Computer Science Research
@ The City University of New York

Home of the Ph.D in Computer Science
+ M.S. in Data Science

https://www.gc.cuny.edu/computer-science
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Big Data Interaction Lab</td>
<td>1-3</td>
</tr>
<tr>
<td>2. CCNY Robotics Lab</td>
<td>4-5</td>
</tr>
<tr>
<td>3. Computer Networks and Mobile Systems (NeMo)</td>
<td>6-7</td>
</tr>
<tr>
<td>4. Computer Simulation Stochastic Modeling and Optimization (CoSSMO)</td>
<td>8</td>
</tr>
<tr>
<td>5. Computer Vision &amp; Robotics Lab</td>
<td>9-10</td>
</tr>
<tr>
<td>6. CSI Computational Vision and Learning (CVL) Lab</td>
<td>11-13</td>
</tr>
<tr>
<td>7. Distributed Artificial Intelligence Research (DAIR) Lab</td>
<td>14-15</td>
</tr>
<tr>
<td>8. Distributed Network Management Lab</td>
<td>16</td>
</tr>
<tr>
<td>9. Dr. Jun Li’s Research Group</td>
<td>17</td>
</tr>
<tr>
<td>10. Dr. Saad Mneimneh’s Research Group</td>
<td>18-19</td>
</tr>
<tr>
<td>11. Hunter Speech Lab</td>
<td>20</td>
</tr>
<tr>
<td>12. Machine Learning, Data Science &amp; Informatics</td>
<td>21-22</td>
</tr>
<tr>
<td>13. Queens College Healthcare Informatics Lab</td>
<td>23</td>
</tr>
<tr>
<td>14. Research Laboratory for Logic and Computation</td>
<td>24-25</td>
</tr>
<tr>
<td>15. Smart Cities and Urban Computing (SCUC) Lab</td>
<td>26-27</td>
</tr>
<tr>
<td>16. City College Visual Computing Lab (CCVCL)</td>
<td>28-29</td>
</tr>
<tr>
<td>17. Computer &amp; Network Security Lab</td>
<td>30-31</td>
</tr>
<tr>
<td>18. CUNY Ponder</td>
<td>32-33</td>
</tr>
<tr>
<td>19. Distributed Systems and Security (DiSSect) Lab</td>
<td>34</td>
</tr>
<tr>
<td>20. Machine Learning Lab</td>
<td>35</td>
</tr>
<tr>
<td>21. Topological Data Analysis</td>
<td>36-37</td>
</tr>
<tr>
<td>22. Underwater Network and Intelligent System (UNIS) Lab</td>
<td>38-40</td>
</tr>
</tbody>
</table>
Big Data Interaction Lab (BDI Lab)

Our research focuses on developing efficient algorithms and interactive systems that enable massive data analysis and visualization at scale. Data-driven exploration is a complex, time-consuming process that often requires techniques from multiple disciplines including statistics, machine learning, data management, and visualization. Tools must be built in such a way that they are usable and within reach to domain experts who often lack computer science expertise. Visual analytics, which couples interactive visual representations with analytical processes, aims to tackle this problem in particular.

For big data, the problem is even more challenging due to the fact that data manipulation and analysis techniques must scale to large and complex datasets. However, there is a gap between big data management techniques and how they may be used to empower visual analytics applications. Bridging this gap and providing an end-to-end solution to data-driven exploration is a main focus of our lab.

Our work is supported by NFS, DARPAR, Alfred P. Sloan Foundation, and industrial partners.

