synthesis
 - Going Beyond Global Optima with Bayesian Algorithm Execution
 - Signal Decomposition via Distributed Optimization
 - Equitable Valuation of Data
 - Neural Networks with Feature Sparsity
 - Machine Learning for Big Data Cosmology and High Energy Physics
 - Beyond Deep Learning in Fundamental Physics
 - Machine Learning for Dark Matter
 - Searching for dark matter in the sky with machine learning
 - Online Bayesian Optimization for the SECAR Recoil Mass Separator
 - Quantum Kernel Methods for the Classification of High-dimensional Data on a Superconducting Processor
 - Machine Learning with Quantum Computers
 - Reservoir computing using digital logic gate networks
 - Power efficient hardware accelerators for machine learning, combinatorial optimization, and pattern matching applications
 - Generative Models and Symmetries
 - Multi-Objective Bayesian Optimization for Accelerator Tuning
 - Machine Learning Techniques for Optics Measurements and Corrections
 - Superconducting Radio-Frequency Cavity Fault Classification Using Machine Learning at Jefferson Laboratory
 - Analytical and Parametric Model Fitting for Inverse Problems, Data Reduction, and Pattern Recognition
 - Deep Learning and Quantum Gravity
 - Bayesian Optimization and Machine Learning for Accelerating Scientific Discovery
 - Physics-informed machine learning for accelerated modeling and optimization of complex systems
 - Probabilistic Programming for Inverse Problems in Physical Sciences
 - Discovering Symbolic Models in Physical Systems using Deep Learning
 - Anomaly Detection in Particle Accelerators using Autoencoders
 - Multi-CryoGAN: Reconstruction of Continuous Conformations in Cryo-EM Using Generative Adversarial Networks
 - Neuromorphic Computing: Where Hardware Meets AI
 - Sequence-guided protein structure determination using graph convolutional and recurrent networks
 - Machine Learning-based Beam Size Stabilization at ALS
 - High-dimensional geometry and the landscapes of deep neural networks
 - Learning Hyperbolic Representations for Unsupervised 3D Segmentation
 - Model-based machine learning: from image restoration to 3D particle segmentation
 - Containers! Containers! Containers!
 - PLAsTiCC: Convincing other people to solve your problems
 - Real-time classification of explosive transients using deep recurrent neural networks
 - Shape learning
 - GPU, Slurm and SLAC-Stanford-Scientific Data Facility
 - Black Box Variational Inference: Scalable, Generic Bayesian Computation and its Applications
 - Pushing the Limits of Fluorescence Microscopy with adaptive imaging and machine learning
 - Machine learning applications of quantum annealing in high energy physics
 - NASA Ames Data Sciences Group Overview
 - A Topology Layer for Machine Learning
 - Accelerating Data Science Workflows with RAPIDS
 - Photometric classification of astronomical transients for LSST
 - Analyzing and Applying Uncertainty in Deep Learning
 - Data-driven Discovery of the Governing Equations of Complex Physical Systems
 - Sparse Submanifold Convolution for Physics 2D/3D Image Analysis
 - Applying Convolutional Neural Networks to MicroBooNE
 - Machine Learning, Datascience and Neutrino Physics at Argonne's Leadership Computing Facility
 - Deep Learning for Particle Track Finding in High Energy Physics
 - Machine Learning for medical applications of Physics
 - Machine Learning synthetic data, scanning probe data, and reciprocal space data on quantum materials
 - Local-to-Global Methods for Topological Data Analysis
 - Experience with a Virtual Multi-Slit Phase Space Diagnostic at Fermilab's FAST Facility
 - A Novel Approach - IoT Device Virtualization using ML
 - Sparsity/Undersampling Tradeoffs in Compressed Sensing

- Learned predictive models: integrating large-scale simulations and experiments using deep learning
- Rapid Gaussian Process Training via Structured Low-Rank Kernel Approximation of Gridded Measurements
- Machine learning applications for hospitals
- Beyond Data and Model Parallelism for Deep Neural Networks
- X-ray spectrometer data processing with unsupervised clustering (Sideband signal seeking)
- Experience with FEL taper tuning using reinforcement learning and clustering
- Statistical Learning of Reduced Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics
- Machine Learning for Jet Physics at the Large Hadron Collider
- In situ visualization with task-based parallelism
- Data Reconstruction Using Deep Neural Networks for Liquid Argon Time Projection Chamber Detectors
- Towards a cosmology emulator using Generative Adversarial Networks
- Optimal Segmentation with Pruned Dynamic Programming
- Fast automated analysis of strong gravitational lenses with convolutional neural networks
- MacroBase: A Search Engine for Fast Data Streams
- Object-Centric Machine Learning
- Reconstruction Algorithms for Next-Generation Imaging: Multi-Tiered Iterative Phasing for Fluctuation X-ray Scattering and Single-Particle Diffraction
- Exploratory Studies in Neural Network-based Modeling and Control of Particle Accelerators
- Estimating behind-the-meter solar generation with existing measurement infrastructure
- Development and Application of Online Optimization Algorithms
- Machine Learning at NERSC: Past, Present, and Future
- Optimization for Transportation Efficiency
- Detecting Simultaneous Changepoints Across Multiple Data Sequences
- Low Data Drug Discovery with One-Shot Learning
- Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis
- Models and Algorithms for Solving Sequential Decision Problems under Uncertainty
- Data Programming: A New Framework for Weakly Supervising Machine Learning Models
- ProxImaL: Efficient Image Optimization using Proximal Algorithms
- Energy-efficient neuromorphic hardware and its use for deep neural networks
- Locating Features in Detector Images with Machine Learning
- Tractable quantum leaps in battery materials and performance via machine learning
- Deep Learning and Computer Vision in High Energy Physics
- Label-Free Supervision of Neural Networks with Physics and Domain Knowledge
- Can machine learning teach us physics? Using Hidden Markov Models to understand molecular dynamics.
- On-the-fly unsupervised discovery of functional materials
- Machine Learning and Optimization to Enhance the FEL Brightness
- Automated tuning at LCLS using Bayesian optimization
- Using Deep Learning to Sort Down Data

Upcoming Seminars

Fast AI at the edge for particle physics

Date: January 21, 2021 1:00 pm Pacific

Speaker: Jennifer Ngadiuba (FNAL)

The Large Hadron Collider at CERN provides up to 200 proton-proton interactions every 25 ns leading to the production of thousands of charged and neutral particles per second passing through the detector volume. As the detectors consist of hundreds of millions of sensors to record the passage of each particle, the experiments at the LHC have to deal with extreme data rates of hundreds of TB per second. To bring down these rates to manageable levels for offline processing and storage, the experiments implement a trigger system that analyze and accept collision events in real-time. There is a fundamental challenge in doing so due to the very strict latency and amount of resources available to perform such analysis. Therefore, in order to preserve most of the interesting physics, basic algorithms are executed on Field-Programmable Gate Arrays (FPGA). Most recently we have started exploring and developing new AI techniques to replace these rules with an advanced analysis that could offer enhanced accuracy while meeting such strict system constraints. In this talk, I will discuss the recent developments in the field of fast AI to achieve such goal, focusing on the application to LHC experiments where the "big data" environment is among the most challenging in HEP.

The Learnt Geometry of Collider Events

Date: January 28, 2021 1:00 pm Pacific

Speaker: Jack Collins (SLAC)

Particle collider events, when imbued with a metric which characterizes the 'distance' between two events (such as an Earth Movers Distance), can be thought of as populating a data manifold in a metric space. The geometric properties of this manifold reflect the physics encoded in the distance metric. I will show how the geometry of collider events can be probed at varying scales of interest using a class of machine learning architectures called Variational Autoencoders. I will introduce notions of scaling dimensionality of representations learnt by the VAE that I believe are novel, and which reflect and quantify the underlying complexity of the training dataset. If there is time, I will also describe two potentially novel approaches to unsupervised classification that are inspired by these notions of dimensionality.

Past Seminars

Reconstructing the Subhalo Mass Function from Strong Gravitational Lensing using Simulation-Based Inference

Date: November 19, 2021 1:00 pm Pacific

Speaker: Sebastian Wagner-Carena (Stanford)

Constraining the distribution of small-scale structure in our universe will allow us to probe alternatives to the cold dark matter (CDM) paradigm. Strong gravitational lensing offers a unique window into small dark matter halos because these halos impart a gravitational lensing signal even if they do not host luminous galaxies. However, the millions of free parameters in gravitational lensing by a substructure population makes directly evaluating the likelihood intractable. In this talk, I will present our group's work using simulation-based inference techniques to return posterior estimates of the distribution of subhalos inside galaxy-mass host halos. We combine a hierarchical inference approach with some of the tools used in sequential neural posterior estimation to reliably infer the subhalo mass function across a variety of configurations. We find that our technique scales efficiently to large lens populations; with 10 strong gravitational lenses we forecast a constraining power competitive with current flux ratio statistics, and with 100 lenses we find that our technique returns sensitivities comparable with current Milky Way satellite constraints. In the 1000 lens regime accessible by future surveys, we demonstrate an unprecedented constraining power on the subhalo mass function. Our work reveals the potential of strong lensing imaging to probe dark matter at small scales.

[Video](#)

"All the Lenses": Toward Large-Scale Hierarchical Inference of the Hubble Constant Using Bayesian Deep Learning

Date: October 29, 2021 1:00 pm Pacific

Speaker: Ji Won Park (Stanford)

Precise constraints on the Hubble constant (H_0) can shed light on the nature of dark matter and dark energy, arguably the biggest mysteries of modern cosmology. An astrophysical phenomenon known as strong gravitational lensing enables direct measurements of H_0 . Seven strong gravitational lenses have been "hand-analyzed" over the last ten years, but next-generation telescope surveys will increase the sample size to tens of thousands of lenses, creating a demand for novel methods that can model large volumes of noisy data. I demonstrate the use of Bayesian neural networks (BNNs) in rapidly extracting cosmological information from the image, catalog, and time series data associated with these lenses. Quantifying various sources of uncertainty is key to minimizing systematic bias on H_0 . Being both accurate and efficient, the BNN pipeline is a promising tool that can combine information from all the lenses -- with varying types and signal-to-noise ratios -- into a large-scale hierarchical Bayesian model.

[Video](#)

Black-box optimisation with Local Generative Surrogates and its application in the SHiP experiment

Date: October 22, 2021 10:00 am Pacific

Speaker: Sergey Shirobokov (Twitter)

We propose a novel method for gradient-based optimisation of black-box simulators using local surrogate models (<https://arxiv.org/abs/2002.04632>). In domains such as HEP, many processes are modeled with non-differentiable simulators (such as GEANT4). However, often one wants to optimise some parameters of the detector or other apparatus relying on the knowledge from the simulator. To address such cases, we utilise deep generative models to approximate a simulator in the local neighbourhood and perform optimisation. In cases when the optimised parameter space is constrained to a low dimension sub-space, we observe that our method outperforms Bayesian optimisation, numerical optimisation, and REINFORCE-based approaches.

[Video](#)

Vector Symbolic Architectures for Autonomous Science

Date: October 8, 2021 1:00 pm Pacific

Speaker: Michael Furlong (University of Waterloo)

Automating exploration often involves information theoretic cost functions which can be expensive to compute. Planetary missions are constrained by size, weight, and power concerns, as well as environmental conditions, that limit the type and amount of computing that can be deployed on these missions.

Neuromorphic computing promises to reduce power requirements needed for deploying high-performance computing, enabling constrained systems to be more capable, but they can be challenging to program. Vector Symbolic Architectures, originally developed in the context of cognitive modelling, have proven useful as a paradigm for programming these computers.

In this talk we will be discussing how a particular Vector Symbolic Architecture can be used to efficiently execute two tasks commonly found in autonomous science applications: anomaly detection and Bayesian optimization. We will show how these algorithms can be computed with time and memory complexity that is constant in the number of observations collected, making them favourable algorithms for long-term operations in resource constrained computing environments.

Bayesian Techniques for Accelerator Characterization and Control

Date: October 1, 2021 1:00 pm Pacific

Speaker: Ryan Roussel (SLAC National Accelerator Laboratory)

Accelerators and other large experimental facilities are complex, noisy systems that are difficult to characterize and control efficiently. Bayesian statistical modeling techniques are well suited to this task, as they minimize the number of experimental measurements needed to create robust models, by incorporating prior, but not necessarily exact, information about the target system. Furthermore, these models inherently consider noisy and/or uncertain measurements and can react to time-varying systems. Here we will describe several advanced methods for using these models in accelerator characterization and optimization. First, we describe a method for rapid, turn-key exploration of input parameter spaces using little-to-no prior information about the target system. Second, we highlight how these models can take hysteresis effects into account and create in-situ models of individual magnetic elements.

[Video](#)

Computational Imaging: Reconciling Physical and Learned Models

Date: July 2, 2021 1:00 pm Pacific

Speaker: Ulugbek Kamilov (Washington University in St. Louis)

Computational imaging is a rapidly growing area that seeks to enhance the capabilities of imaging instruments by viewing imaging as an inverse problem. There are currently two distinct approaches for designing computational imaging methods: model-based and learning-based. Model-based methods leverage analytical signal properties and often come with theoretical guarantees and insights. Learning-based methods leverage data-driven representations for best empirical performance through training on large datasets. This talk presents Regularization by Artifact Removal (RARE), as a framework for reconciling both viewpoints by providing a learning-based extension to the classical theory. RARE relies on pre-trained "artifact-removing deep neural nets" for infusing learned prior knowledge into an inverse problem, while maintaining a clear separation between the prior and physics-based acquisition model. Our results indicate that RARE can achieve state-of-the-art performance in different computational imaging tasks, while also being amenable to rigorous theoretical analysis. We will focus on the applications of RARE in biomedical imaging, including magnetic resonance and tomographic imaging.

