A University Center of Excellence supported by the Air Force Research
Laboratory (AFRL).
https://madlab.ml.wisc.edu/

Newsletter – Q4 2019

Table of Contents

Welcome Message from Professor Robert Nowak, University of Wisconsin-Madison and Lee Seversky, Air Force Research Laboratory

On behalf of the MADLab AFRL University Center of Excellence (UCoE) team for Efficient and Robust Machine Learning (ERML) and the Air Force Research Laboratory center's team, we would like to share the center's first quarterly newsletter. The goal of the newsletter is to provide a venue for showcasing research accomplishments, team members and projects, as well as share information about upcoming events with the academic and Air Force machine learning communities.

What is the ERML UCoE? The Efficiency and Robust Machine Learning University Center of Excellence is

Wisconsin-Madison in 2018 for up to 5 years. The center is a joint effort between the Air Force Research Laboratory, Information Directorate (AFRL/RI) and AFOSR. It is focused on advancing the state-of-the-art in efficient and robust machine learning methods as well as fostering a collaborative research environment between the university and AFRL government scientists and engineers.

Air Force Machine Learning Opportunities and Challenges: Machine Learning (ML) continues to emerge as a critical field of Artificial Intelligence (AI) and has a critical role to play in shaping the future Air Force. From shifting from today's high data processing and analysis burden from the Airman to machines, to increasing the overall quality and speed of decision making - advancements in Machine Learning hold great potential for the Air Force. However, for ML systems to be deployed in Air Force operational settings they must perform reliably in complex operational situations as they learn, make decisions, and act. Further, AF systems operate in degraded and uncertain environments. As such, current ML techniques that do not explicitly consider these challenges have little hope of achieving the high performance necessary for trusted intelligence systems to be deployed in operational environments.

The research goals of this UCoE are to study ML techniques under the lens of these operational learning settings and to develop novel, principled techniques that overcome challenges presented by operational constraints. Specifically this UCoE is focused on the challenges of efficient and robust machine learning which are tackled through four research thrusts: Data efficiency, Computational efficiency, Operational robustness, and Adversarial robustness.

The second focus of the UCoE is to establish a close collaboration between the MADLab university team and AFRL government scientists and engineers. The goal of this UCoE is to foster a collaborative environment to support joint research projects, university & government exchanges, and special events such as hack-a-thons, seminars, and challenge problems to maximally engage both university and AFRL participation. This is especially critical for this fast-paced and high demand area of machine learning where often real-world problems and data drive discovery of new methods and the application of machine learning techniques to problems and data identify new challenges. The UCoE seeks to bring the two communities together to jointly advance the fundamental and applied research to Air Force application.

Meet the Team
MADLAB Team
University of Wisconsin-Madison (UW)
Toyota Technological Institute at Chicago (TTIC)
University of Chicago (UC)

Center Director: Robert Nowak (UW)*

Thrust Leads
Data-Efficiency, Robert Nowak (UW) & Greg Shakhnarovich (TTIC)
Computational Efficiency, Mikko Lipasti (UW) & Dimitris Papailiopoulos (UW)
Adversarial Robustness, Jerry Zhu (UW) & Yingyu Liang (UW)
Operational Robustness, Rebecca Willett (UC) and Karen Livescu (TTIC)
For more information please see the MADLab Website

Executive Committee
Dr. Lee Seversky, Lee.Seversky@us.af.mil, AFRL, Information Directorate*
Dr. Erik Blasch, erik.blasch.1@us.af.mil, AFRL, Office of Scientific Research
Dr. Qing Wu, qing.wu.2@us.af.mil, AFRL, Information Directore

Thrust Leads
Data-Efficiency, Dr. Walter Bennette, walter.bennette.1@us.af.mil & Dr. Lee Seversky,
Lee.Seversky@us.af.mil, AFRL, Information Directorate
Computational Efficiency, Clare Thiem, clare.thiem@us.af.mil, AFRL, Information Directorate
Adversarial Robustness, Ryan Luley, ryan.luley@us.af.mil, AFRL, Information Directorate
Operational Robustness, Dr. Ashley Prater-Bennette, ashley.prater-bennette@us.af.mil, AFRL, Information
Directorate

*Team leads

Research Vignettes

Research Vignettes highlight research projects from the prior quarter.