Want more information?
Visit BDI Lab's website.
Towards Building Scalable Applications for Urban Visual Analytics using Game Engines

The objective of this project is to allow the creation of visual analysis applications, particularly for those that need 3D digital twins, to be more accessible to data enthusiasts and domain experts, in a similar way that game engines are made available to independent developers. Simply put, if a game designer can create a game in a week, the same person should be able to help building a 3D urban application to support decision making in the same amount of time, given that she is provided with a capable tool and framework. As such, a team consisting of game designers and domain experts would be sufficiently equipped to creating an urban application, while allowing visualization experts to focus more on visualization tasks where they are needed the most. Game engines have only been used for building visualization on virtual or mixed reality devices (per the requirements on the devices). In this project, we aim to equip game engines with more data-driven capabilities through plugin mechanism. This will enable us to create an urban analysis framework that can seamlessly integrate urban data with semantic city models, interactive analysis, and a 3D visualization environment.
Distributed Partitioning and Processing of Large Spatial Datasets

As a result of the tremendous growth in data collection, particularly due to the rise of IoT devices and urban digitization, higher demand for new techniques emerged to efficiently process and extract valuable insights within a relatively acceptable time frame. Additionally, much of the collected data has spatial components which makes it even more challenging to process. The current standard approach to large-scale data analysis uses distributed parallel processing systems like Apache Hadoop and Apache Spark. However, these systems are designed for general-purpose parallel processing and require an additional layer to recognize and efficiently process spatial datasets. In this project, we investigate novel strategies to partition massive spatial datasets for use with spatial queries like map-matching and kNN spatial join. Additionally, we also propose an accuracy benchmarking framework for classifying the results of two input files to validate the accuracy of a technique. Our proposed work targets batch processing of large spatial datasets, including structured, unstructured, and semi-structured datasets, and prepare them to be consumed interactively later in the analytic pipeline.

Progressive Interactive Visualization of Massive Spatio-Temporal Textual Data

A large share of the massive available unstructured data are generated by social media, news, and forums, such as Twitter, Reddit, Stack Overflow. Besides texts, these data also contain spatial and temporal information. If mined correctly, spatio-temporal information can provide invaluable insights to better understand social activities. Being able to create a real-time visualization to quickly characterize activities in an area during a specified time range is of great interest to domain scientists and is useful for both building and evaluating analysis models. However, it is not yet possible with existing systems. We propose an out-of-core framework for tag cloud visualization that can quickly return a close approximation of the result while improving progressively. To achieve this, our framework rearranges the input data in a spatio-temporal coherent layout and adds a small sketching structure to help bootstrap the computation process. At the same time, our method also allows users to specify spatio-temporal query constraints to customize their tag cloud visualization. The outcome is an interactive system that returns approximate results with less than 20% error rate on average in the first second for most datasets. Our experimental results on real-world datasets show that the proposed technique can efficiently compute tag cloud visualizations with a minimal requirement in time and storage for preprocessing; and has a small memory footprint and can work with datasets that are much larger than the available memory.
CCNY Robotics Lab

The Robotics and Intelligent Systems Lab (CCNY Robotics Lab) is a research laboratory directed by Dr. John (Jizhong) Xiao, in the Electrical Engineering Department of The City College of New York.

The research activities in this lab emphasize system modeling and analysis, embedded system design, control theory and applications, sensor integration and data fusion, computer vision, real-time and embedded computing, intelligent mechatronics systems, 3D mapping, autonomous navigation, and 3D SLAM (simultaneous localization and mapping). Specific initiatives underway include robotic inspection of infrastructure and renewable energy (solar panels and wind turbine blades), non-destructive testing/evaluation (NDT/NDE) to reveal surface flaws and subsurface defects of concrete structure, design and control of climbing robots, autonomous navigation of MAVs (micro aerial vehicles), assistive navigation for visually impaired people, internet-based telerobotics, mobile sensor networks, multi-robot exploration and swarm robotics. The problems studied here involve both theoretical development in new methodologies for control, planning, coordination and machine learning, as well as implementation issues in sensory measurement, hardware design, and computer software development.

The laboratory continues to receive support from the National Science Foundation (NSF), Army Research Office (ARO), Department of Transportation (DOT), Federal Highway Administration (FHWA), National Collegiate Inventors & Innovators Alliance (NCIIA), CUNY research award programs, and industries for its unique effort in theoretical development and experimental validation.

Faculty

John (Jizhong) Xiao
CCNY

Want more information?
Visit CCNY Robotics Lab’s website.