This talk will be based on the following references:

J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," IEEE J. Sel. Topics Signal Process., vol. 14, no. 6, pp. 1088-1099, October 2020.

Z. Wu, Y. Sun, A. Matlock, J. Liu, L. Tian, and U. S. Kamilov, "SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors," IEEE J. Sel. Topics Signal Process., vol. 14, no. 6, pp. 1163-1175, October 2020.

J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg, and U. S. Kamilov, "SGD-Net: Efficient Model-Based Deep Learning with Theoretical Guarantees," IEEE Trans. Comput. Imag., vol. 7, pp. 598-610, June 2021.

[Video](#)

Deep Learning for Anomaly Detection

Date: June 25, 2021 1:00 pm Pacific

Speaker: Ziyi Yang (Stanford)

Anomaly Detection (AD) refers to the process of identifying abnormal observations that deviate from what is defined as normal. With applications in many real-world scenarios, anomaly detection has become an important research field in ML and AI. However, detecting anomalies in high-dimensional space is challenging. In some high-dimensional cases, previous AD algorithms fail to correctly model the normal data distribution. Also the understanding on the detection mechanism of AD models remained limited. To address these challenges and questions, in this talk, first I will present the Regularized Cycle-consistent GAN (RCGAN) that introduces a penalty distribution in the modeling of normal data distribution. We theoretically show that the penalty distribution regularizes the discriminator and generator towards the normal data manifold. Second, we explore anomaly detection with domain adaptation where the normal data distribution is non-static. We propose to extract the common features of source and target domain data and train an anomaly detector using the extracted features.

Slides and [video](#).

Machine-Learning for Modeling Complex Materials and Media

Date: June 18, 2021 1:00 pm Pacific

Speaker: Serveh Kamrava (USC)

In recent years, machine learning (ML) approaches have made it possible to extract and explore intricate patterns from big data. One of the fields that can benefit from the computational advantages that ML offers is materials characterization where we have complex heterogeneous morphology. The morphology of complex systems is one of the determinant elements that control a variety of their properties, such as flow, transport, and mechanical behaviors. Such properties are often estimated using experimental and computational methods, which can be very costly and time-demanding. As such, faster and more automatic methods are required. Machine learning provides an alternative solution for this problem. In this presentation, I will present a deep learning method that can take the 3D morphology of complex materials and estimate their transport properties. Then, I will talk about a novel method using which one can quantify the accuracy of augmentation methods for adding more data to ML and identify the method that can provide the best set of data by minimizing the discrepancy and expanding the variability. For the next topic, I will discuss the application of deep learning for dynamic data when they change with time for a transport problem on a complex membrane system. I close this particular topic by describing how the governing equations can be used in ML for filling the gap in data and reducing the amount of data for ML. These results will be compared with a fully data-driven ML method.

Autonomous analysis of synchrotron X-ray experiments with applications to metal nanoparticle synthesis

Date: May 7, 2021 1:00 pm Pacific

Speaker: Sathya Chitturi (Stanford)

A critical step in developing autonomous pipelines for materials synthesis experiments is automatic interpretation of characterization experiments. In this talk, we present an example of a closed-loop bayesian optimization pipeline for metal nanoparticle synthesis using real-time information from Small-angle X-ray Scattering (SAXS) experiments. This approach has previously successfully created libraries of monodisperse Pd nanoparticles with user-specified sizes. In addition, we describe a CNN-based method used to interpret complementary X-ray diffraction data. Here CNN regression models are trained for each crystal class to predict lattice parameters for the corresponding unit-cell. A key component of this work involves data augmentation schemes which capture sources of experimental noise in order to improve model generalizability. The lattice parameter estimates are subsequently refined using an automatic whole-pattern fitting algorithm.



SLAC_AI_Seminar_Sathya_final.pdf

[Video](#)

Going Beyond Global Optima with Bayesian Algorithm Execution

Date: April 30, 2021 1:00 pm Pacific

Speaker: Willie Neiswanger (Stanford)

In many real world problems, we want to infer some property of an expensive black-box function f , given a budget of T function evaluations. One example is budget constrained global optimization of f , for which Bayesian optimization is a popular method. Other properties of interest include local optima, level sets, integrals, or graph-structured information induced by f . Often, we can find an algorithm A to compute the desired property, but it may require far more than T queries to execute. Given such an A , and a prior distribution over f , we refer to the problem of inferring the output of A using T evaluations as Bayesian Algorithm Execution (BAX). In this talk, we present a procedure for this task, InfoBAX, that sequentially chooses queries that maximize mutual information with respect to the algorithm's output. Applying this to Dijkstra's algorithm, for instance, we infer shortest paths in synthetic and real-world graphs with black-box edge costs. Using evolution strategies, we yield variants of Bayesian optimization that target local, rather than global, optima. We discuss InfoBAX, and give background on other information-based methods for Bayesian optimization as well as on the probabilistic uncertainty models which underlie these methods.



BAX - SLAC AI Se... - Neiswanger.pdf

[Video](#)

Signal Decomposition via Distributed Optimization

Date: April 23, 2021 1:00 pm Pacific

Speaker: Bennet Meyers (Stanford)

We consider the well-studied problem of decomposing a time series signal into some components, each with different characteristics. We propose a simple and general framework for decomposition of a signal into a number of signal classes, each defined by a loss function and possibly constraints, via optimization. We describe a number of useful signal classes, and give a distributed optimization method for computing the decomposition, that scales well and is extensible. The method finds the optimal decomposition when the signal class constraints and loss functions are convex, and appears to be a good heuristic when they are not.



Meyers_Signal_D...composition.mp4



Meyers_SignalDe...resentation.pdf

Equitable Valuation of Data

Date: April 16, 2021 1:00 pm Pacific

Speaker: Amirata Ghorbani (Stanford)

As data becomes the fuel driving technological and economic growth, a fundamental challenge is how to quantify the value of data in algorithmic predictions and decisions. For example, in healthcare and consumer markets, it has been suggested that individuals should be compensated for the data that they generate, but it is not clear what is an equitable valuation for individual data. In this talk, we discuss a principled framework to address data valuation in the context of supervised machine learning. Given a learning algorithm trained on a number of data points to produce a predictor, we propose data Shapley as a metric to quantify the value of each training datum to the predictor performance. Data Shapley value uniquely satisfies several natural properties of equitable data valuation. We introduce Monte Carlo and gradient-based methods to efficiently estimate data Shapley values in practical settings where complex learning algorithms, including neural networks, are trained on large datasets. We then briefly discuss the notion distributional Shapley, where the value of a point is defined in the context of underlying data distribution.



Equitable Valuation of Data.pdf



Ghorbani_Equita...ationofData.mp4

Neural Networks with Feature Sparsity

Date: April 2, 2021 1:00 pm Pacific

Speaker: Ismael Lemhadri (Stanford)

Much work has been done recently to make neural networks more interpretable, and one approach is to arrange for the network to use only a subset of the available features. In linear models, Lasso (or L1-regularized) regression assigns zero weights to the most irrelevant or redundant features, and is widely used in data science. However the Lasso only applies to linear models. Here we introduce LassoNet, a neural network framework with global feature selection. Our approach enforces a hierarchy: specifically a feature can participate in a hidden unit only if its linear representative is active. Unlike other approaches to feature selection for neural nets, our method uses a modified objective function with constraints, and so integrates feature selection with the parameter learning directly. As a result, it delivers an entire regularization path of solutions with a range of feature sparsity. On systematic experiments, LassoNet significantly outperforms state-of-the-art methods for feature selection and regression. The LassoNet method uses projected proximal gradient descent, and generalizes directly to deep networks. It can be implemented by adding just a few lines of code to a standard neural network.



IsmaelLemhadri_20210402.mp4



LassoNetSlides.pdf

Machine Learning for Big Data Cosmology and High Energy Physics

Date: February 23, 2021 1:00 pm Pacific

Speaker: Agnes Ferte

In the context of future galaxy surveys such as the Legacy Survey of Space and Time (LSST), I proposed an application of unsupervised learning algorithms such as Self-Organizing Maps to efficiently explore the theory space of cosmological models. In the first part of my talk, I will explain the challenges motivating this research and present our first results aiming at categorizing theories of gravity probed by weak gravitational lensing, one of the main cosmological observables that will be measured by LSST. Many experiments of the FPD at SLAC present computational challenges such as data reduction on the fly or physics simulations that require similar machine learning applications and developments. In the second part of my talk, I will present how I will expand the use of unsupervised learning algorithms to other areas at the FPD and contribute to the application of machine learning to LSST, other cosmology experiments and high energy physics experiments.

Beyond Deep Learning in Fundamental Physics

Date: February 16, 2021 1:00 pm Pacific

Speaker: Lukas Heinrich

The experiments at the Large Hadron Collider (LHC) are testament to the success of the reductionist approach to science: the analytical modelling of the 100 million data channels of HEP is patently hard but through a deep, hierarchical stack of simulation across many length and energy-scales and a physics-driven, expert-designed dimensionality reduction procedure, inference on the fundamental parameters of quantum field theory is achievable. In recent years, advancements in Machine Learning techniques have provided physicists promising new tools to analyze the LHC data. To exploit them fundamental questions need to be addressed: How do we formulate ML optimization goals to align with our science goals? How can we translate known constraints in the data into appropriate inductive biases of the trained algorithms? Can we express and incorporate uncertainties and maintain interpretability to achieve safe inference? In light of these challenges I will discuss in this talk recent progress in end-to-end gradient-based optimization, Active Learning, simulator-assisted probabilistic programming.

Machine Learning for Dark Matter

Date: February 12, 2021 1:00 pm Pacific

Speaker: Bryan Ostdiek (Harvard University)

There is five times more dark matter than ordinary matter in the universe, but we have almost no idea what it is. To learn about the possible interactions of dark matter, physicists use complementary data from cosmological probes, astroparticle observations, and particle colliders. There is an increasing need for advanced analytics and machine learning to process these vastly growing datasets. This talk details examples using machine learning in each of the three realms. First, I demonstrate using image recognition techniques on images of strongly lensed galaxies to constrain dark matter properties. Second, I use machine learning to uncover the phase space distribution of dark matter near the Earth, which directly impacts the interpretation of direct detection experiments. Finally, I examine how unsupervised learning methods can aid collider searches for dark matter. The talk concludes with comments on the intersection of machine learning and physics.

Searching for dark matter in the sky with machine learning

Date: February, 2021 1:00 pm Pacific

Speaker: Siddharth Mishra Sharma (New York University)

The next decade will see a deluge of new cosmological data that will enable us to accurately map out the distribution of matter in the local Universe, image billion of stars and galaxies to unprecedented precision, and create high-resolution maps of the Milky Way. Signatures of new physics may be hiding in these observations, offering significant discovery potential for uncovering physics beyond the Standard Model, in particular the nature of dark matter. At the same time, the complexity of astrophysical data provides significant challenges to carrying out these searches using conventional methods. I will describe how overcoming these issues will require a qualitative shift in how we approach modeling and inference in cosmology, connecting particle physics properties to cosmological observables and bringing together several recent advances in machine learning and simulation-based inference. I will present several applications of these methods. I will show how they can be used to combine information from tens of thousands of strong gravitational lensing systems in order to infer structural properties of our Universe that can be directly linked to the microphysical properties of dark matter. Finally, I will present an application to the long-standing problem of understanding the nature of the Galactic Center gamma-ray excess, highlighting challenges associated with analyzing real data and discussing ways to overcome them

For the slides and the recording of Siddharth's seminar, [contact Kazu](#) as it was requested not to make a publicly-open access.

Online Bayesian Optimization for the SECAR Recoil Mass Separator

Date: December 11, 2020 11:00 am Pacific

Speaker: Sara Miskovich (Michigan State University)

The SEparator for CAPture Reactions (SECAR) is a next-generation recoil separator system under commissioning at the National Superconducting Cyclotron Laboratory (NSCL) and Facility for Rare Isotope Beams (FRIB) at Michigan State University. SECAR is optimized for the direct measurement of capture reactions on unstable nuclei that drive some stars to explode and synthesize crucial nuclei that make up our universe. Once SECAR is operational, these precise measurements will improve our understanding of astrophysical processes such as X-ray bursts, novae and supernovae. To maximize the performance of the device, ion optical optimizations and careful beam alignment need to be achieved, which can be time consuming and difficult to achieve through manual tuning. This talk will focus on the first development of an online Bayesian optimization that utilizes a Gaussian process model to tune the beam through the complex system and improve its ion optical properties by optimizing magnet settings. The method is shown to improve recoil separator performance and save operational time for future scientific experiments.