Computational Efficiency via Bitstreams

by Mikko Lipasti, Professor of Electrical and Computer Engineering at the University of Wisconsin-Madison.

Lipasti is the Philip Dunham Reed Professor of Electrical and Computer Engineering at UW-Madison. His primary research interests include high-performance, low-power, and reliable processor cores; networks-on-chip for many-core processors; and fundamentally new, biologically-inspired models of computation. He earned his Ph.D. at Carnegie Mellon University in 1997, is the recipient of the Eta Kappa Nu Outstanding Young Electrical Engineer Award and the NSF CAREER Award, and was named an IEEE Fellow (class of 2013) "for contributions to the microarchitecture and design of high-performance microprocessors and computer systems." Lipasti leads the MadLAB thrust on Computational Efficiency

In extremely low-power applications of machine learning, such as pico-scale unmanned aerial vehicles, power efficiency becomes a primary design constraint. Lipasti's group has been exploring the use of stochastic computing [1] as a means for representing and operating on continuous values on such platforms. In stochastic computing, numeric values are constrained to the range [0,1] and are represented as the probability of a bit in a bitstream taking on the value of 0 or 1. For example, the value 0.25 could be represented with a bitstream of length 12, with three of the bits set to 1, e.g. 0010_1100_0000. This representation mimics the firing patterns or spike trains of biological neurons, and enables extremely low-cost datapath hardware (i.e. multiplication is accomplished with a single AND gate).

```scala
4 import bitstream.simulator.units._
5 import math._
6
7 trait Parameters {
8   val m: Int
9   val n: Int
10 }
11
12 object DefaultParams extends Parameters {
13   val m = 2
14   val n = 2
15 }
16
17 case class Module (params: Parameters) {
18
19   // Define outputs
20   val outputList = List (("v", params.n, 1),
21                          ("u", params.m, 1),
22                          ("sigma", 1, 1))
23
24   def loop(A: Matrix[SBitstream], v: Matrix[SBitstream]):
25       (Matrix[SBitstream], Matrix[SBitstream], SBitstream) = {
26     // Update right singular vector
27     var w = A * v
28     var wScaled = w :/ math.sqrt(params.m)
29     var u = wScaled / Matrix.norm(wScaled)
30
31     // Update left singular vector
32     var z = A.T * u
33     var zScaled = z :/ math.sqrt(params.n)
34     var sigma = Matrix.norm(zScaled)
35     var _v = zScaled / sigma
36
37     (u, _v, sigma)
38   }
39 }
```

proven to be an extremely challenging, time-consuming, and error-prone process, which discourages designers from adopting this approach. In our experience with developing bitstream-based linear solvers [2], algorithms are first developed in Matlab using floating-point data types, then rewritten to emulate stochastic logic in a high-level simulator, and finally prototyped using Verilog RTL on an FPGA substrate. Implementation and validation of three separate realizations is error-prone and requires significant duplication of effort.

To address this challenge, we have developed BitSAD, the first domain-specific language for bitstream computing [3,4]. As seen to the left, BitSAD allows users to write high-level algorithms like an iterative SVD in a Matlab-like syntax, utilizing high-level programming language features, custom data types for bitstream computing and libraries to perform convenient linear algebra matrix operations. Any algorithm developed in BitSAD can be simulated at the software-level to debug bitstream-level issues quickly, and the code automatically generates synthesizable Verilog that implements the algorithm. The Verilog can then be effortlessly deployed on a field-programmable gate array (FPGA) or even synthesized into a custom chip. In order to facilitate widespread adoption of this language, we have released the BitSAD compiler under an open source license, and it can be downloaded from https://github.com/UW-PHARM/BitSAD.

The stochastic computing research community has also struggled with standardized ways of evaluating and comparing stochastic computing implementations. To address this challenge, we developed BitBench, the first standardized benchmark suite for bitstream computing [5]. BitBench includes six key numeric kernels and two end-to-end applications, along with representative input data sets and run rules, and enables a standardized approach for benchmarking stochastic computing implementations. Like BitSAD, BitBench has been released under an open source license and can be downloaded from https://github.com/UW-PHARM/BitBench.

We are actively using these tools in our own group to investigate ultra-low power ML algorithms. Our hope is that releasing these open-source tools will stimulate wide-spread interest and further research into bitstream computing. For more information on BitSAD, and BitBench, please download the source code from the links above, or refer to the published papers that describe them in detail [3,4,5].