Students

Yang He
Ali Salem
Ejup Hoxha
Jingjun Feng
Autonomous Wall-climbing Robots for Inspection and Maintenance of Concrete Bridges

**PROJECT DESCRIPTION**

The project aims to develop novel wall-climbing robots for inspection and maintenance of concrete bridges. The robots are designed to autonomously navigate on inclined surfaces, climb, and perform tasks such as crack detection, leakage identification, and surface defect assessment. The robots are equipped with advanced sensors and actuators to ensure safe and efficient operation. The project involves interdisciplinary research at the intersection of robotics, computer vision, and civil engineering.

**METHODS**

The project employs a multidisciplinary approach, combining expertise from robotics, civil engineering, and computer science. Key components include:

1. **Sensors and Actuators**: Robotic arms, cameras, and other sensors are integrated to enable autonomous navigation and inspection.
2. **Control Systems**: Advanced control algorithms are developed for precise movement and task-specific operations.
3. **Data Processing**: Image processing and machine learning techniques are used to identify and analyze defects in real-time.

**RESULTS**

Initial prototypes have demonstrated the feasibility of the approach, with successful navigation and inspection in simulated environments. Further testing is required to validate performance in real-world scenarios.

**CONCLUSIONS**

The project has potential to revolutionize bridge inspection and maintenance, offering a safer and more efficient alternative to manual inspections. Future work includes the integration of additional sensors and the development of more robust control algorithms.

---

**ISANA: Intelligent Situation Awareness and Navigation Aid for Visually Impaired People**

**Motivation**

- 220 million people are estimated to be visually impaired worldwide.
- Mobility and navigation are significant challenges for visually impaired individuals.

**Goal**

- To develop a system that provides real-time situation awareness and navigation assistance to visually impaired individuals.

**Approach**

- Integration of sensors and actuators to provide real-time situational awareness.
- Development of an intuitive user interface for seamless interaction.

**System Overview**

- **Motivation** and **Goal**
- **Approach**
- **Experimental Setup**
- **ISANA Prototype**
- **Conclusion**

---

**A Field Deployable Wall-Climbing Robot for Bridge Inspection using Vision and Impact Sounding Techniques**

**PROJECT DESCRIPTION**

This project focuses on developing a field-deployable wall-climbing robot for bridge inspection. The robot integrates advanced visual and impact sounding technologies to assess the structural integrity of bridges effectively. The system is designed to navigate along bridge walls, detect cracks, and estimate their severity, providing critical data for maintenance planning.

**METHODS**

- **Vision-based Inspection**: Robotic arms equipped with cameras capture high-resolution images to identify structural anomalies.
- **Impact Sounding**: Piezoelectric transducers emit acoustic signals to assess the integrity of the bridge's material.

**RESULTS**

Preliminary tests have shown promising results, with accurate detection rates for cracks and other defects. Further field trials are planned to validate the system's performance.

**CONCLUSIONS**

The project demonstrates the potential of integrating visual and impact sounding technologies for efficient bridge inspection. Further development is needed to refine the system's accuracy and reliability in various environmental conditions.

---

**ACKNOWLEDGMENTS**

The authors acknowledge the support of various stakeholders, including funding agencies, industry partners, and academic institutions. Special thanks to the team members who contributed to the project's success.
Computer Networks and Mobile Systems (NeMo)

The Computer Networks and Mobil Systems (NeMo) lab of the City University of New York (CUNY) is home based at John Jay College of Criminal Justice and include faculty and students who are affiliated with both CUNY’s John Jay College and the Graduate Center. The NeMo Lab was established by Prof. Ping Ji, and has conducted research activities in various fields of Computer Networking, mostly with Wireless and Mobile Network Measurement, Internet of Things, Mobile Systems Analysis, and Network Security & Forensics. NeMo lab has published research papers on the topics of: mobile device privacy, activity recognition with smartphone sensors, multi-model sensor data analysis, sensor network coverage efficiency, and many more.