Quantum Kernel Methods for the Classification of High-dimensional Data on a Superconducting Processor

Date: December 11, 2020 1:00 pm Pacific

Speaker: Evan Peters (Fermilab, University of Waterloo IQC)

We present a quantum kernel method for high-dimensional data analysis using the Google Sycamore superconducting quantum computer architecture. Our experiment utilizes the largest number of qubits to date compared to prior quantum kernel method experiments. We study an application in the domain of cosmology - a benchmark supernova type classification problem using 67 features with no dimensionality reduction and without vanishing kernel elements. While most experimental work to date has considered synthetic datasets of low dimension, and disregarded the importance of shot statistics and mean kernel element size, we show that the analysis of real, high dimensional datasets requires careful attention to these features when constructing a circuit

Machine Learning with Quantum Computers

Date: December 4, 2020 10:00 am Pacific

Speaker: Maria Schuld (Xanadu, University of KwaZulu-Natal)

A growing number of papers are searching for intersections between High Energy Physics and the emerging field of Quantum Machine Learning. This talk gives an introduction to the latter, while critically discussing potential connections to HEP. A focus lies on the most popular approach to machine learning with quantum computers, which interprets quantum circuits as machine learning models that load input data and produce predictions. By optimizing the quantum circuit, the "quantum model" can be trained like a neural network. To offer a glimpse of the opportunities and challenges of this approach, I will discuss different aspects of such "variational quantum machine learning algorithms", including their close links to kernel methods and integration into modern machine learning pipelines.



20201204_Schuld.pdf



GMT20201204-1...0x1080(1).mp4

Reservoir computing using digital logic gate networks

Date: November 20, 2020 11:00 am Pacific

Speaker: Heidi Komkov (The Institute for Research in Electronics and Applied Physics, University of Maryland)

As Moore's law is coming to an end, new types of computing architectures must be explored to continue the pace of advancement in computing power. At the same time, applications of machine learning are exploding. Reservoir computing is a brain-inspired machine learning method which has shown promise for very rapid time series prediction. The reservoir functions as a recurrent neural network, and substituting a physical system for a computer-based simulation has the potential to allow computation at high speed and very low power. We use an autonomous Boolean network as a reservoir, which uses individual CMOS digital logic gates to implement the nonlinear elements used in machine learning architectures. In this talk I'll show results from an field programmable gate array (FPGA) reservoir and my designs of a 180nm application specific integrated circuit (ASIC) that has been fabricated this year

Power efficient hardware accelerators for machine learning, combinatorial optimization, and pattern matching applications

Date: November 13, 2020 11:00 am Pacific

Speaker: Cat Graves (Hewlett Packard Labs)

The dramatic rise of data-intensive workloads has revived special-purpose hardware and architectures for continuing improvements in computational speed and energy efficiency. While traditional CMOS ASICs deliver some performance gains, typically by limiting data movement or implementing "in-memory computation", such approaches still suffer from low power efficiency. New proposals leveraging emerging non-volatile resistive RAM (ReRAM) devices for in-memory computation are highly attractive in a variety of application domains. While originally developed for as digital (binary) high density non-volatile memories, ReRAM devices have demonstrated a wide range of behaviors and properties – such as a wide range of tunable analog resistance and non-linear dynamics – which motivate their use in novel functions and new computational models. Many recent in-memory compute studies have focused on crossbar circuit architectures, demonstrating their application for neural networks, scientific computing and signal processing. However, other circuit primitives – such as content addressable memories (CAMs) and combined systems such as crossbar arrays and non-linear elements– have shown further promise for mapping a diverse range of complimentary computational models such as finite state machines, pattern matching, hashing algorithms and Hopfield neural networks for tackling optimization problems. In this talk, I will review the exciting opportunities for in-memory computational primitives leveraging non-volatile ReRAM devices and their circuits and architectures for enabling low power, high-throughput computation in a variety of application domains. Recent lab demonstrations of various applications mapped to these in-memory computational circuit primitives based on memristor devices will be shown and I will also give an outlook on performance.

Generative Models and Symmetries

Date: November 5, 2020 10:00 am Pacific

Speaker: Danilo Rezende (Google DeepMind)

The study of symmetries in Physics has revolutionized our understanding of the world. Inspired by this, I will focus on our recent work on incorporating Gauge symmetries into normalizing flow generative models and its potential applications in the sciences and ML.

Multi-Objective Bayesian Optimization for Accelerator Tuning

Date: October 30, 2020 1:00 pm

Speaker: Ryan Roussel (University of Chicago)

Particle accelerators require constant tuning during operation to meet beam quality, total charge and particle energy requirements for use in a wide variety of physics, chemistry and biology experiments. Maximizing the performance of an accelerator facility often necessitates multi-objective optimization, where

operators must balance trade-offs between multiple objectives simultaneously, often using limited, temporally expensive beam observations. Usually, accelerator optimization problems are solved offline, prior to actual operation, with advanced beamline simulations and parallelized optimization methods (NSGA-II, Swarm Optimization). Unfortunately, it is not feasible to use these methods for online multi-objective optimization, since beam measurements can only be done in a serial fashion, and these optimization methods require a large number of measurements to converge to a useful solution. Here, we introduce a multi-objective Bayesian optimization scheme, which finds the full Pareto front of an accelerator optimization problem efficiently in a serialized manner and is thus a critical step towards practical online multi-objective optimization in accelerators. This method uses a set of Gaussian process surrogate models, along with a multi-objective acquisition function, which reduces the number of observations needed to converge by at least an order of magnitude over current methods. We demonstrate how this method can be modified to specifically solve optimization challenges posed by the tuning of accelerators. This includes the addition of optimization constraints, objective preferences and costs related to changing accelerator parameters.



20201030_Roussel.pdf



GMT20201030-2...2736x1824.mp4

Machine Learning Techniques for Optics Measurements and Corrections

Date: October 28, 2020 8:00 am

Speaker: Elena Fol (CERN)

Recently, the application of ML has grown in accelerator physics, in particular in the domain of diagnostics and control. One of the first applications of ML at the LHC is focused on optics measurements and corrections. Unsupervised Learning has been applied to automatic detection of beam position monitors faults to improve optics analysis, demonstrating successful results in operation. A novel ML-based approach for the estimation of magnet errors is developed, using supervised regression models trained on a large set of LHC optics simulations. Also, autoencoder neural networks have found their application in denoising of measurements data and reconstruction of missing data points. The results and future plans for these studies will be discussed following a brief introduction to relevant ML concepts.



SLAC_seminar_slides.pdf

Superconducting Radio-Frequency Cavity Fault Classification Using Machine Learning at Jefferson Laboratory

Date: October 23, 2020 1:00 pm

Speaker: Chris Tennant (Jefferson Laboratory)

We report on the development of machine learning models for classifying C100 superconducting radio-frequency (SRF) cavity faults in the Continuous Electron Beam Accelerator Facility (CEBAF) at Jefferson Lab. CEBAF is a continuous-wave recirculating linac utilizing 418 SRF cavities to accelerate

electrons up to 12 GeV through 5-passes. Of these, 96 cavities (12 cryomodules) are designed with a digital low-level RF system configured such that a cavity fault triggers waveform recordings of 17 RF signals for each of the 8 cavities in the cryomodule. Subject matter experts (SME) are able to analyze the collected time-series data and identify which of the eight cavities faulted first and classify the type of fault. This information is used to find trends and strategically deploy mitigations to problematic cryomodules. However manually labeling the data is laborious and time-consuming. By leveraging machine learning, near real-time – rather than post-mortem – identification of the offending cavity and classification of the fault type has been implemented. We discuss the development and performance of the ML models as well as valuable lessons learned in bringing a ML system to deployment.

Analytical and Parametric Model Fitting for Inverse Problems, Data Reduction, and Pattern Recognition

Date: October 21, 2020 8:00 am

Speaker: Youssef Nashed (ANL, Stats Perform)

Many scientific and engineering challenges can be formulated as fitting a model to existing data. Whether it is comparing a scientific simulation to known experimental observations, finding a continuous representation of sparse/discrete data points, or the values of model parameters which generalize to unforeseen data examples given historical data; all these tasks share a common underlying principle of model fitting, but with different choices made in the model formulation (parametric or analytical) and the assumptions made about the data (acquisition scheme, noise to signal ratio, continuity, or information locality). In this talk I will highlight a few use cases under this framework. Specifically, I will address research conducted at Argonne National Laboratory for X-ray image reconstruction problems, data reduction for scientific simulations, and deep learning approaches for replacing expensive iterative optimization. Additionally, I will present more recent work for sports computer vision applications that enable real time player detection, tracking, and activity prediction from broadcast video.

Deep Learning and Quantum Gravity

Date: October 15, 2020 4:00 pm

Speaker: Koji Hashimoto (Osaka University)

Formulating quantum gravity is one of the final goals of fundamental physics. Recent progress in string theory brought a concrete formulation called AdS/CFT correspondence, in which a gravitational spacetime emerges from lower-dimensional non gravitational quantum systems, but we still lack in understanding how the correspondence works. I discuss similarities between the quantum gravity and deep learning architecture, by regarding the neural network as a discretized spacetime. In particular, the questions such as, when, why and how a neural network can be a space or a spacetime, may lead to a novel way to look at machine learning. I implement concretely the AdS/CFT framework into a deep learning architecture, and show the emergence of a curved spacetime as a neural network, from a given training data of quantum systems.



Hashimoto20201015.pdf



GMT20201015-2...1686x1024.mp4

Bayesian Optimization and Machine Learning for Accelerating Scientific Discovery

Date: October 9, 2020 1:00 pm

Speaker: Stefano Ermon (Stanford)

Applications of AI in the physical sciences require new advances in representing, reasoning about, and acquiring knowledge from data and domain expertise. Motivated by these challenges, I will present new approaches for calibrating ML systems so that predicted probabilities are more reflective of real-world uncertainty, i.e., better capture what is or isn't known by the system. I will discuss approaches to automatically acquire data to reduce uncertainty through maximally informative experiments, focusing on the design of charging protocols for electric batteries and other challenging problems in science and engineering. Finally, I will discuss opportunities for incorporating domain knowledge to further accelerate the process.



GMT20201009-2...0x1440(1).mp4

Physics-informed machine learning for accelerated modeling and optimization of complex systems

Date: October 2, 2020 1:00 pm

Speaker: Paris Perdikaris (University of Pennsylvania)

The towering empirical success of machine learning is promising a pathway for transforming observations to actionable knowledge. Specific to modeling and optimizing complex physical and engineering systems, there is a need for methods that can seamlessly synthesize data of variable fidelity, leverage prior domain knowledge, respect the laws of physics, and provide robust predictions with quantified uncertainty. In this talk I will provide an overview of data-driven techniques that aim to address these needs, and highlight their advantages and limitations through the lens of different application studies. Specifically, we will discuss the effectiveness of Gaussian processes in integrating multi-fidelity data to accelerate the prediction of large scale computational models, as well as the potential of physics-informed deep learning models in tackling a diverse range of forward and inverse problems in computational physics. Finally, I will also discuss the role of predictive uncertainty in closing the observations-to-predictions loop as a proxy for judicious data acquisition and experimental design.



Paris_slides.pdf



GMT20201002-1..._1760x900.mp4

Probabilistic Programming for Inverse Problems in Physical Sciences

Date: September 25, 2020 1:00 pm

Speaker: Atillim Gunes Baydin (University of Oxford)

Machine learning enables new approaches to inverse problems in many fields of science. We present a novel probabilistic programming framework that couples directly to existing scientific simulators through a cross-platform probabilistic execution protocol, which allows general-purpose inference engines to record and control random number draws within simulators in a language-agnostic way. The execution of existing simulators as probabilistic programs enables highly interpretable posterior inference in the structured model defined by the simulator code base. We demonstrate the technique in particle physics, on a scientifically accurate simulation of the tau lepton decay, which is a key ingredient in establishing the properties of the Higgs boson.

Inference efficiency is achieved via amortized inference where a deep recurrent neural network is trained to parameterize proposal distributions and control the stochastic simulator in a sequential importance sampling scheme, at a fraction of the computational cost of a Markov chain Monte Carlo baseline.



Probabilistic Pro...ical Sciences.pdf



GMT20200925-1...1920x1080.mp4

Discovering Symbolic Models in Physical Systems using Deep Learning

Date: September 18, 2020 1:00 pm
Speaker: Shirley Ho (Flatiron Institute)

We develop a general approach to distill symbolic representations of a learned deep model by introducing strong inductive biases. We focus on Graph Neural Networks (GNNs). The technique works as follows: we first encourage sparse latent representations when we train a GNN in a supervised setting, then we apply symbolic regression to components of the learned model to extract explicit physical relations. We find the correct known equations, including force laws and Hamiltonians, can be extracted from the neural network. We then apply our method to a non-trivial cosmology example---a detailed dark matter simulation---and discover a new analytic formula that can predict the concentration of dark matter from the mass distribution of nearby cosmic structures. The symbolic expressions extracted from the GNN using our technique also generalized to out-of-distribution-data better than the GNN itself. Our approach offers alternative directions for interpreting neural networks and discovering novel physical principles from the representations they learn.

