COMPUTER SCIENCE

GRADUATE RESEARCH DAY

October 28, 2022

RevBits
Cyber Security Solutions

Stony Brook University

DEPARTMENT OF COMPUTER SCIENCE

Keynote Talk

Computer Science, Empiricism, and Better Neural Networks

Jonathan Frankle

Chief Scientist at MosaicML and Assistant Professor of Computer Science at Harvard

In computer science, we are trained to think like mathematicians. Something only becomes knowledge when we can prove it to be true. In this talk, I will argue that this narrow conception of knowledge is insufficient for us to make progress in deep learning. Practical neural network systems defy formalization: they are extraordinarily complex, they rely on messy real-world data, and the aspects that make them exciting are the strange phenomena that emerge from this complexity. I will tell this story through the lens of my research over the past several years. During that time, I have sought to understand how practical neural networks learn in hopes of improving our training algorithms to bring down the mind-boggling costs of training them. This includes focused work on neural network pruning - removing unnecessary connections early in training - and broad work on building entirely new training recipes involving dozens of changes like this. The findings suggest we have significant room for progress: early in neural network training, it is possible to remove up to 90% of connections without affecting final model quality, and new recipes can yield 7x reductions in overall training cost. All of this work rests on careful empiricism, not proofs. I have pursued this work both during my PhD at MIT and at my startup MosaicML; I will close by discussing what work is and isn't possible in academia and in industry.

Machine Learning Session Presentations

Best Presentation: Intelligent Systems

Understanding Human Hands in Visual Data

Supreeth Narasimhaswamy

Faculty Advisor
Minh Hoai

Hands are the central means by which humans interact with their surroundings. Understanding hands help human behavior analysis and facilitate other tasks such as action recognition. Recently, there has been a surge of interest in understanding first-person visual data, and hands are dominant interaction entities in such activities. Also, there is an explosion of interest in developing computer vision methods for augmented and virtual reality. To deliver an authentic AR/ VR experience, we must enable humans to interact with virtual world and allow virtual avatars to interact like real humans. Since hands are the dominant interaction entities in such cases, a thorough understanding of human hands is essential.

The first step toward the visual understanding of hands is to detect hands, and we propose a contextual attention method to detect them. While it is essential to detect hands, this is insufficient. When humans interact with their surroundings, they contact objects and other humans. Therefore we need to understand contact information to have a meaningful understanding of hands. We address this in hand contact recognition. To truly understand how humans interact with the world, we need to know where hands move across time and what objects interact with them. Therefore we address hand tracking. While it is essential to detect, track and obtain contact states of hands, this is insufficient for human understanding. For a scene containing multiple people, we need to know what object is manipulated by whom and which person is performing what activity. Therefore we address hand-body association.

Learning Topological Interactions for Multi-Class Medical Image Segmentation

Saumya Gupta

Deep learning methods have achieved impressive performance for multi-class medical image segmentation. However, they are limited in their ability to encode topological interactions among different classes (e.g., containment and exclusion). These constraints naturally arise in biomedical images and can be crucial in improving segmentation quality. In this paper, we introduce a novel topological interaction module to encode the topological interactions into a deep neural network. The implementation is completely convolution-based and thus can be very efficient. This empowers us to incorporate the constraints into end-to-end training and enrich the feature representation of neural networks. The efficacy of the proposed method is validated on different types of interactions. We also demonstrate the generalizability of the method on both proprietary and public challenge datasets, in both 2D and 3D settings, as well as across different modalities such as CT and Ultrasound. Code is available at: \url{https://github.com/TopoXLab/TopoInteraction}. This work has been accepted at ECCV 2022 for an oral presentation.

Faculty Advisor
Chao Chen

MagNET: Modality-Agnostic Network for Brain Tumor Characterization with Missing Modalities

Aishik Konwer

Multiple modalities provide complementary information in medical image segmentation tasks. However, in practice, not all modalities are available during inference. Missing modalities may affect the performance of segmentation and other downstream tasks. Previous approaches either use naive statistical computation to fuse multi-modal features, or attempt to synthesize missing modalities in image or feature space. We propose an end-to-end robust segmentation pipeline, MagNET, to handle heterogeneous modality combinations. A novel attention-based fusion module is designed to generate a modality-agnostic tumor-aware representation. We design an adversarial training strategy to improve the quality of the representation. A missing-modality detector is used as a discriminator to push the encoded feature representation to mimic a full modality setting. In addition, we introduce a loss function to maximize inter-modal correlations; this helps generate the modality-agnostic representation. MagNET significantly outperforms state-of-the-art segmentation methods under several missing modality scenarios, as demonstrated on a large scale brain tumor MRI dataset.

Faculty Advisor
Prateek Prasanna

Detecting Dissonant Stance in Social Media: The Role of Topic Exposure

We address dissonant stance detection, classifying conflicting stance between two input statements. Computational models for traditional stance detection have typically been trained to indicate pro/con for a given target topic (e.g. gun control) and thus do not generalize well to new topics.
In this work, we systematically evaluate the generalizability of dissonant stance detection to situations where examples of the topic have not been seen at all or only a few times.
We show that dissonant stance detection models trained only on a small number of non-target topics can perform as well as those trained on a target topic and that adding non-target topics boosts performance up to approximately 32 topics, suggesting dissonant stance detection models can generalize to unanticipated topics.