Facets: Fast Comprehensive Mining of Co-evolving High-order Time Series

Ubiquitous Coevolving* Time Series

- a) Room condition monitoring in a smart building
- b) Intelligent transportation systems
- c) Stock Market
- d) Climate Monitoring

Mining time series data has been a very active research area, exactly because of its prevalence in many high-impact applications. It has posed many fascinating research questions. Among others, three prominent challenges shared by a variety of real applications are (a) high-order; (b) contextual constraints and (c) temporal smoothness. The state-of-the-art mining algorithms are rich in addressing each of these challenges, but relatively short of comprehensiveness in attacking the coexistence of multiple or even all of these three challenges. In this project, we propose a comprehensive method, Facets, to simultaneously model all these three challenges. We formulate it as an optimization problem from a dynamic graphical model perspective. The key idea is to use tensor factorization to address multi-aspect challenges, and perform careful regularization to attack both contextual and temporal challenges. Based on that, we propose an effective and scalable algorithm to solve the problem. Our experimental evaluations on three real datasets demonstrate that our method (1) outperforms its competitors in two common data mining tasks (imputation and prediction); and (2) enjoys a linear scalability w.r.t. the length of time series. A paper coming out of this work was published in KDD2015.
Activity Recognition with Smartphone Sensors

The ubiquity of smartphones together with their ever-growing computing, networking, and sensing powers have been changing the landscape of people’s daily life. Among others, activity recognition, which takes the raw sensor reading as inputs and predicts a user’s motion activity, has become an active research area in recent years. It is the core building block in many high-impact applications, ranging from health and fitness monitoring, personal bio-metric signature, urban computing, assistive technology, and elder-care, to indoor localization and navigation, etc. This paper presents a comprehensive survey of the recent advances in activity recognition with smartphones’ sensors. We start with the basic concepts such as sensors, activity types, etc. We review the core data mining techniques behind the main stream activity recognition algorithms, analyze their major challenges, and introduce a variety of real applications enabled by activity recognition.

The Privacy Leaky Bucket: Your Mobile Devices

Booming wireless technology makes personal mobile devices, such as smartphones, tablets, wearable devices, must-have gadgets in peoples' lives. Along with the broad deployment of WiFi networks, mobile devices can access the Internet not only through cell phone networks, but more and more frequently through WiFi (802.11x) channels. Since WiFi traffic is propagated through open wireless spectrum, it is relatively easy to monitor the device communication activities in such networks. The goal of this project thus is to assess how much personal information can be leaked (or can be extrapolated) from systematically collected open-in-air message exchanges among mobile devices. We monitor the wireless traffic and apply data mining technique to answer the question of whether and how much personal information can be observed and inferred through monitoring the net work activities conducted by mobile devices, and small devices of Internet of Things, in WiFi networks.
Computer Simulation
Stochastic Modeling and Optimization (CoSSMO)

The Institute for Computer Simulation, Stochastic Modeling and Optimization (CoSSMO) maintains a portfolio of research projects dealing with computer simulation, stochastic modeling and optimization. We focus on the breadth of applications of stochastic modeling and optimization that are a direct benefit to government and private industry.

The academic disciplines within its scope include: mathematics, computer science, statistics, applied mathematics, geography, business, economics and engineering.

Faculty

Felisa Vazquez-Abad
Hunter College

Ted Brown
Queens College

Saad Mneimneh
Hunter College

Students

Alexei Nikolaev
Syed Ali Ahmed
Saman Farhat
Thomas Flynn
Alexandra Diamond
Silvano Bernabel

Rescue Robots

The multifactorial complexity of clinical data complicates prediction and prevention of undesired outcomes. This project aims to investigate the value of more advanced machine learning methods by simultaneously considering all the factors, to develop better predictive and prevention methods. Our approach to preterm birth prevention brings to bear stochastic methods to derive accurate, multidimensional prevention models from large collections of observational data. These methods will support prevention relative to different stages of pregnancy.