Anomaly Detection in Particle Accelerators using Autoencoders

Date: September 11, 2020 1:00 pm
Speaker: Jonathan Edelen (RadiaSoft, LLC)

The application of machine learning (ML) techniques for anomaly detection in particle accelerators has gained popularity in recent years. These efforts have ranged from the analysis of quenches in RF cavities [1, 2] and superconducting magnets [3] to anomalous beam position monitors [4], and even losses in rings [5]. Using ML for anomaly detection can be challenging owing to the inherent imbalance in the amount of data collected during normal operations as compared to during faults. Additionally, the data are not always labeled and therefore supervised learning is not possible. Autoencoders, neural networks that form a compressed representation and reconstruction of the input data, are a useful tool for such situations. Here we explore the use of autoencoders for two types of problems: dimensionality reduction and reconstruction analysis. In the former case, we study machine data from the Fermilab LINAC and correlate changes in the RF parameters to changes in beam loss. For the latter case, we also study the Fermilab LINAC but extend this work to the evaluation of magnet faults in the APS storage ring. [1] A. S. Nawaz, S. Pfeiffer, G. Lichtenberg, and H. Schlarb, "Self-organized critical control for the european xfel using black box parameter identification for the quench detection system," in 2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol), Sep. 2016, pp. 196–201. [2] A. Nawaz, S. Pfeiffer, G. Lichtenberg, and P. Rostalski, "Anomaly detection for the european xfel using a nonlinear parity space method," IFAC-PapersOnLine, vol. 51, no. 24, pp. 1379 – 1386, 2018, 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2018. [3] M. Wielgosz, A. Skoczea, and M. Mertik, "Using lstm recurrent neural networks for monitoring the lhc superconducting magnets," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 867, pp. 40 –50, 2017. [4] Elena Fol, "Evaluation of Machine Learning Methods for LHC Optics Measurements and Corrections Software" CERN-THESIS-2017-336, Aug 2017. [5] G. Valentino, R. Bruce, S. Redaelli, R. Rossi, P. Theodoropoulos, and S. Jaster-Merz, "Anomaly detection for beam loss maps in the large hadron collider," Journal of Physics: Conference Series, vol. 874, p. 012002, Jul 2017.



AnomalyDetection_jedelen.pdf



GMT20200911-2...1920x1200.mp4

Multi-CryoGAN: Reconstruction of Continuous Conformations in Cryo-EM Using Generative Adversarial Networks

Date: September 4, 2020 1:00 pm

Speaker: Harshit Gupta

In this talk, I will present Multi-CryoGAN, a deep-learning-based reconstruction method for cryo-electron microscopy (Cryo-EM). It can reconstruct continuous conformations of a biomolecule from Cryo-EM images in a fully unsupervised and standalone manner. Cryo-EM produces many noisy projections from separate instances of the same but randomly oriented biomolecule. Current methods rely on pose and conformation estimation which are inefficient for the reconstruction of continuous conformations that carries valuable information. Multi-CryoGAN sidesteps the additional processing by casting the volume reconstruction into the distribution matching problem. By introducing a manifold mapping module, Multi-CryoGAN can learn continuous structural heterogeneity without pose estimation nor clustering. It is also backed by a theoretical guarantee of recovery of the true conformations. This method can successfully reconstruct 3D protein complexes on synthetic 2D Cryo-EM datasets for both continuous and discrete structural variability scenarios.

Neuromorphic Computing: Where Hardware Meets AI

Date: August 21, 2020 10:00 am

Speaker: Catherine Schuman (Oak Ridge National Lab)

Though neuromorphic systems were introduced decades ago, there has been a resurgence of interest in recent years due to the looming end of Moore's law, the end of Dennard scaling, and the tremendous success of AI and deep learning for a wide variety of applications. With this renewed interest, there is a diverse set of research ongoing in neuromorphic computing, ranging from novel hardware implementations, device and materials to the development of new training and learning algorithms. There are many potential advantages to neuromorphic systems that make them attractive in today's computing landscape, including the potential for very low power, efficient hardware that can perform neural network computation. In this talk, an overview of the current state of neuromorphic computing will be presented, including a brief background on neuromorphic models, algorithms, hardware, and applications in the literature. An approach for training neuromorphic systems will be described, and several real-world applications will be discussed.



Schuman_SLAC_S...August2020.pdf



GMT20200821-1..._1760x900.mp4

Sequence-guided protein structure determination using graph convolutional and recurrent networks

Date: August 14, 2020 1:00pm (remote)

Speaker: Po-Nan Li (Stanford)

Single particle imaging performed at cryogenic electron microscopy (cryo-EM) facilities, including the S2C2 at SLAC, now routinely outputs high-resolution data for large proteins and their complexes. Building an atomic model into a cryo-EM density map, however, remains challenging, particularly when no structure for the target protein is known a priori. Existing protocols for this type of task often rely on significant human intervention and can take hours to days to produce an output. Here, we present a fully automated, template-free model building approach that is based entirely on neural networks. We use a graph convolutional network (GCN) to generate an embedding from a set of rotamer-based amino acid identities and candidate 3-dimensional C-alpha locations. Starting from this embedding, we use a bidirectional long short-term memory (LSTM) module to order and label the candidate identities and atomic locations consistent with the input protein sequence to obtain a protein structural model. Our approach paves the way for determining protein structures from cryo-EM densities at a fraction of the time of existing approaches and without the need for human intervention.

Machine Learning-based Beam Size Stabilization at ALS

Date: August 7, 2020 1:00pm (remote)

Speaker: Simon Leeman (LBNL)

In state-of-the-art synchrotron light sources the overall source stability is limited by the achievable level of electron beam size stability. This source size stability is presently on the few- percent level, which is still 1-2 orders of magnitude larger than already demonstrated stability of source position/angle (slow /fast orbit feedbacks) and current (top-off injection). Until now, source size stabilization has been achieved through corrections based on a combination of static predetermined physics models and lengthy calibration measurements (feed-forward tables), periodically repeated to counteract drift in the accelerator and instrumentation. We now demonstrate for the first time [PRL 123 194801 (2019)], how application of machine learning allows for a physics- and model-independent stabilization of source size relying only on previously existing instrumentation in ALS. Such feed-forward correction based on neural networks that can be continuously online-retrained achieves source size stability as low as 0.2 microns rms (0.4%) which results in overall source stability approaching the sub-percent noise floor of the most sensitive experiments.



ML-based_Beam_...AI_Seminar.pdf

High-dimensional geometry and the landscapes of deep neural networks

Date: July 31, 2020 1:00pm (remote)

Speaker: Stanislav Fort (Stanford)

When we train a deep neural network on a dataset using gradient descent, we are exploring an extremely high-dimensional landscape of weight configurations looking for a rare solution to our task, while using only the local gradients as a guide. Given how complicated these landscape can be, how exactly do deep neural networks manage to converge to good, generalizable solutions at all, and can we say anything more concrete about the types of landscapes they navigate during training? In this talk, I will focus on recent geometric insights into the structure of neural network loss landscapes -- I will discuss a phenomenological approach to modelling their large-scale structure [1,2], and its consequences for ensembling, calibration, uncertainty estimates and Bayesian methods in general [3]. I will conclude with an outlook on several interesting open questions in understanding artificial deep networks. [1] Fort, Stanislav, and Adam Scherlis. "The Goldilocks zone: Towards better understanding of neural network loss landscapes." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019. arXiv 1807.02581, [2] Stanislav Fort, and Stanislaw Jastrzebski. "Large Scale Structure of Neural Network Loss Landscapes." Advances in Neural Information Processing Systems 32 (NeurIPS 2019). arXiv 1906.04724, [3] S Fort, H Hu, B Lakshminarayanan. "Deep Ensembles: A Loss Landscape Perspective." arXiv 1912.02757

Learning Hyperbolic Representations for Unsupervised 3D Segmentation

Date: June 12, 2020 1:00pm (remote)

Speaker: Joy Hsu (Stanford)

There exists a need for unsupervised 3D segmentation on complex volumetric data in the case of limited annotations or for tasks of object discovery -- especially in biomedical fields. We efficiently learn representations for unsupervised segmentation through *hyperbolic embeddings* that model hierarchy innate to 3D input. Our method learns hyperbolic representation through a novel *gyroplane convolutional layer* as well as a *hierarchical triplet loss*, and retrieves multi-level segmentations from clustering on hyperbolic space. We show the effectiveness of our method on three biologically-inspired datasets, including one on cryogenic electron microscopy (cryo-EM), with images supplied by SLAC.



slac ml seminar.pdf



GMT20200612-2..._1760x900.mp4

Model-based machine learning: from image restoration to 3D particle segmentation

Date: May 21st, 2020 1:00pm (remote)

Speaker: Jizhou Li (Stanford)

In this talk, I will present my recent efforts on the model-based machine learning with applications to two typical problems: 1) image restoration in fluorescence microscopy. The restoration process is parameterized as a linear combination of elementary functions and then optimized by minimizing a robust estimate of the true mean squared error through modeling the noise distribution. This way allows us to get the optimal results by simply solving a linear system of equations, without training through a large set of image pairs. 2) 3D particle segmentation in nano-CT images of lithium-ion battery cathodes. The shape of particles is embedded into the U-Net segmentation network to improve the performance, and a multi-view fusion strategy of 2D results is taken to reduce the annotation efforts and training uncertainty.

Containers! Containers! Containers!

Date: April 30th, 2020 1:00 pm (remote)

Speaker: Yee-ting Li

I'll do a walkthrough of container usage; from why to how and everything in between.

[Link](#) to the zoom recording



Containers and GPUs.pdf

PLAsTiCC: Convincing other people to solve your problems

Date: February 20th, 2020 3:00pm at Bldg 53 Rm 4002 (Toluca)

Speaker: Kara Ponder (Berkeley Center for Cosmological Physics)

The Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) tackled a major issue for the upcoming Rubin Observatory Legacy Survey of Space and Time (LSST): how to classify the hundreds of thousands of transients and variables that will be observed over 10 years. To prepare for this massive amount of data, the PLAsTiCC team produced the the largest public data set of synthetic astronomical light curves to date. The problem was presented to a broad data science community using the Kaggle platform and was able to engage more than 1000 teams in this photometric classification task. In this talk, I will show the steps we took towards building an effective data challenge that brought together the variable and transient science communities. I will briefly describe the winning methods, discuss the future of PLAsTiCC and demonstrate how its results are already influencing the astronomical community.



Ponder_AI_Seminar_at_SLAC.pdf

Real-time classification of explosive transients using deep recurrent neural networks

Date: April 27th, 2020 3:00pm at Bldg 53 Rm 4002 (Toluca)

Speaker: Daniel Muthukrishna

Astronomical transients are stellar objects that become temporarily brighter on various timescales and have led to some of the most significant discoveries in cosmology and astronomy. New and upcoming wide-field surveys such as the Zwicky Transient Facility (ZTF) and the Large Synoptic Survey Telescope (LSST) will record millions of multi-wavelength transient alerts each night. To meet this demand, we have developed a novel machine learning approach, RAPID (Real-time Automated Photometric Identification using Deep learning), that automatically classifies transients as a function of time. Using a deep recurrent neural network (RNN) with Gated Recurrent Units (GRUs), we are able to quickly classify multi-channel, sparse, time-series datasets into 12 different astrophysical types. The classification accuracy improves over the lifetime of the transient as more photometric data becomes available. In this talk, I will explain the main parts of our deep neural network architecture and describe our approach's classification performance on simulated and real data streams.

Shape learning

Date: February 13th, 2020 3:00pm at Bldg 53 Rm 4002 (Toluca)

Speaker: Nina Miolane (Stanford)

In medicine, the advances in bioimaging techniques have enabled us to access the 3D shapes of a variety of structures: organs, cells, proteins. Since biological shapes are related to physiological functions, biomedical research is poised to incorporate more shape data. In experimental physics, elementary particles traversing a detector leave "tracks" whose shapes are characteristics of their properties and interactions. Particle physics research also analyzes shape data to advance fundamental science. Therefore, two scientific fields ask the same machine learning question: how can we build quantitative descriptions of shapes and shape variabilities?



miolane_slac2020..._learning(1).pdf

GPU, Slurm and SLAC-Stanford-Scientific Data Facility

Speaker: Yee-Ting Lee (SLAC)

<https://confluence.slac.stanford.edu/display/SCSPub/Slurm+Batch>



SDF, GPUs and Slurm.pptx

Black Box Variational Inference: Scalable, Generic Bayesian Computation and its Applications

Date: August 12th 3:00pm at Bldg 53 Rm 1350 (Trinity)

Speaker: Rajesh Ranganath (NYU)

Probabilistic generative models are robust to noise, uncover unseen patterns, and make predictions about the future. Probabilistic generative models posit hidden structure to describe data. They have addressed problems in neuroscience, astrophysics, genetics, and medicine. The main computational challenge is computing the hidden structure given the data --- posterior inference. For most models of interest, computing the posterior distribution requires approximations like variational inference. Classically, variational inference was feasible to deploy in only a small fraction of models. We develop black box variational inference. Black box variational inference is a variational inference algorithm that is easy to deploy on a broad class of models and has already found use in neuroscience and healthcare. The ideas around black box variational inference also facilitate new kinds of variational methods such as hierarchical variational models. Hierarchical variational models improve the approximation quality of variational inference by building higher-fidelity approximations from coarser ones. Black box variational inference opens the doors to new models and better posterior approximations. Lastly, I will discuss some recent generic techniques in finding important features in predictive models like neural networks or random forests.