Vasudha Varadarajan

Faculty Advisor
H. Andrew Schwartz

Brain Cancer Survival Prediction on Pre-treatment MRI using Deep Anchor Attention Learning

Image-based brain cancer prediction models, based on radiomics, quantify the radiologic phenotype from magnetic resonance imaging (MRI). However, these features are difficult to reproduce because of variability in acquisition and preprocessing pipelines. Despite evidence of intra-tumor phenotypic heterogeneity, the spatial diversity between different
slices within an MRI scan has been relatively unexplored using such methods. In this work, we propose a deep anchor attention aggregation strategy with a Vision Transformer to
predict survival risk for brain cancer patients. A Deep Anchor Attention Learning (DAAL) algorithm is proposed to assign different weights to slice-level representations with trainable distance measurements. We evaluated our method on N = 326 MRIs. Our results outperformed attention multiple instance learning-based techniques. DAAL highlights the importance of critical slices and corroborates the clinical intuition that inter-slice spatial diversity can reflect disease severity and is implicated in outcome.

Xuan Xu

Faculty Advisor
Prateek Prasanna

Taming Entangled Accessibility Forum Threads for Efficient Screen Reading

Anand Aiyer

Faculty Advisor
I.V. Ramakrishnan

Blind Screen Reader users are forced to navigate Accessibility Forum discussion threads serially one line at a time using specialized keyboard shortcuts. This is tedious and cognitively overwhelming.

We propose partitioning forum threads semantically by disentangling chains of posts that talk about the same topic. The forum thread can be represented as a two-level hierarchical list with posts starting new conversations (conversation starts) at one level and a list of all posts that reply to a particular conversation start (replies) as another level. Users can now browse and select from a short list of conversation starts and focus only on replies of interest instead of going through all the posts one by one. We implement this solution as a chrome extension. It replaces the existing thread with an accessible interface containing a two-level list of conversation starts and replies representing that thread and supports navigation with keyboard arrow keys.

A user study with 11 blind screen reader users shows the proposed interface makes screen reading on accessibility forum threads more efficient. Results show a 57.19 percent reduction on average in total user input actions, a 45.27 percent reduction on average in task completion times, and a 34.59 percent reduction on average in perceived cognitive load while navigating the proposed interface with a screen reader compared to the status quo of navigating the existing accessibility forum web interface with a screen reader.

Systems and Security Session Presentations

Participant's Choice: Best Presentation

Too Good To Be True: Exploring the Ecosystem of Cryptocurrency Giveaway Scams

Xigao Li

Faculty Advisor
Nick Nikiforakis,
Amir Rahmati

As cryptocurrencies increase in popularity and users obtain and manage their own assets, attackers are pivoting from just abusing cryptocurrencies as a payment mechanism, to stealing crypto assets from end users. In this paper, we report on the first large-scale analysis of cryptocurrency giveaway scams. Giveaway scams are deceptively simple scams where attackers set up webpages advertising fake events and promising users to double or triple the funds that they send to a specific wallet address. To understand the population of these scams in the wild we design and implement our tool, that uses Certificate Transparency logs to identify likely giveaway scams. Through a 6-month-long experiment, our tool identified a total of 10,079 giveaway scam websites targeting users of all popular cryptocurrencies. Next to analyzing the hosting and domain preferences of giveaway scammers, we perform the first quantitative analysis of stolen funds using the public blockchains of the abused cryptocurrencies, extracting the transactions corresponding to 2,266 wallets belonging to scammers. We find that just for the scams discovered in our reporting period, attackers have stolen the equivalent of tens of millions of dollars, organizing large-scale campaigns across different cryptocurrencies. Lastly, we find evidence that attackers try to re-victimize users by offering fund-recovery services and that some victims send funds multiple times to the same scammers.

eAudit: A Deployable System for Audit Data Collection

Hanke Kimm

We are in an era of long-running cyber attack campaigns, known as "Advanced and Persistent Threats" (APTs). APT campaigns can remain hidden within a victim enterprise's network while exfiltrating data, installing malware, etc. Audit logs provide a crucial resource for after-the-fact detection, the primary recourse of combating APTs. However, existing systems, both enterprise and experimental, suffer from performance problems stemming from audit log generation, such as high data loss and performance overhead, as well as excessive data volume and log tampering vulnerability. Higher performance audit data collection systems require potentially unsafe changes to kernel OS code, brought on by system stability or compatibility issues. To address these drawbacks, we propose a novel approach for audit data collection called eAudit, a system that relies on the Extended Berkeley Packet Filter (eBPF) framework. Through eBPF, eAudit can be deployed safely on stock Linux distributions through the use of safe probes at various hooks within the kernel. We demonstrate the efficacy of eAudit by providing a multi-level buffer design and preliminary performance evaluation. We show that eAudit is a highly performant, robust and compatible auditing system in contrast to current audit logging systems.

Faculty Advisor
R. Sekar

Best Presentation: Systems and Security

AccessWear: Making Smartphone Applications Accessible to Blind Users

Prerna Khanna

AccessWear is a system that improves the accessibility of smartphone touchscreen interactions for blind users using smartwatch gestures. Our system design is human-centered, it incorporates the design goals that were learned from a formative user study with 9 blind users. Our study showed that blind users prefer to use smartwatch gestures to interact with their phones because it allows them to perform gestures with one hand and they do not have to get their phones out in public. In addition, the gestures performed by blind users have different patterns compared to sighted users and there is more variability amongst the users. To this end, AccessWear makes two contributions.
The first is a gesture recognition system that works specifically for blind users that is lightweight and does not require per-person training. The second is a near-zero-effort gesture replacement system. AccessWear uses input virtualization techniques so that a given gesture can replace the touchscreen input without requiring changes to the application. We implement AccessWear on an Android smartphone and watch. We perform a quantitative and qualitative study with 8 blind users. Our study shows that AccessWear can recognize gestures with a 92% accuracy and end-to-end latency when using an alternate gesture is less than 100ms. The qualitative study shows that when users perform a task, consisting of a series of gestures, the system is robust, does not have perceived delays, and does not add physical or mental load on the users. Macro benchmark study with sighted users shows that the false positive rate when using AccessWear during everyday activities is low and AccessWear can be used to interact with common applications without requiring changes to the application.