This is collaborative work with Collaborators at Columbia University and CUMC funded by the NIH. The team recently won a 2021 National Institutes of Health’s Decoding Maternal Morbidity Challenge Prize for presenting a new methodology that identified patients with a high risk of preeclampsia early in pregnancy.

Want more information?
Visit CoSSMO’s website.
Our research is within the broad area of 3D processing, reconstruction, classification and detection from range and image data. In the past, we were interested in 3D photorealistic modeling of urban areas from terrestrial Lidar and color images, as well as urban object classification (for example detection of facades, windows, etc.). Currently, we are working on 3D object detection applying deep learning techniques, with applications in robotic perception, manipulation and planning.

Faculty

Ioannis Stamos
Hunter College

Students

Zhujun Li
Youngsu Jung
Jaime Canizales

Want more information? Visit Computer Vision Lab’s website.
3D Object Detection and Instance Segmentation

Instance segmentation and object detection are significant problems in the fields of computer vision and robotics. They are important parts for robotic perception as well robotic grasping applications. We address those problems by proposing a novel object segmentation and detection system. First, we detect 2D objects based on RGB, depth only, or RGB-D images. A 3D convolutional-based system, named Frustum VoxNet, has been developed. This system generates frustums from 2D detection results, proposes 3D candidate voxelized images for each frustum, and uses a 3D convolutional neural network (CNN) based on these candidate voxelized images to perform the 3D instance segmentation and object detection. Results on the SUN RGB-D dataset show that our RGB-D-based system’s 3D inference is much faster than state-of-the-art methods, without a significant loss of accuracy. At the same time, we can provide segmentation and detection results using depth only images, with accuracy comparable to RGB-D-based systems.

Real-time perception on a robotic platform

We have been demonstrating various robotic perception algorithms on our lab robots (Fetch). Based on the ROS architecture we have integrated navigation, localization and perception modules. Our robot can detect objects in real-time using novel 3D detection algorithms that utilize 3D and 2D sensing. The detection can be performed on the robot, or at a more powerful GPU-based computer. In our scenario, the robot automatically moves towards detected objects of interest. We are currently working on 6D pose detection algorithms in order to demonstrate detecting, picking and manipulating household objects.
Our research addresses computational vision's theoretical foundations and practical applications to science, biomedicine, art, consumer products, entertainment, business, manufacturing, and defense. We are focused on discovering and proposing the fundamental principles, algorithms, and implementations for solving high-level visual perception and cognition problems involving computational geometry, automated image and video analysis, and visual reasoning.

The primary goal of our research is
1. To give the computational devices the ability to "see, learn, and understand of the real world "just as humans do.
2. To develop new sensing instruments (portable 3D microscopes and multi-purpose imaging sensors
3. To make more effective the computational device and human collaborations.
4. To assist in the process of commercializing a discovery.

Our highly interdisciplinary research integrates theories and techniques from mathematics and computational neuroscience, psychology, physics, statistics, statistics, computer vision, machine learning, and computational cameras, allowing Agaian to investigate these research areas thoroughly and efficiently teach multidisciplinary courses (including AI, medical imaging, computational vision and image processing, etc.). CVL’s current areas of active study include Computer Vision and Visual Analytics/Understanding, Multimedia Data Mining, Analytics, and Security, Autonomous Robots and Intelligent Machines, Image Quality Assessments, Machine Learning and Pattern Recognition, Multi-modal Perception and Digital Imaging, and Biomedical Image Processing and Medical Applications. This has been possible thanks to the support of DARPA, NSF, US Army, Air Force, NJU (including CRADA), and Industry, and in collaboration with multiple research labs worldwide, through multi-university research programs. Alums of the Lab have moved on to academic positions at Tuts, MIT, Johns Hopkins, Tufts University, UTSA, China, India, Armenia, UTSA, Tampere Institute of Technology, etc., or to leading positions at industrial research labs at Google, Facebook, Amazon, Intel, Sysco, Boeing, Raytheon, Nokia, Southwest Research Institute (SwRI), DARPA, National Instrument, Lincoln Laboratories, Air Force, the Central Intelligence Agency, etc.). Some of them have become a Fellow SPIE, a Fellow IEEE, the National Academy of Inventors Fellows, or have started tech companies.
Recent projects