Pushing the Limits of Fluorescence Microscopy with adaptive imaging and machine learning

Date: September 5th 3:00pm at Bldg 53 Rm 4002 (Toluca)

Speaker: Dr. Loic A. Royer (Chan Zuckerberg Biohub)

Fluorescence microscopy lets biologist see and understand the intricate machinery at the heart of living systems and has led to numerous discoveries. Any technological progress towards improving image quality would extend the range of possible observations and would consequently open up the path to new findings. I will show how modern machine learning and smart robotic microscopes can push the boundaries of observability. One fundamental obstacle in microscopy takes the form of a trade-off between imaging speed, spatial resolution, light exposure, and imaging depth. We have shown that deep learning can circumvent these physical limitations: microscopy images can be restored even if 60-fold fewer photons are used during acquisition, isotropic resolution can be achieved even with a 10-fold under-sampling along the axial direction, and diffraction-limited structures can be resolved at 20-times higher frame-rates compared to state-of-the-art methods. Moreover, I will demonstrate how smart microscopy techniques can achieve the full optical resolution of light-sheet microscopes — instruments capable of capturing the entire developmental arch of an embryo from a single cell to a fully formed motile organism. Our instrument improves spatial resolution and signal strength two to five-fold, recovers cellular and sub-cellular structures in many regions otherwise not resolved, adapts to spatiotemporal dynamics of genetically encoded fluorescent markers and robustly optimises imaging performance during large-scale morphogenetic changes in living organisms.

Machine learning applications of quantum annealing in high energy physics

Date: August 22nd 3:00pm at Bldg 53 Rm 4002 (Toluca)

Speaker: Alexander Zlokapa (Caltech)

Due to their limitations, noisy intermediate-scale quantum (NISQ) devices often pose challenges in encoding real-world problems and in achieving sufficiently high fidelity computations. We present methodologies and results for overcoming these challenges on the D-Wave 2X quantum annealer for two problems in high energy physics: Higgs boson classification and charged particle tracking. Each problem is solved with a different construction, offering distinct perspectives on applications of quantum annealing. The quantum annealing for machine learning (QAML) algorithm ensembles weak classifiers to create a strong classifier from the excited states in the vicinity of the ground state, taking advantage of the noise that characterizes NISQ devices to help achieve comparable results to state-of-the-art classical machine learning methods in the Higgs signal-versus-background classification problem. Under a Hopfield network formulation, we also find successful results for charged particle tracking on simulated Large Hadron Collider data. Novel classical methods are proposed to overcome the limited size and connectivity of the D-Wave architecture, enabling the analysis of events with pileup at the scale of the Large Hadron Collider during its discovery of the Higgs boson. Furthermore, the time complexity of these classical pre-processing procedures is found to scale better with track density than current state-of-the-art tracking techniques, leaving open the possibility of a quantum speedup for tracking in the future.

NASA Ames Data Sciences Group Overview

Date: August 15th 3:00pm at Bldg 53 Rm 4002 (Toluca)

Speaker: Dr. Nikunj Oza (NASA Ames Research Center)

The Data Sciences Group (DSG) at NASA Ames Research Center performs research and development of machine learning and data mining methods for application to problems of interest to NASA, including Earth Science, Aeronautics, Space Science, and Human Space Exploration, as well as related problems involving work of interest to and funded by other organizations. This talk will give an overview of the Data Sciences Group's research and applications.

A Topology Layer for Machine Learning

Date: August 12th 3:00pm at Bldg 53 Rm 1350 (Trinity)

Speaker: Brad Nelson (Stanford/SLAC)

Topology applied to real world data using persistent homology has started to find applications within machine learning, including deep learning. We present a differentiable topology layer that computes persistent homology based on level set filtrations and distance-based filtrations. We present three novel applications: the topological layer can (i) serve as a regularizer directly on data or the weights of machine learning models, (ii) construct a loss on the output of a deep generative network to incorporate topological priors, and (iii) perform topological adversarial attacks on deep networks trained with persistence features. The code is publicly available and we hope its availability will facilitate the use of persistent homology in deep learning and other gradient based applications

Accelerating Data Science Workflows with RAPIDS

Date: July 24th 3:00pm at Bldg 53 Rm 1350

Speaker: Zahra Ronaghi (NVIDIA)

The RAPIDS suite of open source software libraries gives you the ability to accelerate and execute end-to-end data science workflows entirely on GPUs. RAPIDS relies on NVIDIA CUDA® primitives for low-level compute optimization, GPU parallelism, and high-bandwidth memory speed through user-friendly Python interfaces. Learn how to use the RAPIDS software stack from Python, including cuDF (a DataFrame library interoperable with Pandas), Dask-cuDF (for distributing DataFrame processing across GPUs) and cuML (GPU-accelerated versions of machine learning algorithm in Scikit-Learn)

Photometric classification of astronomical transients for LSST

Date: July 25th 3:00pm at Bldg 53 Rm 4002

Speaker: Kyle Boone (Berkeley)

Upcoming astronomical surveys such as the Large Synoptic Survey Telescope (LSST) will discover up to 10,000 new astronomical transients per night, and will need to use machine learning algorithms to classify them. Transient classification has several major challenges for machine learning algorithms, including sparse measurements with heteroskedastic noise, highly unbalanced classes, and unrepresentative training samples. To address these issues, I developed an algorithm that models the light curves of transients using Gaussian process regression, and then classifies them using gradient boosted decision trees. This model took first place out of over 1000 entries in the recent LSST PLAsTiCC photometric classification challenge that was hosted on the Kaggle platform. In this talk, I will provide an overview of the LSST PLAsTiCC challenge and dataset, and I will describe the algorithm that I developed for transient classification.

Analyzing and Applying Uncertainty in Deep Learning

Date: June 27th 3:00pm at Bldg 53 Rm 4002 (Zoom: <https://stanford.zoom.us/j/954139340>)

Speaker: Dustin Tran

Having reliable uncertainty estimates in neural networks is fundamental for a number of properties: from encouraging exploration outside the training distribution during active learning and control tasks; to having well-calibrated predictions; and to better understanding how to make actual decisions in healthcare given a model's patient-prediction. In this talk, we will primarily discuss noise contrastive priors (NCPs). The key idea is to train the model to

output high uncertainty for data points outside of the training distribution, improving its exploration properties. NCPs do so using an input prior, which adds noise to the inputs of the current minibatch, and an output prior, which is a wide distribution given these inputs. As time permits, we will also discuss recent efforts in measuring calibration for deep learning and also in analyzing the role of uncertainty for electronic health records.

Data-driven Discovery of the Governing Equations of Complex Physical Systems

Date: June 20th 3:00pm at Bldg 53 Rm 4002 (Zoom: <https://stanford.zoom.us/j/8036931498>)

Speaker: Paulo Alves

The increasing rate of production of scientific data, from both high-rep-rate experiments and large supercomputer simulations, is stimulating new opportunities in the way we do science. I will discuss how modern machine learning (ML) regression techniques can be exploited to uncover accurate and interpretable physical models directly from (high-fidelity simulation or experimental) data of complex physical systems. In particular, I will show how recent sparse learning methods can be used to discover partial differential equations (PDEs) that describe the spatio-temporal dynamics of the data in an interpretable form. The potential of these techniques will be demonstrated by extracting reduced physical models of plasma dynamics from high-fidelity Particle-in-Cell (PIC) simulations. Plasmas provide a complex and challenging test bed for data-driven discovery techniques due to their multi-scale and multivariate dynamics. I will demonstrate the recovery of the fundamental hierarchy of plasma physics equations, from the kinetic Vlasov equation to magnetohydrodynamics, based solely on spatial and temporal data of plasma dynamics from first-principles PIC simulations. The challenges associated with correlated plasma variables and the noise intrinsic to PIC simulation data will also be presented, and I will discuss strategies to overcome these issues for the robust recovery of the underlying plasma dynamics. I will conclude with an outlook on how such data-driven model discovery techniques can accelerate scientific research of complex physical systems.

Sparse Submanifold Convolution for Physics 2D/3D Image Analysis

Date: June 11th 3:00pm at Bldg 53 Rm 4002

Speaker: Laura Domine

Liquid Argon Time Projection Chambers (LArTPC) are a class of particle imaging detectors which records the trajectory of charged particles in either 2D or 3D imaging data with a breathtaking resolution (~3mm/pixel). Convolutional Neural Networks (CNNs), a powerful deep learning technique to extract physics features from images, were successfully applied to the data reconstruction and analysis of LArTPC. Yet a unique feature of LArTPC data challenges traditional CNN algorithms: it is locally dense (no gap in a particle trajectory) but generally sparse. A typical 2D and 3D LArTPC image has less than 1% and 0.1% of pixels occupied with non-zero value respectively. This makes standard CNNs with dense matrix operations very inefficient. Submanifold sparse convolutional networks (SSCN) have been proposed to address exactly this class of sparsity challenges by keeping the same level of sparsity throughout the network. We demonstrate their strong performance on some of our data reconstruction tasks which include 3D semantic segmentation for particle identification at the pixel-level. SSCN can address the problem of computing resource scalability for 3D deep learning-based data reconstruction chain R&D for LArTPC detectors.

Applying Convolutional Neural Networks to MicroBooNE

Date: June 6th 3:00pm at Bldg 53 Rm 4002

Speaker: Taritree Wongjirad

The MicroBooNE experiment consists of liquid argon time projection chamber(LArTPC) situated in the path of the Booster Neutrino Beam (BNB) at Fermilab. The goals of the experiment are to (1) investigate the observation of an excess of a possible electron-neutrino and anti-neutrino events by the MiniBooNE experiment, (2) measure argon-nucleus cross sections, and (3) perform R&D for LArTPCs. The data from MicroBooNE, and other LArTPCs, can be naturally arranged as high-resolution images of particle tracks traversing the detector. This has spurred effort on MicroBooNE towards applying convolutional neural networks (CNNs), a type of deep learning algorithm shown to be extremely effective in numerous computer vision problems, to our data. I'll talk about the ways in which MicroBooNE uses CNNs with a focus on recent results demonstrating their performance on real data. I'll also discuss future directions MicroBooNE is exploring to further apply CNNs.



2019_06_06_SL...Sem_UBCNN.pdf

Machine Learning, Datascience and Neutrino Physics at Argonne's Leadership Computing Facility

Date: February 28th 3:00pm at Bldg 53 Rm 4002

Speaker: Corey Adams

Abstract: As machine learning and deep learning have become important tools for science, both in analysis and simulation of data, high performance computing workflows have expanded to target learning applications. The Leadership Computing Facility at Argonne National Laboratory (ALCF) will be the home to the first US Exascale computer in 2021, Aurora, and the datascience group at ALCF works to scale applications in simulation and learning to

current and future systems across broad scientific domains. In this talk, I will describe the ALCF center and its suite of datascience projects, and as a neutrino physicist I will describe the neutrinos-on-HPC deep learning applications at ALCF. In particular, I will cover many tools and best practices for getting good performance of machine learning workflows on HPC systems.

[SLAC_ML_AI_Seminar_CAdams.pdf](#)

Deep Learning for Particle Track Finding in High Energy Physics

Date: February 21st 3:00pm at Bldg 53 Rm 4002

Speaker: Steve Farrell

Abstract: Particle track reconstruction is a challenging but critical component of data analysis in high energy physics experiments such as those at the Large Hadron Collider. In the high-luminosity LHC era, experiments will need to identify $O(10k)$ particle tracks in every proton-proton collision event using detectors with $O(100M)$ readout channels. Today's solutions use Kalman Filters and combinatorial searches which are inherently serial and are expected to consume a substantial amount of compute resources at the HL-LHC. Meanwhile, deep learning methodologies are enjoying increasing success in many domains in industry and science. These approaches provide the means to learn complex representations of high-dimensional data for solving complex tasks while also efficiently using modern parallel hardware devices such as GPUs. The HEP.TrkX project has therefore been exploring deep learning solutions using convolutional, recurrent, and graph neural networks for image-like and spacepoint representations of tracking detector data. In this presentation I will describe the approaches that have been studied in the HEP.TrkX project, including an in-depth focus on a graph neural network (GNN) model which operates on a graph-structured representation of tracking detector data. This GNN model is able to uncover particle trajectories in entire collision events by resolving pairwise associations between position measurements in the detector and has shown promising results on simulated data from the Kaggle tracking machine learning challenge. I will also discuss ongoing and future directions for this line of research.



Farrell_trackml_slac.pdf

Machine Learning for medical applications of Physics

Date: January 16th at 12:30pm at Bldg 53 Rm 4002

Speaker: Carlo Mancini (INFN, Rome)

Abstract:

Deep Neural Networks (DNNs) techniques are applied to a vast number of cases, such as human face recognition, image segmentation, self-driving cars, and even playing Go. In this talk, I present our first steps in using DNNs in medical applications, i.e.: to segment Magnetic Resonance (MR) images and to reproduce the final state of a low energy nuclear interaction model, BLOB (Boltzmann Langevin One Body). The first application tries to give an answer to the necessity, expressed by clinicians, of identifying rectal cancer patients who do not need radical surgery after the chemo-radiotherapy prescribed by the clinical protocol. The second one aims at exploring the possibility of using a Variational Auto Encoder (VAE) to simulate accurately low energy nuclear interactions in order to reduce the computation time with respect to the full model. Once trained, the VAE could be used in Monte Carlo simulation of patients' treatments with ion beams.

Machine Learning synthetic data, scanning probe data, and reciprocal space data on quantum materials

Date: October 19 at 1pm (note change in time!)

Speaker: Eun-Ah Kim (Cornell)

Abstract: The scientific questions in the field of electronic quantum matter require fundamentally new approaches to data science for two reasons: (1) quantum mechanical imaging of electronic behavior is probabilistic, (2) inference from data should be subject to fundamental laws governing microscopic interactions. In this talk, I will discuss our approaches to synthetic data, scanning probe data and comprehensive X-ray data.