Faculty Advisor
Aruna Balasubramanian

Best Overall Presentation

Leveraging Earables for Unvoiced Command Recognition

Tanmay Srivastava

Faculty Advisor
Shubham Jain

This poster presents MuteIt, an ear-worn system for recognizing unvoiced human commands. MuteIt presents an intuitive alternative to voice-based interactions that can be unreliable in noisy environments, disruptive to those around us, and compromise our privacy. We propose a twin-IMU setup to track the user's jaw motion and cancel motion artifacts caused by head and body movements. MuteIt processes jaw motion during word articulation to break each word signal into its constituent syllables, and further each syllable into phonemes (vowels, visemes, and plosives). Recognizing unvoiced commands by only tracking jaw motion is challenging. As a secondary articulator, jaw motion is not distinctive enough for unvoiced speech recognition. MuteIt combines IMU data with the anatomy of jaw movement as well as principles from linguistics, to model the task of word recognition as an estimation problem. Rather than employing machine learning to train a word classifier, we reconstruct each word as a sequence of phonemes using a bi-directional particle filter, enabling the system to be easily scaled to a large set of words. We validate MuteIt for 20 subjects with diverse speech accents to recognize 100 common command words. MuteIt achieves a mean word recognition accuracy of 94.8% in noise-free conditions. When compared with common voice assistants, MuteIt outperforms them in noisy acoustic environments, achieving higher than 90% recognition accuracy. Even in the presence of motion artifacts, such as head movement, walking, and riding in a moving vehicle, MuteIt achieves a mean word recognition accuracy of 91% over all scenarios.

Access Control for Augmented Reality Systems

Sanket Goutam

Faculty Advisor
Amir Rahmati

Augmented Reality (AR) is widely considered the next evolution in personal devices, enabling seamless integration of the digital world into our reality. Such integration, however, often requires unfettered access to sensor data, causing significant overprivilege for applications that run on these platforms. Through analysis of two AR frameworks and 45 popular AR applications, we identify key trends in how these applications use sensor data and pinpoint unique threats users face in AR environments. Using these findings, we present Erebus, an access control framework for AR platforms that enables seamless fine-grained control over data used by AR applications. Erebus introduces a domain-specific language (DSL) for permission control in AR platforms, allowing every class of applications to access data needed for their functionality without requiring unrestricted access to sensors. To enable fine-tuning of permissions in real-time, Erebus provides a natural language translation module, converting users' voice commands to access control rules. We implement Erebus on Unity's AR Foundation Framework and port five existing AR applications to demonstrate Erebus's capability to secure various classes of apps. Performance results using these applications and various microbenchmarks show that Erebus achieves its security goals while being practical, introducing negligible performance overhead to the AR system.

Theory Session Presentations

Subset Node Anomaly Tracking over Large Dynamic Graphs

Xingzhi Guo

Faculty Advisor
Steven Skiena

Tracking a targeted subset of nodes in an evolving graph is important for many real-world applications. Existing methods typically focus on identifying anomalous edges or finding anomaly graph snapshots in a stream way. However, edge-oriented methods cannot quantify how individual nodes change over time while others need to maintain representations of the whole graph all the time, thus computationally inefficient.
This paper proposes DynAnom, an efficient framework to quantify the changes and localize per-node anomalies over large dynamic weighted-graphs. Thanks to recent advances in dynamic representation learning based on Personalized PageRank, DynAnom is 1) efficient: the time complexity is linear to the number of edge events and independent of node size of the input graph; 2) effective: DynAnom can successfully track topological changes reflecting real-world anomaly; 3) flexible: different type of anomaly score functions can be defined for various applications. Experiments demonstrate these properties on both benchmark graph datasets and a new large real-world dynamic graph. Specifically, an instantiation method based on DynAnom achieves the accuracy of 0.5425 compared with 0.2790, the best baseline, on the task of node-level anomaly localization while running 2.3 times faster than the baseline. We present a real-world case study and further demonstrate the usability of DynAnom for anomaly discovery over large-scale graphs.

Best Presentation: Theory

GraphZeppelin: Processing Enormous, Changing Graphs

Evan West

Faculty Advisor
Michael Bender

Finding the connected components of a graph is a fundamental problem with uses throughout computer science and engineering. The task of computing connected components becomes more difficult when graphs are very large, or when they are dynamic, meaning the edge set changes over time subject to a stream of edge insertions and deletions. A natural approach to computing the connected components on a large, dynamic graph stream is to buy enough RAM to store the entire graph. However, the requirement that the graph fit in RAM is prohibitive for very large graphs. Thus, there is an unmet need for systems that can process dense dynamic graphs, especially when those graphs are larger than available RAM.

We present a new high-performance streaming graph-processing system for computing the connected components of a graph. This system, which we call GraphZeppelin, uses new linear sketching data structures (CubeSketch) to solve the streaming connected components problem and as a result requires space asymptotically smaller than the space required for a lossless representation of the graph. GraphZeppelin is optimized for massive dense graphs: GraphZeppelin can process millions of edge updates (both insertions and deletions) per second, even when the underlying graph is far too large to fit in available RAM. As a result GraphZeppelin vastly increases the scale of graphs that can be processed.