[1] B. R. Gaines. Stochastic Computing Systems, pages 37–172. Springer US, Boston, MA, 1969.
[2] R. Shukla, S. Khoram, E. Jorgensen, J. Li, M. Lipasti, and S. Wright. Computing generalized matrix inverse on spiking neural substrate. Frontiers in Neuroscience, 12:115, 2018.
[3] K. Daruwalla, H. Zhuo, and M. Lipasti. Bitsad: A domain-specific language for bitstream computing. In Proceedings of the First ISCA Workshop on Unary Computing (WUC'19), June 2019.

[5] K. Daruwalla, H. Zhuo, C. Schulz, and M. Lipasti. Bitbench: A benchmark for bitstream computing. In Proceedings of the International Conference on Languages Compilers, Tools and Theory of Embedded Systems (LCTES '19), June 2019.

Exploiting Class Learnability in Noisy Data

by Matthew Klawonn, Air Force Research Laboratory, Information Directorate

Matthew Klawonn received his PhD in computer science in 2019 from Rensselaer Polytechnic Insitute (RPI), where he was a recipient of the Department of Defense SMART scholarship. He is now a research scientist at the Air Force Research Laboratory working under the Autonomous Command and Control (AC2) competency. Matt's thesis focused on the combination of supervised machine learning and knowledge representation and reasoning techniques for the completion of difficult perceptual tasks. His current interests lie in the same area, specifically focusing on using structured auxiliary knowledge to help with a few short learning scenarios. What follows is an abridged highlight of his work with collaborators from AFRL and RPI published in [1] at AAAI '19.

Exploiting Class Learnability in Noisy Data

Many state of the art machine learning models require large amounts of data to avoid overfitting. In constructing large scale data sets, one often turns to accessible sources that can quickly provide vast amounts of data, including crowdsourcing and Webly supervised learning (scraping data and labels from the Web). Such processes frequently result in data that is noisy. For example, consider a data set of images and tags scraped from Flickr, intended to be used as training data for an image tagger. Tags on Flickr are assigned by users of the site without restriction, and the assignment of tags can range from the very literal to the abstract. When attempting to learn an image tagger, training on noisy classes is potentially harmful to model performance on classes with strong signal (see Figure 1 for an example). Rather than try to achieve moderate performance on all classes at the expense of those with strong signal, the classifier designer may want to avoid noisy classes entirely. With this in mind, in order to develop an accurate classifier in the presence of noisy classes, we hypothesize that not all classes ought to be treated equally during training. Rather, it may be beneficial to focus training on classes for which the model appears capable of achieving a low generalization error. We call this characteristic of classes appearing in the data set learnability.
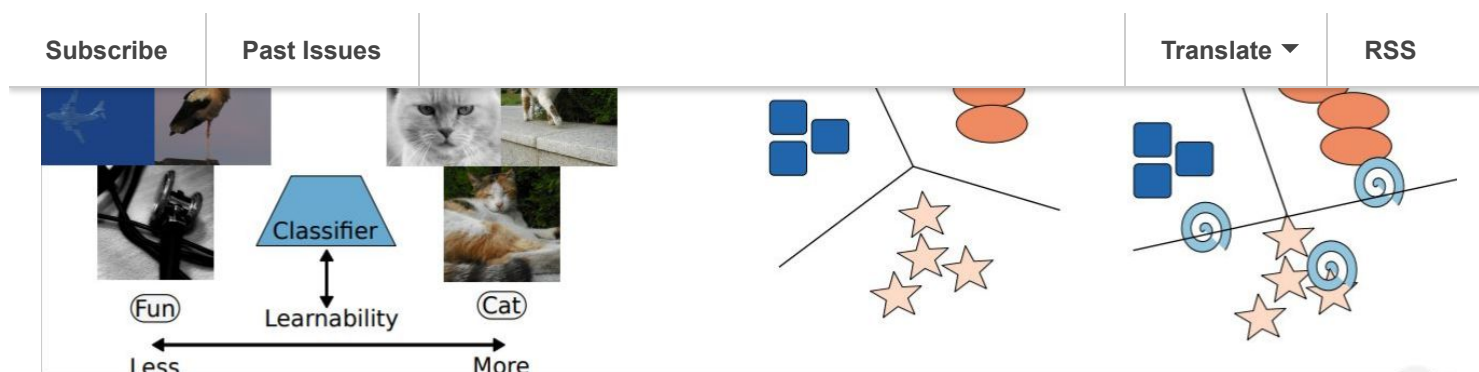
Figure 1

Figure 1: Left: Example of labeled Flickr images. The common features of the "Cat" images suggest that a low generalization error can be achieved for this class. Right: An illustration of how a class with poor signal can negatively affect a classifier.