2018 SLAC-to-share.pdf

Local-to-Global Methods for Topological Data Analysis

Date: October 1, 2018 at 3pm

Speaker: Brad Nelson

Abstract: Topological data analysis (TDA) seeks to understand and utilize the shape of data. An important task in TDA is the design and computation of topological invariants of sampled data, which can be transformed into features for machine learning, or used to inform an exploratory modeling process. This talk will begin with an introduction to the core methods of TDA and some scientific applications. We will then present our recent work on the use of hierarchical models to investigate larger and more complex data sets. Video of presentation: https://stanford.zoom.us/recording/share/E0gJd-JheyJin0gCRNOeWsVWLRbAqqpJjOL_pNM4mnGwlumekTziMw?startTime=1538431327000



slides.pdf

Experience with a Virtual Multi-Slit Phase Space Diagnostic at Fermilab's FAST Facility

Date: August 27, 2018 at 3pm

Speaker: Auralee Edelen

Abstract: We discuss an ML-based virtual diagnostic for multi-slit beam measurements at the Fermilab Accelerator Science and Technology Facility (FAST). At FAST, a multi-slit diagnostic setup is used to obtain information about the transverse electron beam phase space. In addition to being a destructive measurement, the process of conducting this measurement at multiple points along the machine (e.g. before and after the chicane) is time-consuming. Here, we use a learned representation of the low-energy portion of the machine to provide estimates of the multi-slit measurement after the second superconducting capture cavity. The result is a fast-executing tool that uses present machine settings to give an online prediction of the resultant multi-slit measurement without the need to repeatedly insert the intercepting diagnostic. This work is particularly relevant to ongoing efforts at FACET-II and LCLS/LCLS-II to create similar kinds of virtual diagnostic tools for prediction and, ultimately, control of the longitudinal phase space.



20180826_Edelen_SeminarAI.pdf

A Novel Approach - IoT Device Virtualization using ML

Date: July 19, 2018 at 11am (Note special time!)

Speaker: Knowthings

Abstract: Developing and testing IoT solutions is challenging today due to many factors: complexity around available platforms, heterogeneous environments, voluminous device requirements, multiple protocols, and communication synchronization. In this session we'll find out how IoT Device Virtualization can help expedite development and testing of IoT applications, and how machine learning can play a very important role in that. We'll see how sequence alignment techniques and data mining methods, like Needleman Wunsch algorithm, enable the means of learning device behavior, and generate a data model out of it. This model can then be played back at runtime to behave like an actual IoT device.

Sparsity/Undersampling Tradeoffs in Compressed Sensing

Date: July 9, 2018 at 3pm

Speaker: Hatef Monajemi (Stanford)

Abstract: Compressed sensing (CS) is a sampling technique that exploits sparsity to speed up acquisition. An important question in CS is "how much undersampling is allowed for a given level of sparsity?" This question has been answered when sampling is done using Gaussian random matrices. Unfortunately, the theories for Gaussian matrices are not directly applicable to certain real life applications such as magnetic resonance imaging /spectroscopy where unique experimental considerations may impose extra constraints on the sampling matrix. In this talk, we will review the literature on sparsity/undersampling tradeoffs for Gaussian matrices and then present new predictions that are applicable to MR spectroscopy/imaging.

Learned predictive models: integrating large-scale simulations and experiments using deep learning

Date: June 25, 2018 at 3pm

Speaker: Brian Spears (LLNL)

Abstract: Across scientific missions, we regularly need to develop accurate models that closely predict experimental observation. Our team is developing a new class of model, called the learned predictive model, that captures theory-driven simulation, but also improves by exposure to experimental observation. We begin by designing specialized deep neural networks that can learn the behavior of complicated simulation models from exceptionally large simulation databases. Later, we improve, or elevate, the trained models by incorporating experimental data using a technique called transfer learning. The training and elevation process improves our predictive accuracy, provides a quantitative measure of uncertainty, and helps us cope with limited experimental data volumes. To drive this procedure, we have also developed a complex computational workflow that can generate hundreds of thousands to billions of simulated training examples and can steer the subsequent training and elevation process. These workflow tasks require a heterogeneous high-performance computing environment supporting computation on CPUs, GPUs, and sometimes specialized, low-precision processors. We will present a global view of our deep learning efforts, our computational workflows, and some implications that this computational work has for current and future large-scale computing platforms.

Rapid Gaussian Process Training via Structured Low-Rank Kernel Approximation of Gridded Measurements

Date: June 4, 2018 at 3pm

Speaker: Franklin Fuller

Abstract: The cubic scaling of matrix inversion with the number of data points is the main computational cost in Gaussian Process (GP) regression. Sparse GP approaches reduce the complexity of matrix inversion to linear complexity by making an optimized low rank approximation to the kernel, but the quality of the approximation depends (and scales with) how many "inducing" or representative points are allowed. When the problem at hand allows the kernel to be decomposed into a kronecker product of lower dimensional kernels, many more inducing points can be feasibly processed by exploiting the kronecker factorization, resulting in a much higher quality fit. Kronecker factorizations suffer from exponentially scaling with the dimension of the input, however, which has limited this approach to problems of only a few input dimensions. It was recently shown how this problem can be circumvented by making an additional low-rank approximation across input dimensions, resulting in an approach that scales linearly in both data points and the input dimensionality. We explore a special case of this recent work wherein the observed data are measured on a complete multi-dimensional grid (not necessarily uniformly spaced), which is a very common scenario in scientific measurement environments. In this special case, the problem decomposes over axes of the input grid, making the cost linearly scale mainly with the largest axis of the grid. We apply this approach to deconvolve linearly mixed spectroscopic signals and are able to optimize kernel hyper parameters on datasets containing billions of measurements in minutes with a laptop.



June42018_Struc...edMatrices.pptx

Machine learning applications for hospitals

Date: May 21, 2018 at 3pm

Speaker: David Scheinker

Abstract: Academic hospitals and particle accelerators have a lot in common. Both are complex organizations; employ numerous staff and scientists; deliver a variety of services; research how to improve the delivery of those services; and do it all with a variety of large expensive machines. My group focuses on helping the Stanford hospitals, mostly the Children's Hospital, seek to improve: throughput, decision-support, resource management, innovation, and education. I'll present brief overviews of a variety of ML-based approaches to projects in each of these areas. For example, integer programming to optimize surgical scheduling and Neural Networks to interpret continuous-time waveform monitor data. I will conclude with a broader vision for how modern analytics methodology could potentially transform healthcare delivery. More information on the projects to be discussed is available at surf.stanford.edu/projects

Beyond Data and Model Parallelism for Deep Neural Networks

Date: May 7, 2018 at 3pm

Speaker: Zhihao Jia

Abstract: Existing deep learning systems parallelize the training process of deep neural networks (DNNs) by using simple strategies, such as data and model parallelism, which usually results in suboptimal parallelization performance for large scale training. In this talk, I will first formalize the space of all possible parallelization strategies for training DNNs. After that, I will present FlexFlow, a deep learning framework that automatically finds efficient parallelization strategies by using a guided random search algorithm to explore the space of all possible parallelization strategies. Finally, I will show that FlexFlow significantly outperforms state-of-the-art parallelization approaches by increasing training throughput, reducing communication costs, and achieving improved scalability.

X-ray spectrometer data processing with unsupervised clustering (Sideband signal seeking)

Date: April 9, 2018 at 3pm

Speaker: Guanqun Zhou

Abstract: Online spectrometer plays an important role in the characterization of the free-electron laser (FEL) pulse spectrum. With the help of beam synchronization acquisition (BSA) system, the spectrum of independent shot can be stored, which helps the downstream scientific researchers a lot. However, because of spontaneous radiation, FEL intrinsic fluctuations and other stochastic effects, the data from spectrometer cannot be fully utilized. A specific case is sideband signal resolution in hard-xray self-seeding experiment. During the seminar, I will present my exploration of employing unsupervised clustering algorithm to mine the latent information in the spectrometer data. In this way, sideband signal starts to appear.

Experience with FEL taper tuning using reinforcement learning and clustering

Date: April 2, 2018 at 3pm

Speaker: Juhao Wu

Abstract: LCLS, world's first hard X-ray Free Electron Laser (FEL) is serving multiple users. It commonly happens that different scientific research requires very different parameters of the X-Ray pulses, therefore setting up the system in a timing fashion meeting these requests is a nontrivial task. Artificial intelligence is not only very helpful to conduct well defined task towards definitive goal, it also helps to find new operating regime generating unexpected great results. Here in this talk, we will report experience with FEL taper tuning using reinforcement learning and clustering. Such study opens up novel taper configuration such as a zig-zag taper which takes full advantages of the filamentation of the electron bunch phase space in the deep saturated regime.



ASLAC_TaperRLCluster_Sub.pptx

Statistical Learning of Reduced Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics

Date: Feb. 26, 2018 at 3pm

Speaker: Qian Yang (Stanford)

Complex chemical processes, such as the decomposition of energetic materials and the chemistry of planetary interiors, are typically studied using large-scale molecular dynamics simulations that can run for weeks on high performance parallel machines. These computations may involve thousands of atoms forming hundreds of molecular species and undergoing thousands of reactions. It is natural to wonder whether this wealth of data can be utilized to build more efficient, interpretable, and predictive models of complex chemistry. In this talk, we will use techniques from statistical learning to develop a framework for constructing Kinetic Monte Carlo (KMC) models from molecular dynamics data. We will show that our KMC models can not only extrapolate the behavior of the chemical system by as much as an order of magnitude in time, but can also be used to study the dynamics of entirely different chemical trajectories with a high degree of fidelity. Then, we will discuss a new and efficient data-driven method using L1-regularization for automatically reducing our learned KMC models from thousands of reactions to a smaller subset that effectively reproduces the dynamics of interest.



SLAC_AI_slides_redacted.pdf

Machine Learning for Jet Physics at the Large Hadron Collider

Date: February 12, 2018 at 3pm

Speaker: Ben Nachman (CERN)

Abstract: Modern machine learning (ML) has introduced a new and powerful toolkit to High Energy Physics. While only a small number of these techniques are currently used in practice, research and development centered around modern ML has exploded over the last year(s). I will highlight recent advances with a focus on jets: collimated sprays of particles resulting from quarks and gluons produced at high energy. Themselves defined by unsupervised learning algorithms, jets are a prime benchmark for state-of-the-art ML applications and innovations. For example, I will show how deep learning has been applied to jets for classification, regression, and generation. These tools hold immense potential, but incorporating domain-specific knowledge is necessary for optimal performance. In addition, studying what the machines are learning is critical for robustness and may even help us learn new physics!



Nachman_AI@SLAC.pdf

Tutorial – Implementation Practice of Deep Neural Network Technique into Our Research

Date: Wednesday Jan. 24, 2018 at 3pm-5pm in Mammoth B53-3036 (Note time and place!)

Speaker: Kazuhiro Terao

Abstract: The progress of machine learning techniques in recent years has been impactful in many fields of research including physics science. While learning more about machine learning subject exciting, an implementation in our research can face a huge learning curve and take time. In this first AI seminar of 2018, I will show our implementation of convolution neural network for analyzing liquid argon time projection chamber (LArTPC) detector data. In particular, we will look at an instance of convolutional neural network trained for semantic-segmentation task (i.e. pixel-level object classification task). This seminar will be in an interactive tutorial style using Jupyter python notebook. In the first part of the seminar, I will give an overview of software stacks and workflow. Then I will demonstrate training the algorithm. Both the software and data will be made available in advance, and the audience is welcome to participate in training the algorithm (which may require an NVIDIA GPU enabled Linux interactive shell such as SLAC computing or AWS service).



kazu_tutorial.pdf

In situ visualization with task-based parallelism

Date: Nov. 27, 2017 at 3pm

Speaker: Alan Heirich

Abstract: This short paper describes an experimental prototype of in situ visualization in a task-based parallel programming framework. A set of reusable visualization tasks were composed with an existing simulation. The visualization tasks include a local OpenGL renderer, a parallel image compositor, and a display task. These tasks were added to an existing fluid-particle-radiation simulation and weak scaling tests were run on up to 512 nodes of the Piz Daint supercomputer. Benchmarks showed that the visualization components scaled and did not reduce the simulation throughput. The compositor latency increased logarithmically with increasing node count.

Data Reconstruction Using Deep Neural Networks for Liquid Argon Time Projection Chamber Detectors

Date: Oct. 16, 2017 at 3pm

Speaker: Kazuhiro Terao

Deep neural networks (DNNs) have found a vast number of applications ranging from automated human face recognition, real-time object detection for self-driving cars, teaching a robot Chinese, and even playing Go. In this talk, I present our first steps in exploring the use of DNNs to the task of analyzing neutrino events coming from Liquid Argon Time Projection Chambers (LArTPC), in particular the MicroBooNE detector. LArTPCs consist of a large volume of liquid argon sandwiched between a cathode and anode wire planes. These detectors are capable of recording images of charged particle tracks with breathtaking resolution. Such detailed information will allow LArTPCs to perform accurate particle identification and calorimetry, making it the detector of choice for many current and future neutrino experiments. However, analyzing such images can be challenging, requiring the development of many algorithms to identify and assemble features of the events in order to identify and remove cosmic-ray-induced particles and reconstruct neutrino interactions. This talk shows the current status of DNN applications and our future direction.



Kazu_2017_10_16_AI@SLAC.pdf

Towards a cosmology emulator using Generative Adversarial Networks

Date: Oct 3, 2017 at 2pm

Speaker: Mustafa Mustafa

The application of deep learning techniques to generative modeling is renewing interest in using high dimensional density estimators as computationally inexpensive emulators of fully-fledged simulations. These generative models have the potential to make a dramatic shift in the field of scientific simulations, but for that shift to happen we need to study the performance of such generators in the precision regime needed for science applications. To this end, in this talk we apply Generative Adversarial Networks to the problem of generating cosmological weak lensing convergence maps. We show that our generator network produces maps that are described by, with high statistical confidence, the same summary statistics as the fully simulated maps.