Provable Observation Noise Robustness for Neural Network Control Systems

Veena Krish

Andrew Mata

Faculty Advisor
Amir Rahmati
Stanley Bak

Neural networks are known to be vulnerable to adversarial perturbations: slight changes to inputs that can result in unexpected outputs. In neural network control systems, these inputs are often noisy sensor readings. In such settings, natural sensor noise—or an adversary who can manipulate the environment—might be able to cause the system to fail. In this paper, we introduce the first technique to provably compute the minimum magnitude of sensor noise that can cause a neural network control system to violate a safety property from a given initial state. Our algorithm constructs a tree of possible successors with increasing amounts of noise until a specification is violated. We build upon open-loop neural network verification methods to determine the least amount of noise that could change actions at each step of a closed-loop execution. We prove that this method identifies the unsafe trajectory with the minimum magnitude of noise that leads to a safety violation. We evaluate our method on four systems: the Cart Pole and Lunar Lander environments from OpenAI gym, an aircraft collision avoidance system based on a neural network compression of ACAS Xu, and the SafeRL Aircraft Rejoin scenario. Our analysis produces unsafe trajectories where deviations under 1% of the sensor noise range make the systems behave erroneously.

Single and Multiple Secretary Selection with ML Advice

Arghya Bhattacharya

Faculty Advisor
Michael Bender,
Rezaul Chowdhury

Traditional online algorithms are designed to make decisions online in the face of uncertainty to perform well compared to the optimal offline algorithm. These algorithms are rather pessimistic in design, focusing on the worst-case inputs. On the other hand, Machine Learning (ML) algorithms try to extrapolate the pattern from past inputs to predict the future. They make decisions online based on the predictions to perform well for the average-case inputs. Recent studies have augmented several traditional online algorithms (e.g., ski-rental, competitive caching, job scheduling, etc.) with machine learning oracles to serve all possible inputs better.

In the traditional secretary problem, the algorithm's objective is to choose the best candidate for the secretary position without knowing the distribution of the qualitative values associated with the candidates. The value maximization secretary problem approaches maximizing the selected candidate's expected value. The k-secretary selection algorithm extends the problem to choosing k candidates and maximizing the expected sum of their qualitative values.

We propose an ML-advised secretary selection algorithm. The algorithm works for the k-secretary selection problem where k = 1 is a special case for the algorithm. We empirically show that for a sufficiently accurate predictor, the ML-advised algorithm performs better than its traditional online counterparts. The algorithm is robust because when we introduce error in prediction from a normal distribution, with comparatively large prediction errors, the ML-advised algorithm still outperforms the online algorithms.

Runtime-Assured Microgrid Control

Shouvik Roy

Faculty Advisor
Scott A. Smolka,
Scott D. Stoller

We present Barrier-based Simplex (Bb-Simplex), a new, provably correct design for runtime assurance of continuous dynamical systems. Bb-Simplex is centered around the Simplex Control Architecture, which consists of a high-performance advanced controller which is not guaranteed to maintain safety of the plant, a verified-safe baseline controller, and a decision module that switches control of the plant between the two controllers to ensure safety without sacrificing performance. In Bb-Simplex, Barrier certificates are used to prove that the baseline controller ensures safety. Furthermore, Bb-Simplex features a novel automated method for deriving, from the barrier certificate, the conditions for switching between the controllers. Our method is based on the Taylor expansion of the barrier certificate and yields computationally inexpensive switching conditions. We consider a significant application of Bb-Simplex to a microgrid featuring an advanced controller in the form of a neural network trained using reinforcement learning. Our results demonstrate that Bb-Simplex can automatically derive switching conditions for complex systems, the switching conditions are not overly conservative, and Bb-Simplex ensures safety even in the presence of adversarial attacks on the neural controller.

Poster Session Presentations

Aligning Large Natural Language Documents

Tanzir Pial

Faculty Advisor
Steven Skiena

Creating semantic alignment of text sequences is a research area with many use-cases in Natural Language Processing (NLP), e.g., semantic plagiarism detection, extracting relevant portions from a large database of documents using a query document, etc. A very similar problem is well-studied in bioinformatics: aligning DNA/Protein sequences and querying a database sequences with newfound sequences of the respective type. This project explores the feasibility of adapting bioinformatics alignment algorithms to the NLP domain. We design algorithms inspired by bioinformatics for semantic alignment of large text documents, both at the pair-wise level and the database level. We work with three different datasets, i) Book-to-Movie Adaptation alignment, ii) Book-to-Book alignment, and iii) Book-to-Summary alignment. We observe that dynamic programming based sequence alignment algorithms utilizing embeddings of text as a similarity metric produce strong performance for text alignment. Our methods achieve better accuracy and efficiency than previous approaches in the domain of book-movie alignment. We also adapt a heuristic algorithm inspired by BLAST for aligning query text sequence with database sequences. Finally, we create a general purpose alignment tool that anyone can use for semantically aligning texts. Along with the aligned segments we also produce statistical significance of the alignment score. This helps the end users to interpret the significance of the alignments produced.