To this end, we seek to develop an algorithm that uses the concept of learnability to focus the training of a classifier on specific classes. In this work we propose an online algorithm that acts in concert with the stochastic learning algorithm used to train a classifier. For every class, our algorithm assesses the generalization performance of the classifier, and learns online to spend more iterations on learnable classes. We pose this learning scheme as a Multi-Armed Bandit (MAB, bandit) problem. By doing so, we are able to leverage prior MAB work that has developed algorithms with theoretical guarantees on performance. Our MAB algorithm is able to assess the learnability of each class by sampling only from a single class at each time-step. As a result, we are able to efficiently select classes from which to learn, incurring only slight overhead to the normal stochastic training process.

Our bandit algorithm, based on prior work in [2], estimates generalization performance of a given classifier on a particular class via training and validation batches. Specifically, when a class is sampled, the difference in performance on a validation set before and after training is measured, with the relative improvement indicating how well the model has generalized on examples from the given class. Such improvement is known as self prediction gain (SPG) in the literature [3]. At a given training iteration t, data from a single class is sampled via an upper confidence bound, i.e the class with the highest sum of historical SPG and uncertainty is chosen. Given that the classifier parameters change over time, we hypothesize that the ability of the classifier to generalize on batches of a given class may also change. Therefore we employ the discounting scheme of [2], an approach that, for calculation of class upper confidence bound at iteration t, ensures the SPG observed at time t − n decreases in influence as n increases. An illustration of our algorithm can be found in Figure 2.

We empirically evaluate our model in two scenarios, one synthetic and one with existing data sets. We construct a synthetic data set from Cifar100 [4] by augmenting the existing data with noisy classes. Specifically, we create a new classification task with 200 classes total: 100 original classes with around 600 instances each, and 100 new classes with 600 instances uniformly sampled at random from existing classes. Intuitively, our bandit selection mechanism should work in conjunction with the classifier to identify the original 100 classes as having a stronger signal than the 100 new classes. Our results align
with this intuition, showing that out of the 100 classes selected most.
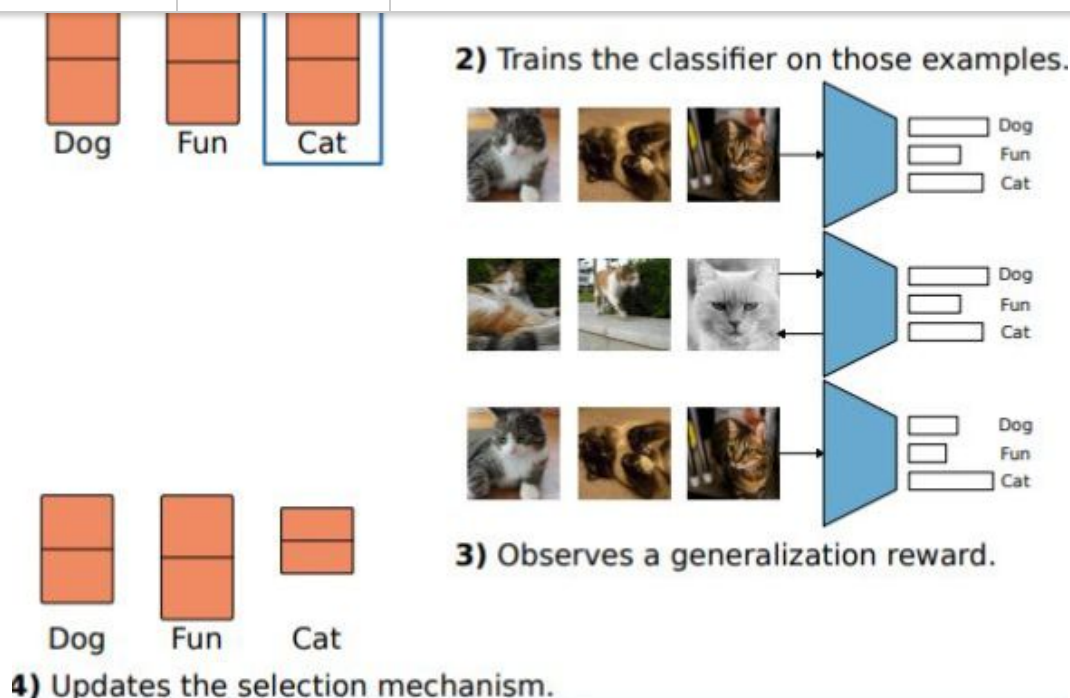
Figure 2