Tiny Vision Transformer and Deep Learning Hybrid Models with Applications

Machine learning has recently become a widespread technology because of its applicability in various fields and capacity to create automatic models that handle large amounts of currently collected data. When applied to larger datasets, deep learning has moved to more complex models with higher accuracy. However, it increases computing power/memory and requires training more parameters with bigger models on larger datasets. This project focuses on deploying and applying lightweight, high-efficiency tiny Vision Transformer and CNN hybrid models for resource-limited devices with high inference speed and acceptable accuracy by combining quantization and signal processing techniques.

Tracking Wildlife via Infrared equipped UAVs

Wildlife trading is a significant cause of species loss, where wild animals are widely commodified and subjected to unsustainable harvesting practices that lead to precipitous global population declines. Using Unmanned Aerial Vehicles (UAVs) offers a cost-effective opportunity to monitor wildlife to estimate population size and protect them from loss due to poaching. The UAVs are usually equipped with cameras or other sensors, conferring a specified field of view beneath the UAV. Two challenges continuous surveillance faces are automatic detection of animal herds and poachers on live video feeds and finding an optimal UAV flight path to cover the entire area of interest through a process known as Coverage Path Planning (CPP).

This project focuses on the former challenge: (i) Identify animal herds and poachers in live video streams, (ii) Track and predict paths of identified objects, and (iii) Develop a poaching risk alert model based on the collected information. Our approach combines image feature extraction, motion detection, and analysis as input into various models trained to identify objects of interest and their type. Generated paths are input into various secondary models to predict future locations and assess the risk of animal/poacher colocation.
Digitized Painting Art Analytics with a New Portrait Dataset

The increasing availability of extensive digitized art collections opens new research directions. In particular, correctly identifying the artistic style or portrait emotions/sentiment analysis or art movement of paintings is essential for painter authentication, art movement classification, large artistic database indexing, and mobile recognition of painters. Also, facial emotion recognition (FER) is an essential topic in computer vision and artificial intelligence due to its significant academic and commercial potential. Instead of focusing on classifying and recognizing natural facial emotions, we plan to develop AI tools to classify and recognize art portrait emotions. This work is more complex than emotional "natural" faces since portraits have a wider variation in color and texture.

The project will (i) investigate art portrait emotions/sentiment analysis, (ii) develop art movement classification using portrait paintings, (iii) generate engineering and CNN features to assist different tasks in project related research, and (iv) introduce datasets related to portrait faces of the well-known artist.
The Distributed Artificial Intelligence Research Lab directed by Professor Anita Raja, is concerned with the design and development of intelligent single and multi-agent systems. Lab members conduct research in distributed computing, convention formation, cascading risks, clinical informatics, monitoring and control of computation, complex networks, machine learning, resource-bounded reasoning, and reasoning under uncertainty.

Faculty

Anita Raja  
Hunter College

Students

Arezoo Bybordi  
Nan Jia  
Alisa Leshchenko

Eric Li  
Daniel Mallia  
Matthew Carr

Want more information? Visit DAIR Lab’s website.

Advanced Machine Learning for Clinical Informatics

The multifactorial complexity of clinical data complicates prediction and prevention of undesired outcomes. This project aims to investigate the value of more advanced machine learning methods by simultaneously considering all the factors, to develop better predictive and prevention methods. Our approach to preterm birth prevention brings to bear stochastic methods to derive accurate, multidimensional prevention models from large collections of observational data. These methods will support prevention relative to different stages of pregnancy.