Best Poster:
Intelligent Systems

The Effects of Color and Familiar Visual Context on Chart Comprehension and Memorability

Amit Kumar Das

Information chart design guidelines advise removing “chartjunk”—visual flourishes that are unnecessary for comprehending data. Despite this attitude towards chartjunk, many chart designers still employ it and promote its inclusion for improved memorability. Due to this disagreement, opinions on chartjunk are currently widely divided, and questions have been raised regarding its advantages and role in visualization. Our research focuses on developing a chart that can reduce chartjunk and increase memorability. To achieve this goal, we modify a minimalist chart by adding colors and recognizable context and compare it with both embellished and minimalist charts. We conducted an experiment where we assessed the interpretation accuracy immediately following the study and three days afterward. According to the findings, color combined with a recognizable context has an impact on a participant’s performance in terms of chart interpretation and memorability. Additionally, the proposed chart significantly impacted long-term recall performance while remembering the chart information.

Faculty Advisor
Klaus Mueller

Robust Mental Health Assessments in Time and Space Through Social Media Language Analyses

Siddharth Mangalik

Current standards for monitoring mental health in the U.S. population lack the kinds of temporal and geospatial resolution that are available for other conditions. For example, estimates of the prevalence of anxiety used by the National Institutes of Health are 20 years old and are not representative at the local level. Until recently, tracking mental health was difficult because we lacked an objective method for tracking symptoms, but recent advances in language-based assessments provide us with an opportunity to profile the geographic distribution and temporal changes in mental health symptoms. In the present work, we validate a pipeline to estimate mental health changes across place and over time using language based assessments. Our results show that this method are comparable at the national level to contemporaneous polling-based efforts while also providing reliable resolution down to the county-week level.

Faculty Advisor
H. Andrew Schwartz

ViTag: Online WiFi Fine Time Measurements Aided Vision-Motion Identity Association in Multi-person Environments

Bryan Bo Cao

Faculty Advisor
Shubham Jain

We present ViTag to associate user identities across multimodal data, particularly those obtained from cameras and smartphones. ViTag associates a sequence of vision tracker generated bounding boxes with Inertial Measurement Unit (IMU) data and Wi-Fi Fine Time Measurements (FTM) from smartphones. We formulate the problem as association by sequence to sequence (seq2seq) translation. In this two-step process, our system first performs cross-modal translation using a multimodal LSTM encoder-decoder network X-Translator that translates one modality to another, e.g. reconstructing IMU and FTM readings purely from camera bounding boxes. Second, an association module finds identity matches between camera and phone domains, where the translated modality is then matched with the observed data from the same modality. In contrast to existing works, our proposed approach can associate identities in multi-person scenarios where all users may be performing the same activity. Extensive experiments in real-world indoor and outdoor environments demonstrate that online association on camera and phone data (IMU and FTM) achieves an average Identity Precision Accuracy (IDP) of 88.39% on a 1 to ~3 seconds window, outperforming the state-of-the-art Vi-Fi (82.93%). Further study on modalities within the phone domain shows the FTM can improve association performance by 12.56% on average. Finally, results from our sensitivity experiments demonstrate the robustness of \system~ under different noise and environment variations.

Best Overall Poster

How To Accelerate Training Certifiably Robust Neural Networks

Pratik Vaishnavi

Faculty Advisor
Amir Rahmati

Training deep neural network classifiers that are certifiably robust against adversarial attacks is critical to ensuring the security and reliability of AI-controlled systems. Although numerous state-of-the-art certified training methods have been developed, they are computationally expensive and scale poorly with respect to both dataset and network complexity. Widespread usage of certified training is further hindered by the fact that periodic retraining is necessary to incorporate new data and network improvements. In this paper, we propose Certified Robustness Transfer (CRT), a general-purpose framework for reducing the computational overhead of any certifiably robust training method through knowledge transfer. Given a robust teacher, our framework uses a novel training loss to transfer the teacher's robustness to the student. We provide theoretical and empirical validation of CRT. Our experiments on CIFAR-10 show that CRT speeds up certified robustness training by 8× on average across three different architecture generations while achieving comparable robustness to state-of-the-art methods. We also show that CRT can scale to large-scale datasets like ImageNet.

Learning Viewpoint-Agnostic Visual Representations by Recovering Tokens in 3D Space

Jinghuan Shang

Faculty Advisor
Michael Ryoo

Humans are remarkably flexible in understanding viewpoint changes due to visual cortex supporting the perception of 3D structure. In contrast, most of the computer vision models that learn visual representation from a pool of 2D images often fail to generalize over novel camera viewpoints. Recently, the vision architectures have shifted towards convolution-free architectures, visual Transformers, which operate on tokens derived from image patches. However, these Transformers do not perform explicit operations to learn viewpoint-agnostic representation for visual understanding, as in convolutions. To this end, we propose a 3D Token Representation Layer (3DTRL) that estimates the 3D positional information of the visual tokens and leverages it for learning viewpoint-agnostic representations. The key elements of 3DTRL include a pseudo-depth estimator and a learned camera matrix to impose geometric transformations on the tokens. These enable 3DTRL to recover the 3D positional information of the tokens from 2D patches. In practice, 3DTRL is easily plugged-in into a Transformer. Our experiments demonstrate the effectiveness of 3DTRL in many vision tasks including image classification, multi- view video alignment, and action recognition. The models with 3DTRL outperform their backbone Transformers in all the tasks with minimal added computation.