Figure 2: A single iteration of the bandit supervised training procedure. Generalization reward is based on improvement on a validation batch after training.

In addition to evaluating the algorithm's ability to find classes with strong signal, we intend to show that the resulting trained classifier has high accuracy on classes frequently selected for training. To this end, we employ our training algorithm on two data sets; the YFCC100m data set [5] of images and tags from Flickr, and the Recipe1m data set [6] containing images of prepared food dishes labeled with their ingredients. Our results show that in contrast to classifiers trained via more traditional algorithms, our bandit trained classifier is able to achieve very high performance on a subset of classes.

For further results and discussion, please see our paper [1].

References

[1] M. Klawonn, E. Heim, and J. Hendler, "Exploiting class learnability in noisy data," in Proc. 33rd AAAI Conf. on Artificial Intelligence, (Honolulu, HI, USA), Jan. 2019. to be published.

[2] I. Bogunovic, J. Scarlett, and V. Cevher, "Time-varying gaussian process bandit optimization," in Proc. 19th Annu. Int. Conf. Artificial Intelligence and Statistics, (Cadiz, Spain), pp. 314–323, May 2016.

[3] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," vol. 11, no. 6, pp. 265–286, 2007.

[4] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Master's thesis, Dept. Comput. Sci., Univ. Toronto, Toronto, Canada, Apr. 2009.

[5] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li, "Yfcc100m: The new data in multimedia research," Comm. of the ACM, vol. 59, no. 2, pp. 64–73, 2016.

embeddings for cooking recipes and food images," vol. 720, no. 619-508, p. 2, 2017.

Learning Nearest Neighbor Graphs from Noisy Distance Samples
by Ardhendu Tripathy and Blake Mason

Ardhendu Tripathy is a postdoctoral research associate at the Wisconsin Institute for Discovery. His current research focusses on designing and analyzing learning algorithms and methods in statistical learning theory. He obtained his Ph.D. in Electrical and Computer Engineering from Iowa State University, where he received the Research Excellence Award from the Graduate College for his dissertation.

Blake Mason is a Ph.D. student in Electrical and Computer Engineering at the University of Wisconsin-Madison. His research focus is on active learning and metric learning to analyze human-generated data. His Bachelor's of Science was in Electrical Engineering at the University of Southern California.

Frequently, in machine learning problems we are interested in knowing which item in a dataset is closest or most similar to a given item. For instance, in the case of online retail, a retailer may wish to know which products are most similar to each other to improve their recommendations. In the case of telecommunications, providers often wish to connect phones to the nearest towers that can support the best data rate. The closest item can be identified by defining an appropriate notion of similarity or distance between a pair of items. However, the exact distance function to be used is often unclear. In the two situations described above, we see it is possible to *learn* the distance between items from past customer or user interactions. This approach will have to contend with noisy measurements--- different customers have different similarity notions for products and different wireless towers have different physical and load conditions. Furthermore, the queried item (the clicked product, the pinged tower) is not known a priori.

To efficiently answer these "closest item" queries for any referred item, Blake Mason, Ardhendu Tripathy, and Robert Nowak model this question as the Nearest Neighbor Graph Problem and have developed a novel multi-armed bandit algorithm, ANNTri, to solve it. The algorithm only requires access to an oracle that returns

This matches the known optimal complexity for finding the Nearest Neighbor Graph when the oracle is noise-free. The reduction in queries compared to brute force is obtained by using basic properties of distance metrics, i.e., symmetry and triangle inequality, to reduce the search space of possible graphs.