2017-10-03.pdf

Optimal Segmentation with Pruned Dynamic Programming

Date: Sept. 12, 2017 at 2pm

Speaker: Jeffrey Scargle (NASA)

Bayesian Blocks (1207.5578) is an $O(N^2)$ dynamic programming algorithm to compute exact global optimal segmentations of sequential data of arbitrary mode and dimensionality. Multivariate data, generalized block shapes, and higher dimensional data are easily treated. Incorporating a simple pruning method yields a (still exact) $O(N)$ algorithm allowing fast analysis of series of $\sim 100M$ data points. Sample applications include analysis of X- and gamma-ray time series, identification of GC-islands in the human genome, data-adaptive triggers and histograms, and elucidating the Cosmic Web from 3D galaxy redshift data.



scargle_optimal...egmentation.pdf

Fast automated analysis of strong gravitational lenses with convolutional neural networks

Date: Sept. 12, 2017 at 2pm

Speaker: Yashar Hezaveh

Strong gravitational lensing is a phenomenon in which the image of a distant galaxy appears highly distorted due to the deflection of its light rays by the gravity of a more nearby, intervening galaxy. We often see multiple distinct arc-shaped images of the background galaxy around the intervening (lens) galaxy, just like images in a funhouse mirror. Strong lensing gives astrophysicist a unique opportunity to carry out different investigations, including mapping the detailed distribution of dark matter, or measuring the expansion rate of the universe. All these great sciences, however, require a detailed knowledge of the distribution of matter in the lensing galaxies, measured from the distortions in the images. This has been traditionally performed with maximum-likelihood lens modeling, a procedure in which simulated observations are generated and compared to the data in a statistical way. The parameters controlling the simulations are then explored with samplers like MCMC. This is a time and resource consuming procedure, requiring hundreds of hours of computer and human time for a single system. In this talk, I will discuss our recent work in which we showed that deep convolutional neural networks can solve this problem more than 10 million times faster: about 0.01 seconds per system on a single GPU. I will also review our method for quantifying the uncertainties of the parameters obtained with these networks. With the advent of upcoming sky surveys such as the Large Synoptic Survey Telescope, we are anticipating the discovery of tens of thousands of new gravitational lenses. Neural networks can be an essential tool for the analysis of such high volumes of data.

MacroBase: A Search Engine for Fast Data Streams

Date: Sept. 5, 2017 at 2pm

Speaker: Sahaana Suri (Stanford)

While data volumes generated by sensors, automated process, and application telemetry continue to rise, the capacity of human attention remains limited. To harness the potential of these large scale data streams, machines must step in by processing, aggregating, and contextualizing significant behaviors within these data streams. This talk will describe progress towards achieving this goal via MacroBase, a new analytics engine for prioritizing attention in this large-scale "fast data" that has begun to deliver results in several production environments. Key to this progress are new methods for constructing cascades of analytic operators for classification, aggregation, and high-dimensional feature selection; when combined, these cascades yield new opportunities for dramatic scalability improvements via end-to-end optimization for streams spanning time-series, video, and structured data. MacroBase is a core component of the Stanford DAWN project (<http://dawn.cs.stanford.edu/>), a new research initiative designed to enable more usable and efficient machine learning infrastructure.



macrobase-SLAC_orig.pptx

Object-Centric Machine Learning

Date: Aug. 29, 2017 at 2pm

Speaker: Leo Guibas (Stanford)

Deep knowledge of the world is necessary if we are to have autonomous and intelligent agents and artifacts that can assist us in everyday activities, or even carry out tasks entirely independently. One way to factorize the complexity of the world is to associate information and knowledge with stable entities, animate or inanimate, such as persons or vehicles, etc -- what we generally refer to as "objects."

In this talk I'll survey a number of recent efforts whose aim is to create and annotate reference representations for (inanimate) objects based on 3D models with the aim of delivering such information to new observations, as needed. In this object-centric view, the goal is to learn about object geometry, appearance, articulation, materials, physical properties, affordances, and functionality. We acquire such information in a multitude of ways, both from crowd-sourcing and from establishing direct links between models and signals, such as images, videos, and 3D scans -- and through these to language and text. The purity of the 3D representation allows us to establish robust maps and correspondences for transferring information among the 3D models themselves -- making our current 3D repository, ShapeNet, a true network.

While neural network architectures have had tremendous impact in image understanding and language processing, their adaptation to 3D data is not entirely straightforward. The talk will also briefly discuss current approaches in designing deep nets appropriate for operating directly on irregular 3D data representations, such as meshes or point clouds, both for analysis and synthesis -- as well as ways to learn object function from observing multiple action sequences involving objects -- in support of the above program.

Reconstruction Algorithms for Next-Generation Imaging: Multi-Tiered Iterative Phasing for Fluctuation X-ray Scattering and Single-Particle Diffraction

Date: Aug. 15, 2017 at 2pm

Location: Tulare (B53-4006) (NOTE CHANGE IN ROOM!)

Speaker: Jeffrey Donatelli (CAMERA, Berkeley)

Abstract: The development of X-ray free-electron lasers has enabled new experiments to study uncrystallized biomolecules that were previously infeasible with traditional X-ray sources. One such emerging experimental technique is fluctuation X-ray scattering (FXS), where one collects a series of diffraction patterns, each from multiple particles in solution, using ultrashort X-ray pulses that allow snapshots to be taken below rotational diffusion times of the particles. The resulting images contain angularly varying information from which angular correlations can be computed, yielding several orders of magnitude more information than traditional solution scattering methods. However, determining molecular structure from FXS data introduces several challenges, since, in addition to the classical phase problem, one must also solve a hyper-phase problem to determine the 3D intensity function from the correlation data. In another technique known as single-particle diffraction (SPD), several diffraction patterns are collected, each from an individual particle. However, the samples are delivered to the beam at unknown orientations and may also be present in several different conformational states. In order to reconstruct structural information from SPD, one must determine the orientation and state for each image, extract an accurate 3D model of the intensity function from the images, and solve for the missing complex phases, which are not measured in diffraction images.

In this talk, we present the multi-tiered iterative phasing (M-TIP) algorithm for determining molecular structure from both FXS and SPD data. This algorithm breaks up the associated reconstruction problems into a set of simpler subproblems that can be efficiently solved by applying a series of projection operators. These operators are combined in a modular iterative framework which is able to simultaneously determine missing parameters, the 3D intensity function, the complex phases, and the underlying structure from the data. In particular, this approach is able to leverage prior knowledge about the structural model, such as shape or symmetry, to obtain a reconstruction from very limited data with excellent global convergence properties and high computational efficiency. We show results from applying M-TIP to determine molecular structure from both simulated data and experimental data collected at the Linac Coherent Light Source (LCLS).



Donatelli_Phasing.pdf

Exploratory Studies in Neural Network-based Modeling and Control of Particle Accelerators

Date: Aug 1, 2017 at 2pm

Speaker: Auralee Edelen (CSU)

Particle accelerators are host to myriad control challenges: they involve a multitude of interacting systems, are often subject to tight performance demands, in many cases exhibit nonlinear behavior, sometimes are not well-characterized due to practical and/or fundamental limitations, and should be able to run for extended periods of time with minimal interruption. One avenue toward improving the way these systems are controlled is to incorporate techniques from machine learning. Within machine learning, neural networks in particular are appealing because they are highly flexible, they are well-suited to problems with nonlinear behavior and large parameter spaces, and their recent success in other fields (driven largely by algorithmic advances, greater availability of large data sets, and improvements in high performance computing resources) is an encouraging indicator that they are now technologically mature enough to be fruitfully applied to particle accelerators. This talk will highlight a few recent efforts in this area that were focused on exploring neural network-based approaches for modeling and control of several particle accelerator subsystems, both through simulation and experimental studies.



Aug1_20170801_...LAC_slides.pdf

Estimating behind-the-meter solar generation with existing measurement infrastructure

Date: July 11, 2017 at 2pm

Speaker: Emre Kara

Real-time PV generation information is crucial for distribution system operations such as switching, state-estimation, and voltage management. However, most behind-the-meter solar installations are not monitored. Typically, the only information available to the distribution system operator is the installed capacity of solar behind each meter; though in many cases even the presence of solar may be unknown. We present a method for disaggregating behind-the-meter solar generation using only information that is already available in most distribution systems. Specifically, we present a contextually supervised source separation strategy adopted to address the behind-the-meter solar disaggregation problem. We evaluate the model sensitivities to different input parameters such as the number of solar proxy measurements, number of days in the training set, and region size.



EmreKara_Estima...rastructure.pdf

Development and Application of Online Optimization Algorithms

Date: June 27, 2017 at 3pm

Location: Kings River, B52-306 (Note change in time and place!)

Speaker: Xiabiao Huang

Automated tuning is an online optimization process. It can be faster and more efficient than manual tuning and can lead to better performance. It may also substitute or improve upon model based methods. Noise tolerance is a fundamental challenge to online optimization algorithms. We discuss our experience in developing a high efficiency, noise-tolerant optimization algorithm, the RCDS method, and the successful application of the algorithm to various real-life accelerator problems. Experience with a few other online optimization algorithms are also discussed.



XiaobiaoHuang_RCDS.pdf

Machine Learning at NERSC: Past, Present, and Future

Date: May 16, 2017 at 2pm

Speaker: Prabhat (NERSC)

Modern scientific discovery increasingly relies upon analysis of experimental and observational data. Instruments across a broad range of spatial scales: telescopes, satellites, drones, genome sequencers, microscopes, particle accelerators, gather increasingly large and complex datasets. In order to 'infer' properties of nature, in light of noisy, incomplete measurements, scientists need access to sophisticated statistics and machine learning tools. In order to address these emerging challenges, NERSC has deployed a portfolio of Big Data technologies on HPC platforms. This talk will review the evolution of Data Analytics tools (statistics, machine learning/deep learning) in the recent past, comment on current scientific use cases and challenges, and speculate on the future of AI-powered scientific discovery.

Optimization for Transportation Efficiency

Date: May 2, 2017 at 2pm

Location: Sycamore Conference Room (040-195)

Speaker: John Fox

Abstract: Plug-in hybrid and all-electric vehicles offer potential to transfer energy demands from liquid petroleum fuels to grid-sourced electricity. We are investigating optimization methods to improve the efficiency and resource utilization of Plug-in Hybrid Electric Vehicles (HEVs). Our optimization uses information about a known or estimated vehicle route to predict energy demands and optimally manage on-board battery and fuel energy resources to maximally use grid-sourced electricity and minimally use petroleum resources for a given route. Our convex optimization method uses a simplified car model to find the optimal strategy over the whole route, which allows for re-optimization on the fly as updated route information becomes available. Validation between the simplified model and a more complete vehicle technology model simulation developed at Argonne National Laboratory was accomplished by "driving" the complete car simulation with the simplified control model. By driving on routes with the same total energy demand but different demand profiles we show fuel efficiency gains of 5-15% on mixed urban/suburban routes compared to a Charge Depleting Charge Sustaining (CDCS) battery controller. The method also allows optimizing the economic lifetime of the vehicle battery by considering the stress on the battery from charge and discharge cycles in the resource optimization.



jfoxhybrid.pdf

Detecting Simultaneous Changepoints Across Multiple Data Sequences

Date: April 25, 2017 at 3pm

Location: Kings River, 052-306 (NOTE DIFFERENT LOCATION)

Speaker: Zhou Fan

Abstract: Motivated by applications in genomics, finance, and biomolecular simulation, we introduce a Bayesian model called BASIC for changepoints that tend to co-occur across multiple related data sequences. We design efficient algorithms to infer changepoint locations by sampling from and maximizing over the posterior changepoint distribution. We further develop a Monte Carlo expectation-maximization procedure for estimating unknown prior hyperparameters from data. The resulting framework accommodates a broad range of data and changepoint types, including real-valued sequences with changing mean or variance and sequences of counts or binary observations. We use the resulting BASIC framework to analyze DNA copy number variations in the NCI-60 cancer cell lines and to identify important events that affected the price volatility of S&P 500 stocks from 2000 to 2009.



ZhouFan_BASIC.pdf

Low Data Drug Discovery with One-Shot Learning

Date: April 18, 2017 at 2pm

Speaker: Bharath Ramsundar

Location: Berryessa Conference Room (B53-2002) **(NOTE DIFFERENT ROOM!)**

Abstract: Recent advances in machine learning have made significant contributions to drug discovery. Deep neural networks in particular have been demonstrated to provide significant boosts in predictive power when inferring the properties and activities of small-molecule compounds. However, the applicability of deep learning has been limited by the requirement for large amounts of training data. In this work, we demonstrate how one-shot learning can be used to significantly lower the amounts of data required to make meaningful predictions in drug discovery applications. We introduce a new architecture, the iterative refinement long short-term memory, that, when combined with graph convolutional neural networks, significantly improves learning of meaningful distance metrics over small-molecules. Our models are open-sourced as part of DeepChem, an open framework for deep-learning in drug discovery and quantum chemistry.