This is collaborative work with Collaborators at Columbia University and CUMC funded by the NIH. The team recently won a 2021 National Institutes of Health’s Decoding Maternal Morbidity Challenge Prize for presenting a new methodology that identified patients with a high risk of preeclampsia early in pregnancy.
As cities across the globe continue to grow, traffic congestion has become globally ubiquitous with great economic and environmental costs associated with it. The increasing prevalence of self-driving vehicles creates an opportunity to build smart, responsive traffic infrastructure of the future. Such an infrastructure consisting of connected and autonomous vehicles and smart traffic lights would have the potential to cope with congestion, weather phenomena and accidents, while maintaining safety and ensuring privacy of information. We argue that an approach that leverages multiagent meta-level control (MMLC) to address the challenge of dynamically adjusting traffic to the changes in the environment leads to improvement in travel times as well as decrease in emissions in mixed traffic simulation environments. We develop distributed algorithms to enable connected vehicles to overcome the traffic congestion problem efficiently and cost-effectively.

**Software Engineering for Machine Learning**

Machine Learning (ML), including Deep Learning (DL), systems, i.e., those with ML capabilities, are pervasive in today’s data-driven society. Such systems are complex; they are comprised of ML models and many subsystems that support learning processes. As with other complex systems, ML systems are prone to classic technical debt issues, especially when such systems are long-lived, but they also exhibit debt specific to these systems. Unfortunately, there is a gap of knowledge in how ML systems actually evolve and are maintained. Our recent work indicates that developers refactor these systems for a variety of reasons, both specific and tangential to ML, some refactorings correspond to established technical debt categories, while others do not, and code duplication is a major cross-cutting theme that particularly involved ML configuration and model code, which was also the most refactored. We also introduce new ML-specific refactorings and technical debt categories, respectively, and put forth several recommendations, best practices, and anti-patterns. The results can potentially assist practitioners, tool developers, and educators in facilitating long-term ML system usefulness. This is a collaborative project with Professor Raffi Khatadourian and the Hunter College Ponder Lab funded by the NSF.
Distributed Network Management Lab

The main research direction is on distributed management and control of complex networked systems. Specific areas of research focus are adaptive fault-tolerance & QoS, autonomic network resource management, declarative networks, QoS auditing in clouds, and cyber-physical network systems. A common thread in these research activities pertains to the design and management of trustable computer network systems. Specific research activities are listed below.

Assessment & certification of large-scale networked systems

Network assessment & certification is beneficial in two ways. First, it offers the means to measure the goodness of a complex network system in domain-specific metric spaces and compare it with other competing systems. For instance, a military commander deploying different network systems in a theater of operation can use the numerical scores to reason about the overall effectiveness of the combined system under various external conditions. Second, the certification enables an autonomic controller to improve upon the workflow processes and algorithms embodied in the system to deal with the environment conditions in a better way. The domain-independent nature of the certification methods developed in this research make them employable in diverse application domains: which lowers the software development costs of complex network systems. We are currently studying the certification methods in the domains of content distribution networks (CDN), replicated web services, and adaptive video transport.

QoS auditing in cloud-based distributed services

As a case study of QoS auditing methods, we work on the measurement of available bandwidth estimation on an end-to-end path set up from a client device (e.g., smart-phone) to a cloud data center over a series of routers. This estimate is then reasoned about in light of the SLA negotiated with the cloud provider at the time of path set up. In the future, we plan to employ OpenFlow switches and PlanetLab nodes as part of our experimental platform.
Dr. Jun Li's Research Group

Coding theory plays a critical role in many systems, such as communications, distributed storage, and distributed computing, to tolerate various forms of system noises. Our group focuses on theoretical and system research related to coding-based techniques in various applications, especially distributed machine learning.

Our research seeks to maximize the performance of training of machine learning models in the infrastructure with stragglers. Our work has been supported by NSF, Amazon, and Google.