ANNTri proceeds in rounds and uses the metric properties to share information between rounds. Specifically, in the ith round, the algorithm does the following:

1. Via symmetry, reuse samples from rounds $1, \ldots, i - 1$ to estimate $d(x_i, x_j)$ for all $i < j$ as a warm start.
2. Compute upper and lower bounds on all distances via the triangle inequality and the bounds at rounds $1, \ldots, i - 1$.
3. With triangle bounds and existing distance samples, eliminate points that cannot be the nearest neighbor.
4. Find point i's nearest neighbor among the remaining points using a modified successive elimination procedure that handles the warm start estimates.

To demonstrate the real-world empirical performance of the method, the authors used ANNTri to learn which pairs of shoes are most similar to each other from a set of triplet queries of 85 shoe images previously collected by air force researchers. Using these triplet queries, the authors define nearest neighbors, and show that their method outperforms both random sampling and a state-of-the-art ordinal embedding method (STE) by an order of magnitude, as shown below.
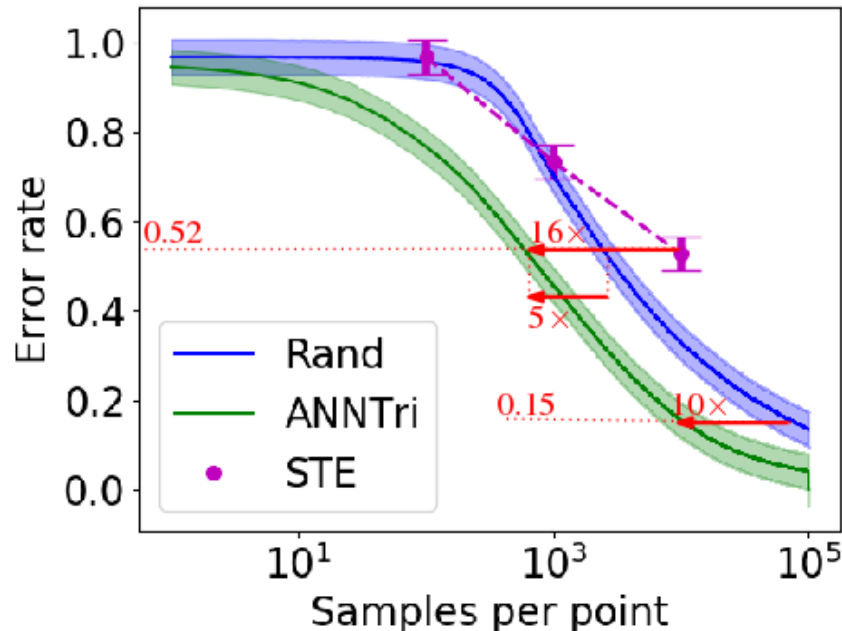
More details can be found in the ArXiv draft: Learning Nearest Neighbor Graphs from Noisy Distance Samples

Looking Ahead

Go here for information on monthly MADLab/AFRL WebEx's featuring a short research discussion

SILO is a weekly seminar series which hosts a catered lunch every Wednesday at the Wisconsin Institute for Discovery for graduate students from across campus. Researchers from Computer Science, Engineering, Mathematics and Statistics make up the core of SILO, but those from other fields are strongly encouraged to participate.