Audio Analysis of Healthy Aging in World War II Veterans

Yunting Yin

Faculty Advisor
Steven Skiena

Researchers are increasingly interested in better methods for assessing the pace of aging in older adults. The present study sought to determine whether paralinguistic vocal attributes improve estimates of the risk of mortality in older adults. To measure vocal age, we curated interviews provided by male US World War II veterans in the Library of Congress collection. We used diarization to identify speakers and measure vocal features; recording data were matched to mortality information. Veterans (N=2,447) were randomly split into testing (n=1,467) and validation (n=980) subsets to generate estimations of vocal age and years of life remaining. Results were replicated to examine out-of-sample utility using Korean War veterans (N=352). WWII veterans' average age was 86.08 at time of recording and 91.28 at time of death. Overall, 7.4% were prisoners of war, 43.3% were Army veterans, and 29.3% were drafted. Vocal-age estimates were within 5 years of chronological age 78.5% of the time. With chronological age held constant, older vocal-age estimation was correlated with shorter time until death. Results imply that paralinguistic analyses might augment other assessments for individuals when oral patient histories are recorded.

Addressing Class Imbalance in Semi-supervised Image Segmentation: A Study on Cardiac MRI

Hritam Basak

Due to the imbalanced and limited data, semi-supervised medical image segmentation methods often fail to produce superior performance for some specific tailed classes. Inadequate training for those particular classes could introduce more noise to the generated pseudo labels, affecting overall learning. To alleviate this shortcoming and identify the under-performing classes, we propose maintaining a confidence array that records class-wise performance during training. A fuzzy fusion of these confidence scores is proposed to adaptively prioritize individual confidence metrics in every sample rather than traditional ensemble approaches, where a set of predefined fixed weights are assigned for all the test cases. Further, we introduce a robust class-wise sampling method and dynamic stabilization for a better training strategy. Our proposed method considers all the under-performing classes with dynamic weighting and tries to remove most of the noises during training. Upon evaluation on two cardiac MRI datasets, ACDC and MMWHS, our proposed method shows effectiveness and generalizability and outperforms several state-of-the-art methods found in the literature.

Faculty Advisor
Zhaozheng Yin

Best Poster:
Systems and Security

Debloating Web Applications

Babak AminAzad

The process of debloating, i.e., removing unnecessary code and features in software, has become an attractive proposition to managing the ever-expanding attack surface of ever-growing modern applications. Researchers have shown that debloating produces significant security improvements in a variety of application domains including operating systems, libraries, compiled software, and, more recently, web applications. Even though the client/server nature of web applications allows the same backend to serve thousands of users with diverse needs, web applications have been approached monolithically by existing debloating approaches. That is, a feature can be debloated only if none of the users of a web application requires it. Similarly, everyone gets access to the same "global" features, whether they need them or not.

Recognizing that different users need access to different features, we propose role-based debloating for web applications. In this approach, we focus on clustering users with similar usage behavior together and providing them with a custom debloated application that is tailored to their needs. Through a user study with 60 experienced web developers and administrators, we first establish that different users indeed use web applications differently. This data is then used by DBLTR, an automated pipeline for providing tailored debloating based on a user's true requirements. Next to debloating web applications, DBLTR includes a transparent content-delivery mechanism that routes authenticated users to their debloated copies.
We demonstrate that for different web applications, DBLTR can be 30-80% more effective than the state-of-the-art in debloating in removing critical vulnerabilities.

Faculty Advisor
Nick Nikiforakis

Domains Do Change Their Spots

Johnny So

Faculty Advisor
Nick Nikiforakis

When domains expire and are released to the public, adversaries can re-register them with the hope of exploiting residual trust from clients that are unaware of the change in ownership. Because domain name resolution is integral to the web, possible clients run the gamut from humans browsing the web to automated and periodic processes such as system updates. For an adversary, this trivially yields access to an attack vector that can affect a multitude of diverse systems. We reason that some domains which experience residual trust and are valuable from a security perspective are not valuable from a dropcatching perspective, and can be re-registered without participating in fiercely competitive auctions for expired domains.

In this paper, we present an investigation into this attack vector using a top-down, opportunistic approach, as opposed to bottom-up approaches used by prior work that target specific systems and infrastructure. We identify and re-register potentially valuable dropped domains using a threshold of passive DNS resolutions, and find instances of residual trust that can be exploited. Our honeypot services recorded 650,737,621 requests from 5,540,379 unique IP addresses situated in 22,744 different autonomous systems to the 201 re-registered domains in four months. Anomalous expired domains that received significantly more, or different, traffic included a torrent tracker, a computer lab statistics API, an Android haptics library API, security company DNS sinkhole servers, and command-and-control servers for malware. These findings demonstrate that adversaries with modest budgets can compromise a wide range of systems by merely registering previously-popular domains that expired.

Belief Miner: A Methodology for Causal Knowledge and Misconception Discovery from General Populations

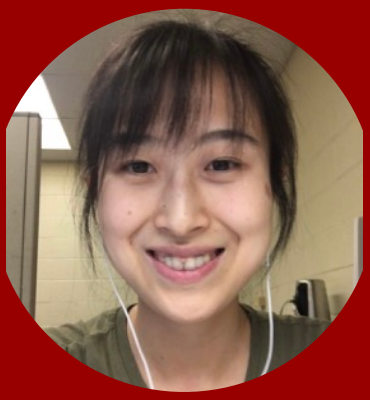
Shahreen Salim Aunti

Faculty Advisor
Klaus Mueller

Several efforts have proposed using causality as a model for collecting peoples' beliefs on a topic of interest. However, we lack a systematic evaluation method for validating causal beliefs. Such evaluation methods are imperative for contended topics such as climate change, as there are various ways misconceptions may persist in collected beliefs (e.g., illusions of correlated factors being causally related). We propose Belief Miner, a methodology for evaluating people's causal beliefs from three angles: Aggregated Evaluation, Causal Link-based Evaluation, and Potential Misconception and Obliviousness Detection. To collect the causal beliefs, we designed an interactive visual interface where people can create causal relations between factors related to climate change, e.g., wildfires, CO2, etc. A crowdsourcing experiment with 94 workers on Amazon Mechanical Turk revealed that workers could create small causal networks through our system. Analyzing these networks using Belief Miner, we discovered a variety of crowd perception and potential misconceptions.