Bio: Bharath Ramsundar received a BA and BS from UC Berkeley in EECS and Mathematics and was valedictorian of his graduating class in mathematics. He is currently a PhD student in computer science at Stanford University with the Pande group. His research focuses on the application of deep-learning to drug-discovery. In particular, Bharath is the creator and lead-developer of DeepChem, an open source package that aims to democratize the use of deep-learning in drug-discovery and quantum chemistry. He is supported by a Hertz Fellowship, the most selective graduate fellowship in the sciences.

Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis

Date: Mar. 28, 2017 at 2pm

Speaker: Michela Paganini

Location: Berryessa Conference Room (B53-2002)

Abstract: We provide a bridge between generative modeling in the Machine Learning community and simulated physical processes in High Energy Particle Physics by applying a novel Generative Adversarial Network (GAN) architecture to the production of jet images -- 2D representations of energy depositions from particles interacting with a calorimeter. We propose a simple architecture, the Location-Aware Generative Adversarial Network, that learns to produce realistic radiation patterns from simulated high energy particle collisions. The pixel intensities of GAN-generated images faithfully span over many orders of magnitude and exhibit the desired low-dimensional physical properties (i.e., jet mass, n-subjettiness, etc.). We shed light on limitations, and provide a novel empirical validation of image quality and validity of GAN-produced simulations of the natural world. This work provides a base for further explorations of GANs for use in faster simulation in High Energy Particle Physics.



gan_presentation_SLAC.pdf

Models and Algorithms for Solving Sequential Decision Problems under Uncertainty

Date: Mar. 21, 2017 at 2pm

Speaker: Mykel Kochenderfer

Location: Sycamore Conference Room (040-195)

Abstract: Many important problems involve decision making under uncertainty, including aircraft collision avoidance, wildfire management, and disaster response. When designing automated decision support systems, it is important to account for the various sources of uncertainty when making or recommending decisions. Accounting for these sources of uncertainty and carefully balancing the multiple objectives of the system can be very challenging. One way to model such problems is as a partially observable Markov decision process (POMDP). Recent advances in algorithms, memory capacity, and processing power, have allowed us to solve POMDPs for real-world problems. This talk will discuss models for sequential decision making and algorithms for solving them.

Data Programming: A New Framework for Weakly Supervising Machine Learning Models

Date: Mar. 7, 2017 at 2pm

Speaker: Alex Ratner

Location: Sycamore Conference Room (040-195)

Abstract: Today's state-of-the-art machine learning models require *massive* labeled training sets--which usually do not exist for real-world applications. Instead, I'll discuss a newly proposed machine learning paradigm--[data programming](#)--and a system built around it, [Snorkel](#), in which the developer focuses on writing a set of *labeling functions*, which are just scripts that programmatically label data. The resulting labels are noisy, but we model this as a generative process--learning, essentially, which labeling functions are more accurate than others--and then use this to train an end discriminative model (for example, a deep neural network in TensorFlow). Given certain conditions, we show that this method has the same asymptotic scaling with respect to generalization error as directly-supervised approaches. Empirically, we find that by modeling a noisy training set creation process in this way, we can take potentially low-quality labeling functions from the user, and use these to train high-quality end models. We see this as providing a general framework for many [weak supervision](#) techniques, and at a higher level, as defining a new programming model for weakly-supervised machine learning systems.



AlexRatner_SLAC...ding_share.pptx

ProxImaL: Efficient Image Optimization using Proximal Algorithms

Date: Feb. 28, 2017 at 2pm

Speaker: Felix Heide

Location: Truckee Room (B52-206) **(NOTE ROOM CHANGE!)**

Abstract: Computational photography systems are becoming increasingly diverse while computational resources, for example on mobile platforms, are rapidly increasing. As diverse as these camera systems may be, slightly different variants of the underlying image processing tasks, such as demosaicking, deconvolution, denoising, inpainting, image fusion, and alignment, are shared between all of these systems. Formal optimization methods have recently been demonstrated to achieve state-of-the-art quality for many of these applications. Unfortunately, different combinations of natural image priors and optimization algorithms may be optimal for different problems, and implementing and testing each combination is currently a time consuming and error prone process.

ProxImaL is a domain-specific language and compiler for image optimization problems that makes it easy to experiment with different problem formulations and algorithm choices. The language uses proximal operators as the fundamental building blocks of a variety of linear and nonlinear image formation models and cost functions, advanced image priors, and different noise models. The compiler intelligently chooses the best way to translate a problem formulation and choice of optimization algorithm into an efficient solver implementation. In applications to the image processing pipeline deconvolution in the presence of Poisson-distributed shot noise, and burst denoising, we show that a few lines of ProxImaL code can generate a highly-efficient solver that achieves state-of-the-art results. We also show applications to the nonlinear and nonconvex problem of phase retrieval.

Energy-efficient neuromorphic hardware and its use for deep neural networks

Date: Feb. 14, 2017 at 2pm

Speaker: Steve Esser (IBM)

Location: Sycamore Conference Room (040-195)

Abstract: Neuromorphic computing draws inspiration from the brain's structure to create energy-efficient hardware for running neural networks. Pursuing this vision, we created the TrueNorth chip, which embodies 1 million neurons and 256 million configurable synapses in contemporary silicon technology, and runs using under 100 milliwatts. Spiking neurons, low-precision synapses and constrained connectivity are key design factors in achieving chip efficiency, though they stand in contrast to today's conventional neural networks that use high precision neurons and synapses and have unrestricted connectivity. Conventional networks are trained today using deep learning, a field developed independent of neuromorphic computing, and are able to achieve human-level performance on a broad spectrum of recognition tasks. Until recently, it was unclear whether the constraints of energy-efficient neuromorphic computing were compatible with networks created through deep learning. Taking on this challenge, we demonstrated that relatively minor modifications to deep learning methods allows for the creation of high performing networks that can run on the TrueNorth chip. The approach was demonstrated on 8 standard datasets encompassing vision and speech, where near state-of-the-art performance was achieved while maintaining the hardware's underlying energy-efficiency to run at > 6000 frames / sec / watt. In this talk, I will present an overview of the TrueNorth chip, our methods to train networks for this chip and a selection of performance results.

<http://science.sciencemag.org/content/345/6197/668>

<http://www.pnas.org/content/113/41/11441.abstract>

Locating Features in Detector Images with Machine Learning

Date: Feb. 7, 2017 at 2pm

Speaker: David Schneider

Location: Sycamore Conference Room (040-195)

Abstract: Often analysis at LCLS involves image processing of large area detectors. One goal is to find the presence, and location of certain features in the images. We'll look at several approaches to locating features using machine learning. The most straightforward is learning from training data that includes feature locations. When location labels are not in the training data, techniques like guided back propagation, relevance propagation, or occlusion can be tried. We'll discuss work on applying these approaches. We'll also discuss ideas based on generative models like GAN's (Generative Adversarial Networks) or VAE's (Variational Auto Encoders).



machine_learnin...ctor_images.pdf

Tractable quantum leaps in battery materials and performance via machine learning

Date: Jan. 17, 2017

Speaker: Austin Sendek

Abstract: The realization of an all solid-state lithium-ion battery would be a tremendous development towards remedying the safety issues currently plaguing lithium-ion technology. However, identifying new solid materials that will perform well as battery electrolytes is a difficult task, and our scientific intuition on whether a material is a promising candidate is often poor. Compounding on this problem is the fact that experimental measurements of performance are often very time- and cost intensive, resulting in slow progress in the field over the last several decades. We seek to accelerate discovery and design efforts by leveraging previously reported data to train learning algorithms to discriminate between high- and poor performance materials. The resulting model provides new insight into the physics of ion conduction in solids and evaluates promise in candidate materials nearly one million times faster than state-of-the-art methods. We have coupled this new model with several other heuristics to perform the first comprehensive screening of all 12,000+ known lithium-containing solids, allowing us to identify several new promising candidates.



sendek_materials.pptx

Deep Learning and Computer Vision in High Energy Physics

Date: Dec 6, 2016

Speaker: Michael Kagan

Location: Kings River 306, B52

Abstract: Recent advances in deep learning have seen great success in the realms of computer vision, natural language processing, and broadly in data science. However, these new ideas are only just beginning to be applied to the analysis of High Energy Physics data. In this talk, I will discuss developments in the application of computer vision and deep learning techniques to the analysis and interpretation of High Energy Physics data, with a focus on the Large Hadron Collider. I will show how these state-of-the-art techniques can significantly improve particle identification, aid in searches for new physics signatures, and help reduce the impact of systematic uncertainties. Furthermore, I will discuss methods to visualize and interpret the high level features learned by deep neural networks that provide discrimination beyond physics derived variables, adding a new capability to understand physics and to design more powerful classification methods in High Energy Physics.



Kagan_MLHEP_Dec2016.pdf

Links to papers discussed:

<https://arxiv.org/abs/1511.05190>

<https://arxiv.org/abs/1611.01046>

Label-Free Supervision of Neural Networks with Physics and Domain Knowledge

Date: Oct 18, 2016

Speaker: Russell Stewart

Abstract: In many machine learning applications, labeled data is scarce and obtaining more labels is expensive. We introduce a new approach to supervising neural networks by specifying constraints that should hold over the output space, rather than direct examples of input-output pairs. These constraints are derived from prior domain knowledge, e.g., from known laws of physics. We demonstrate the effectiveness of this approach on real world and simulated computer vision tasks. We are able to train a convolutional neural network to detect and track objects without any labeled examples. Our approach can significantly reduce the need for labeled training data, but introduces new challenges for encoding prior knowledge into appropriate loss functions.



Russell_constrai...earning slac.key



Russell_constrai...earning slac.pdf

Can machine learning teach us physics? Using Hidden Markov Models to understand molecular dynamics.

Date: Sept 21, 2016

Speaker: T.J. Lane

Abstract: Machine learning algorithms are often described solely in terms of their predictive capabilities, and not utilized in a descriptive fashion. This “black box” approach stands in contrast to traditional physical theories, which are generated primarily to describe the world, and use prediction as a means of validation. I will describe one case study where this dichotomy between prediction and description breaks down. While attempting to model protein dynamics using master equation models — known in physics since the early 20th century — it was discovered that there was a homology between these models and Hidden Markov Models (HMMs), a common machine learning technique. By adopting fitting procedures for HMMs, we were able to model large scale simulations of protein dynamics and interpret them as physical master equations, with implications for protein folding, signal transduction, and allosteric modulation.



TJLane_SLAC_ML_Sem.pptx

On-the-fly unsupervised discovery of functional materials

Date: Aug 31, 2016

Speaker: Apurva Mehta

Abstract: Solutions to many of the challenges facing us today, from sustainable generation and storage of energy to faster electronics and cleaner environment through efficient sequestration of pollutants, is enabled by the rapid discovery of new functional materials. The present paradigm based on serial experimentation and serendipitous discoveries takes decades from initiation of a new search for a material to marketplace deployment of a device based on it. Major road-blocks in this process arise from heavy dependence on humans to transfer knowledge between interdependent steps. For example, currently humans look for patterns in current knowledge-bases, build hypotheses, plan and conduct experiments, evaluate results and extract knowledge to create the next hypothesis. The recent insight, emerging from the materials genome initiative, is that rapid transfer of information between hypothesis building, experimental testing and scale-up engineering can reduce the time and cost of material discovery and deployment by half. Humans, though superb at pattern recognition and complex decision making, are too slow and the major challenge in this new discovery paradigm is to reliably extract high-level actionable information from large and noisy data on-the-fly with minimal human intervention. In here, I will discuss some of the strategies and challenges involved in construction of unsupervised machines that perform these tasks on high throughput and large volume X-ray spectroscopic and scattering data sets.



ApurvaMehta_AI group talkv3.pptx

Machine Learning and Optimization to Enhance the FEL Brightness

Date: Aug 17, 2016

Speakers: Anna Leskova, Hananiel Setiawan, Tanner M. Worden, Juhao Wu

Abstract: Recent studies on enhancing the FEL brightness via machine learning and optimization will be reported. The topics are tapered FEL and improved SASE. The existing popular machine learning approaches will be reviewed and selected based on the characteristics of different tasks. Numerical simulation and preliminary LCLS experiment results will be presented.



Leskova_PresentAI.pptx

Automated tuning at LCLS using Bayesian optimization

Date: July 6, 2016

Speaker: Mitch McIntire

Location: Truckee Room, B52-206 T

Abstract: The LCLS free-electron laser has historically been tuned by hand by the machine operators. Existing tuning procedures account for hundreds of hours of machine time per year, and so efforts are underway to reduce this tuning time via automation. We introduce an approach for automated tuning using Bayesian optimization with statistical models called Gaussian processes. Initial testing has shown that this method can substantially reduce tuning time and is potentially a significant improvement on existing automated tuning methods. In this talk I'll describe Bayesian optimization and Gaussian processes and share some details and insights of implementation, as well as our preliminary results.



McIntire_AI-at-SLAC.pdf

Using Deep Learning to Sort Down Data

Date: June 15, 2016

Speaker: David Schneider

Abstract:

We worked on data from a two color experiment (each pulse has two bunches at different energy levels). The sample reacts differently depending on which of the colors lased and the energy in the lasing. We used deep learning to train a convolutional neural network to predict these lasing and energy levels from the xtcav diagnostic images. We then sorted down the data taken of the sample based on these values and identified differences in how the sample reacted. Scientific results from the experiment will start with an analysis of these differences. We used guided back propagation to see what the neural network identified as important and were able to obtain images that isolate the lasing portions of the xtcav images.



xtcav_mlearn.pdf