Publications

- S Rajput, H Wang, Z Charles, D Papailiopoulos. DETOX: A Redundancy-based Framework for Faster and More Robust Gradient Aggregation. 2019.
- Yuanzhi Li, Yingyu Liang. Learning Overparameterized Neural Networks via Stochastic Gradient Descent on Structured Data. Aug 2019.
- Ayon Sen, Xiaojin Zhu, Liam Marshall, Robert Nowak. Should Adversarial Attacks Use Pixel p-Norm? Jun 2019.
- Sumeet Katariya, Ardhendu Tripathy, Robert Nowak. MaxGap Bandit: Adaptive Algorithms for Approximate Ranking. Jun 2019.
- Rohit Shukla, Mikko Lipasti, Brian Van Essen, Adam Moody, Naoya Maruyama. REMODEL: Rethinking Deep Cnn Models to Detect and Count on a NeuroSynaptic System.
- Kyle Daruwalla, Heng Zhuo, Carly Schulz, and Mikko Lipasti. BitBench: A Benchmark for Bitstream Computing," in Proceedings of Languages, Compilers, Tools and Theory of Embedded Systems. Jun 2019.
- Kyle Daruwalla, Heng Zhuo and Mikko Lipasti. BitSAD: A Domain-Specific Language for Bitstream Computing. Jun 2019.
- Kyle Daruwalla and Mikko Lipasti. Resource Efficient Navigation Using Bitstream Computing. Jun 2019.
- Michael Mishkin. Efficient Inference Acceleration. Jun 2019.
- Gilton, G. Ongie, R. Willett. Neumann Networks for Inverse Problems in Imaging. Jun 2019.
- Blake Mason, Ardhendu Tripathy, Robert Nowak. Learning Nearest Neighbor Graphs from Noisy Distance Samples. May 2019.
- Mina Karzand, Robert D. Nowak. Active Learning in the Overparameterized and Interpolating Regime. May 2019.
- Shengchao Liu, Mehmet Furkan Demirel, Yingyu Liang. N-Gram Graph: A Simple Unsupervised Representation for Molecules. 2019.
- Jiefeng Chen, Xi Wu, Vaibhav Rastogi, Yingyu Liang, Somesh Jha. Robust Attribution Regularization. 2019.
- Zeyuan Allen-Zhu, Yuanzhi Li, Yingyu Liang. Learning and Generalization in Overparameterized Neural Networks, Going Beyond Two Layers. 2019.
- Pasad, B. Shi, H. Kamper, and K. Livescu. On the contributions of visual and textual supervision in low-resource semantic speech retrieval. 2019.
- Kamper, A. Anastassiou, and K. Livescu. Semantic query-by-example speech search using visual grounding. 2019.
- Yuval Bahat, Michal Irani, Gregory Shakhnarovich. Natural and Adversarial Error Detection using Invariance to Image Transformations. Feb 2019.

Shakhnarovich. DIODE: A Dense Indoor and Outdoor Depth Dataset. 2019.

- Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Recurrent Back-Projection Network for Video Super-Resolution.

- Rajput, Z. Feng, Z. Charles, P.-L. Loh, D. Papailiopoulos. Does Data Augmentation Lead to Positive Margin? 2019.

- Liu, D. Papailiopoulos, D. Achlioptas. Bad Global Minima Exist and SGD Can Reach them. 2019.

- J. Mardia, J. Jiao, E. Tánczos, R. Nowak and T. Weissman. Concentration Inequalities for the Empirical Distribution.

- Z. Charles, H. Rosenberg, D. Papailiopoulos. A Geometric Perspective on the Transferability of Adversarial Directions. AISTATS 2019.

- Jiefeng Chen, Xi Wu, Vaibhav Rastogi, Yingyu Liang, Somesh Jha. Towards Understanding Limitations of Pixel Discretization Against Adversarial Attacks. *IEEE European Symposium on Security and Privacy 2019 (EuroS&P), 2019.*

- Hongyang Zhang, Vatsal Sharan, Moses Charikar, Yingyu Liang. Recovery Guarantees for Quadratic Tensors with Limited Observations.
  International Conference on Artificial Intelligence and Statistics (AISTATS), 2019.

- Shengchao Liu, Yingyu Liang, Anthony Gitter. Loss-Balanced Task Weighting to Reduce Negative Transfer in Multi-Task Learning. *AAAI Conference on Artificial Intelligence (AAAI), 2019, Student Abstract and Poster Program.*

- Yuzhe Ma, Xiaojin Zhu, Justin Hsu. Data Poisoning against Differentially-Private Learners: Attacks and Defenses.

- Xuezhou Zhang and Xiaojin Zhu. Online Data Poisoning Attacks.

- Yiding Chen and Xiaojin Zhu. Optimal Adversarial Attack on Autoregressive Models.

- Laurent Lessard, Xuezhou Zhang, and Xiaojin Zhu. An optimal control approach to sequential machine teaching. In The 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), 2019.

- U. Oswal and R. Nowak. Scalable Sparse Subspace Clustering via OWL.

- G. Ongie, L. Balzano, D. Pimentel, R. Willettt and R. Nowak. Tensor Methods for Nonlinear Matrix Completion.

- Y. Ma, R. Nowak, P. Rigollet, X Zhang, and X. Zhu. Teacher Improves Learning by Selecting a Training Set.

- S. Katariya, L. Jain, N. Sengupta, J. Evans and R. Nowak. Adaptive Sampling for Clustered Ranking.

Subscribe to This Newsletter