Best Poster: Theory

A Multistep Frank-Wolfe Method



Zhaoyue Chen

The Frank-Wolfe algorithm has regained much interest in its use in structurally constrained machine learning applications. However, one major limitation of the Frank-Wolfe algorithm is the slow local convergence property due to the zig-zagging behavior. We observe that this zig-zagging phenomenon can be viewed as an artifact of discretization, as when the method is viewed as an Euler discretization of a continuous time flow, that flow does not zig-zag. For this reason, we propose multistep Frank-Wolfe variants based on discretizations of the same flow whose truncation errors decay as O(Delta^p), where p is the method's order. This strategy ``stabilizes" the method, and allows tools like line search and momentum to have more benefit. However, our results suggest that the worst case convergence rate of Runge-Kutta-type discretization schemes cannot improve upon that of the vanilla Frank-Wolfe method for a rate depending on k, suggesting this line of reasoning cannot provably accelerate the method. Still, we believe that this analysis adds to the growing knowledge of flow analysis for optimization methods, and is a cautionary tale on the ultimate usefulness of multistep methods.

Faculty Advisor
Yifan Sun

Multisecant Extensions of Quasi-Newton method

Mokhwa Lee

When dealing with a large-scale optimization problem, classical second-order methods, such as Newton's method, are no longer practical because it requires iteratively solving a large-scale linear system of order n. For this reason, Quasi-Newton(QN) methods, like BFGS or Broyden's method, are introduced because they are more efficient than Newton's method. This project focuses on multi-secant extensions of the BFGS method, to improve its Hessian approximation properties. Unfortunately, doing so sacrifices the matrix estimate's positive semi-definiteness, and steps are no longer assured to be descent directions. Therefore, we apply a perturbation strategy, inspired by the 'Haynsworth inertia additivity formula', to construct an almost-secant positive-definite Hessian estimate matrix. This strategy has a low computational cost, involving only rank-2 updates with variable and gradient successive differences. We also explore several ways of improving this method, accepting and rejecting older updates according to several nondegeneracy metrics. Future goals include extending these techniques to limited memory versions.

Faculty Advisor
Yifan Sun

Incontext Lexicon Utilization for Semantic Parsing of Software Specification

Md. Saqib Hasan

Faculty Advisor
Niranjan Balasubramanian

Training automated systems to translate natural language specifications of software documentation to logic-based formalism for verification (auto-formalization) can be a challenging task for two big reasons: lack of annotated datasets for any problem in the respective domain space; the need for domain expertise for translation and development of datasets to train statistical models through supervised learning. Furthermore, any off-the-shelf system requires domain adaptation via retraining before it can be used in a problem space that is out-of-distribution (OOD) from the system's original training domain. In order to overcome these challenges, we propose a transformer-based approach that learns to incorporate expert annotated information, presented in the form of a lexicon, during inference time by teaching it to do so via synthetically augmented datasets adapted to the domain space. In doing so, our system can solve any auto-formalization tasks in unseen domains by simply using corpus-adapted information "incontext" without the need for any expensive human annotated datasets for retraining. We aim to show the feasibility of the approach, in terms of performance and resources, in a particularly difficult problem setting: conversion of natural language requirements present in the Network File System (NFS) documentation into a domain constrained version of Signal Temporal Logic (STL). As far as we know, this is the first-of-a-kind system that trains a model to infuse previously unseen domain expert knowledge while performing on out-of-distribution domains. We are also the first to propose any such auto-formalization system for the Network File System problem setting.

Tuning Cloud Services

Gagan Somashekar

Faculty Advisor
Anshul Gandhi

Microservice architecture is an architectural style for designing applications that supports a collection of fine-grained and loosely-coupled services, called microservices, enabling independent development and deployment. Today, the configuration in each of
the software layers (microservices, scheduler, OS, etc.,) is chosen in silo, which, not surprisingly, could yield sub-optimal outcomes in terms of dollar-cost or throughput. This presents a significant opportunity for enabling the stake-holders of the cloud service make informed decisions, by coordinating the configuration choices across the layers. The problem is particularly significant for first-party workloads and applications on enterprise-scale cloud services, where (a) there is a mix of workloads with varying SLAs, and (b) it is feasible to make dynamic and coordinated choices across the layers.

This work investigates optimization algorithms to address the problem of configuration tuning of microservices applications. To address the critical issue of large state space, practical dimensionality reduction strategies are developed based on the available system characteristics. The evaluation of the optimization algorithms and dimensionality reduction techniques across three popular benchmarking applications highlights the importance of configuration tuning to reduce tail latency (by as much as 46%). Results show that the right combination of optimization algorithms and dimensionality reduction can provide substantial latency improvements by identifying the right subset of parameters to tune, reducing the search space by as much as 83%. This work also makes the case for tuning parameters across software layers showing preliminary improvements of 20% over tuning specific layers.

Contextualized Medication Event Extraction

Noushin Salek Faramarzi

Faculty Advisor
Ritwik Banerjee

To have a complete view of a patient's medication history, it is crucial to comprehend drug occurrences in clinical notes. The identification of medication changes in clinical notes has been the subject of prior research, but because clinical documentation is longitudinal and narrative in nature, the extraction of medication change information alone without the required clinical context is insufficient for use in practical applications like medication timeline generation and medication reconciliation. To bridge this gap, this track aims to capture multi-dimensional context of medication changes documented in clinical notes.

