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.
address the operational Air Force learning challenges of efficiency and robustness. The ERML UCoE was awarded to the University of 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**
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

AFRL Center Team
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 Director

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 recent research projects.

Towards Better Understanding and Robustness of Deep Learning
by Yingyu Liang, Assistant Professor of Computer Sciences at the University of Wisconsin-Madison

Liang is an assistant professor of Computer Sciences at the University of Wisconsin-Madison since 2017. His research focuses include theoretical analysis of deep learning, robust machine learning, and their applications. He received his Ph.D. in 2014 from Georgia Institute of Technology advised by Maria Florina Balcan, and was a postdoc at Princeton hosted by Sanjeev Arora.

Deep learning has achieved unprecedented success and is a primary driving force behind many current intelligent decision-making systems. Besides the empirical success, provable guarantees have also become a sought-after goal. Unfortunately, the lack of adequate theoretical understanding is limiting our capacity to fully exploit the potential of deep learning in realistic environments, such as in security-sensitive scenarios.
learning achieve provable learning guarantees? A recent line of works on overparameterized networks has provided some insights. It has been observed that neural networks in practice are typically overparameterized (i.e., the model is larger than statistically necessary to perfectly fit the training data), and overparameterization can help the learning (e.g., [1,2]). This contrasts with the prediction by traditional learning theory. Our work [3,4] has studied the optimization and generalization of overparameterized neural networks, which is among the earliest works providing new insights and opening this direction.

The work [3] studied the problem of learning a two-layer ReLU neural network via stochastic gradient descent (SGD) from random initialization. When the data comes from mixtures of well-separated distributions, we prove that SGD learns a network with a small generalization error, albeit the network has enough capacity to fit arbitrary labels. Furthermore, the analysis provides interesting insights into several aspects of deep learning neural networks. It shows that with sufficient overparameterization the training dynamics stay close to the initialization and are almost convex. This insight has been used to prove convergence and generalization guarantees for general deep networks in subsequent works. They can be viewed as approximating overparameterized deep learning by convex learning with the Neural Tangent Kernel (NTK) in [5]. Our work [4] goes beyond this convex approximation, extending the above insights to prove results in a more general setting. We prove that using overparameterized neural networks, one can (improperly) learn some notable hypothesis classes, including two and three-layer neural networks with fewer parameters and smooth activations. Moreover, the learning process can be done efficiently by SGD or its variants, but over a non-convex loss landscape.

We have also studied the robustness of deep learning under adversarial attacks. It has been shown that carefully chosen small perturbation can be added to the input and change the prediction of deep learning models. [6] analyzes a defense method for such adversarial examples called pixel discretization defense and attempt to study when it works and when it does not. Besides robust prediction, we also consider robustness in other aspects such as the interpretation of the model. A prototypical task of interpretation is for a given network and a given input, to produce an attribution vector measuring the relative importance of each feature in the input for prediction. Previous work (e.g., [7]) showed that adversarial attacks could easily fool existing attribution methods. Our work [8] takes a step towards solving this robustness issue. We considered the integrated-gradients attribution framework in [9] as the interpretation and then designed training objectives in classic robust optimization models to achieve robust attribution. Experiments demonstrated that the method significantly improves the robustness of the interpretation; see Figure 1. We also proved that our framework includes previous objectives for robust prediction as special cases, and they naturally degenerate to classic soft-margin training for one-layer neural networks. This connection thus provides an explanation to the interesting observation that robust prediction sometimes also helps robust interpretation [10,11]. It also gives positive support for achieving both simultaneously. Indeed, our robust attribution method already leads to comparable or even better accuracy under attacks when compared to adversarial training.
Figure 1. Our method [8] leads to semantically meaningful and robust attribution under attacks. Top row: original image, bottom row: adversarial example.

Detection and Description of Change in Visual Streams

written by Greg Shakhnarovich

This is a joint project conducted by CoE co-PIs Greg Shakhnarovich (TTIC) and Rebecca Willett (University of Chicago) and PhD students Davis Gilton (Wisconsin) and Ruotian Luo (TTIC). Its focus is on analysis of visual streams - ordered sequences of images, possibly separated by significant time gaps. Our goal is to detect and describe “relevant” changes while ignoring nuisance changes such as lighting or perspective changes or seasonal change. This change detection and description task is relevant to intelligence, insurance, urban planning, and natural disaster relief.

In our initial effort in this direction, we propose a new approach to incorporating unlabeled data into training to generate natural language descriptions of change. This is a key requirement for robust, data-efficient learning in this domain, since manual annotation of change in visual stream is extremely labor- and cost-intensive. We also develop a framework for estimating the time of change in visual stream (i.e., a change point detector). Our approach uses learned representations for change evidence and consistency of perceived change, and combines these in a regularized graph cut based change detector. Experimental evaluation on visual stream datasets, which we release as part of our contribution, shows that representation learning driven by natural language descriptions significantly improves change detection accuracy, compared to methods that do not rely on language. Figure: streetchange.png

A subsequence from “Street Change,” a proposed dataset for change detection and captioning for visual streams. The dotted line denotes the location of a changepoint, where some meaningful element of the scene has changed. Some human captions for the change across the changepoint are: “A road sign is gone.” “The orange construction sign was removed.” “The safety sign is not in the road.” Our approach learns to both detect the change point (dotted line) and provide automated captions that captures the change description as would be provided by a human (rather than "nuisance" change, like seasonal variations).
Looking Ahead

Virtual AFRL ERML-COE Workshop May 19-20, 10am-3pm EST
AFRL and the ERML CoE have organized this workshop to bring together academia and government researchers, scientists, engineers, program managers, and senior leaders to share and discuss adversarial robustness in the context of artificial intelligence and machine learning.

For registration information and more details please see the website.

Recent and Upcoming SILOS http://silo.ece.wisc.edu/web/content/seminars
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.

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

Publications


M Karzand, RD Nowak. MaxiMin Active Learning in Overparameterized Model Classes IEEE Journal on Selected Areas in Information Theory.


Subscribe to This Newsletter

Want to change how you receive these emails? You can update your preferences or unsubscribe from this list.