The main objective of this task is to find all drug mentions in a clinical note, to indicate if a change has been mentioned or not, and to categorize change occurrences along 5 contextual aspects. We approach the smaller tasks in this challenge in the following manner:

Medical Extraction - Employed Med7 Named- Entity Recognition model that ascertains seven concepts pertinent to medication viz. Dosage, drug names, duration, form, frequency, route of administration, and strength.

Event Classification - In this task medication mentions in clinical notes categorizes as either Disposition (medication change mentioned), NoDisposition (no change discussed), or Undetermined (need more information). To begin with, we use the drug name offsets throughout the preprocessing stage to extract the sentences that contain the drug name. Then, utilizing smart-batching schema, we fine-tune various transformer-based models.

Context Classification – In this task we classify the contextual information for Disposition events along 5 orthogonal dimensions: Action (e.g. start, stop), Negation (e.g. negated), Temporality (e.g. past, present), Certainty (e.g. hypothetical, conditional), and Actor (e.g. patient, physician). For each event classification, we employ a transformer-based model, such as BERT, trained using the smart-batching schema.

Incentivizing Rideshare Sensing Agents via Reinforcement Learning

Mohib Azam

Faculty Advisor
Susu Xu

Mobile Crowdsensing Systems (MCS) often utilize non-dedicated mobile agents (e.g., ridesharing cars) equipped with sensors to collect urban data as they move around a city. Utilizing these non- dedicated agents provides a cheap means for mobile crowdsensing. However, driver agents often focus on traveling to areas where they
can pick up more passengers, which can cause the distribution of uploaded data to mismatch our intended distribution. In this work, we propose a passive incentivization scheme in which incentives are applied to different map locations for agents who travel to that location at that time and upload data. We design a simulated envi-
ronment for multi-agent mobile crowdsensing systems based on real-world taxi data and model large amount of agents with hetero-geneous utility functions and routing behaviors, namely myopic, semi-myopic, and farsighted. With this environment, we propose using reinforcement learning to search for an ideal incentivization policy. Finally, we show through experiments that a Proximal Policy Optimization-based method is capable of learning such a policy in spite of the complexities and high dimensionality of the environment and action spaces.

Smudged Fingerprints: Characterizing the Performance of Modern Web Application Fingerprinting Techniques

Brian Kondracki

Faculty Advisor
Nick Nikiforakis

Web applications are the backbone of the modern-day Internet. Open-source web application software in particular has given everyone the ability to deploy complex web applications on their site(s), ranging from blogs and personal clouds, to server administration tools and webmail clients. Given that there exists millions of deployments of this software in the wild, the ability to fingerprint a particular release of a web application residing at a web endpoint provides attackers with the ability to trivially exploit known vulnerabilities. Conversely, defenders also rely on the ability to identify vulnerable software among their infrastructure so that they can selectively patch outdated deployments.

In this work, we study modern web application finger- printing techniques and identify their inherent strengths and weaknesses. We develop a web application testing framework, we call WASABO, and use it to measure the performance of six web application fingerprinting tools against 1,196 releases of popular web applications. To measure the robustness of each fingerprinting technique against common anti-fingerprinting measures taken by real-world websites, we deploy each web application both in their "out-of-the-box" settings, and with minor anti-fingerprinting measures taken to the content, analyzing the performance drop-off encountered.

We find that the web application fingerprinting techniques utilized today are not sufficient for real-world use. While 94.8% of all web application releases were correctly labeled by at least one fingerprinting tool in ideal conditions, only 55.3% were identified when anti-fingerprinting measures were applied to the content of each web application. Moreover, we find that even in ideal fingerprinting conditions, many tools are unable to produce a single version prediction for a particular release, leading to instances where a release is labeled as multiple disparate versions, distributed years apart. Alarm- ingly, we identify 61 instances in which a web application release that contains a severe vulnerability is predicted to be a non-vulnerable version, leading to a false sense of security for any administrator who relies on web application fingerprinting to identify vulnerable web hosts on their network.

Participant's Choice: Best Poster

Discrete Outcome Quantum Sensor Networks

Caitao Zhan

Faculty Advisor
Himanshu Gupta

We model a quantum sensor network using techniques from quantum state discrimination. The interaction between a qubit detector and the environment is described by a unitary operator, and we will assume that at most one detector does interact. The task is to determine which one does or if none do. This involves choosing an initial state of the detectors and a measurement. We consider global measurements in which all detectors are measured simultaneously. We find that an entangled initial state can improve the detection probability, but this advantage decreases as the number of detectors increases.

Using Commonsense Knowledge to Answer Why Questions

Yash Kumar Lal

Answering questions in narratives about why events happened often requires commonsense knowledge external to the text. What aspects of this knowledge are available in large language models? What aspects can be made accessible via external commonsense resources? We study these questions in the context of answering questions in the TellMeWhy dataset using COMET as a source of relevant commonsense relations. We analyze the effects of model size (T5 variants and GPT-3) along with methods of injecting knowledge (COMET) into these models. Results show that the largest models, as expected, yield substantial improvements over base models. Injecting external knowledge helps models of various sizes, but the amount of improvement decreases with larger model size. We also find that the format in which knowledge is provided is critical, and that smaller models benefit more from larger amounts of knowledge. Finally, we develop an ontology of knowledge types and analyze the relative coverage of the models across these categories.

Faculty Advisor
Niranjan Balasubramanian