**Faculty**

Julian Li  
Queens College

**Students**

Xiaodi Fan  
Yuchun Zou  
Xian Su

**Lightweight Projective Derivative Codes for Compressed Asynchronous Gradient Descent**

Coded distributed computation has become common practice for performing gradient descent on large datasets to mitigate stragglers and other faults. This project proposes a novel algorithm that encodes the partial derivatives themselves and furthermore optimizes the codes by performing lossy compression on the derivative codewords by maximizing the information contained in the codewords while minimizing the information between the codewords. The utility of this application of coding theory is a geometrical consequence of the observed fact in optimization research that noise is tolerable, sometimes even helpful, in gradient-descent-based learning algorithms since it helps avoid overfitting and local minima. This stands in contrast with much current conventional work on distributed coded computation which focuses on recovering all of the data from the workers. A second further contribution is that the low-weight nature of the coding scheme allows for asynchronous gradient updates since the code can be iteratively decoded; i.e., a worker’s task can immediately be updated into the larger gradient.

**Rook Coding for Batch Matrix Multiplication**

Matrix multiplication is a fundamental building block in various distributed computing algorithms. In order to multiply large matrices, it is common practice to distribute the computation into multiple tasks running on different nodes. In order to tolerate stragglers among such nodes, various coding schemes have been proposed by adding additional coded tasks. However, most existing coding schemes for matrix multiplication are constructed for only one matrix multiplication, while batch matrix multiplication is common in large-scale distributed computing workloads. In this project, we design Rook Coding (RC), a novel polynomial-based coding framework for computing the multiplication of $n$ pairs of matrices in batch. Designed to achieve lower encoding time in practice, we construct RC as polynomials of much simpler forms than existing coding schemes for batch matrix multiplication, achieving a recovery threshold of $O(n^{\frac{1}{2}} \log^2 n)$. Compared to existing coding schemes, RC achieves a lower encoding complexity in practice, because of its simpler forms in the encoding polynomials. Through extensive experiments, we show that RC can save the time of the whole job thanks to its low overhead of encoding.

Want more information? Visit [Prof. Li’s research page](#).
Dr. Saad Mneimneh's Research Group

Prof. Mneimneh’s research focus is in algorithms and the mathematical modeling of various physical and biological systems with applications to networks, graphs, optimization and approximation algorithms, computational biology, probabilistic analysis, and discrete mathematics. Some of his current research includes (but is not limited to) the following topics:

- Networks: Switching and load balancing in communication networks, and mathematical modeling of networks, e.g. preferential attachment in hypergraphs and simplicial complexes.
- Biology: Optimization problems inspired by RNA interactions and mathematical models for hydrophobicity in proteins.
- Computational geometry: Effect of random sampling on Frechet distance, finding monotonicity (almost straight lines) in walks.
- Probabilistic analysis: Power law in affiliation networks, Blindsort (a sorting algorithm that is not allowed to read) and random formations of islands (and pools) on a two dimensional surface.

Some of these topics can also provide rich material for undergraduate student engagement through small projects. In particular, several aspects related to blindsort, Skolem sequences, and the Tower of Hanoi make interesting simulation and analysis projects.

Faculty

- Saad Mneimneh
  Hunter College

Students

- Syed Ali Ahmad
- Alexey Nikolaev
- Saman Farhat

Want more information?
Visit Prof. Mneimneh's website, or email him at saad@hunter.cuny.edu
Preferential Attachment in Affiliation Networks

Preferential attachment describes a variety of graph-based models in which a network grows incrementally via the sequential addition of new nodes and edges, and where existing nodes acquire new neighbors at a rate proportional to their degree. Some networks are better described as groups of nodes rather than a set of pairwise connections. These groups are called affiliations, and the corresponding networks affiliation networks. When viewed as graphs, affiliation networks do not necessarily exhibit the power law distribution of node degrees that is typically associated with preferential attachment. They do, however, when viewed as hypergraphs and simplicial complexes (a special case of hypergraphs), suggesting an underlying evolutionary preferential attachment mechanism not at the graph level. Below we show a Facebook friendship network viewed as a graph and as a simplicial complex.

Multiple RNA Interaction

One of our research topics in computational biology is on the interaction of RNA molecules. The goal is to find an optimal structure resulting from the folding of the individual RNAs on the one hand, and the simultaneous binding of the RNAs on the other hand, resulting in a RNA complex. The traditional approach was to concatenate all RNAs into one and use folding to simulate the interaction. This approach cannot produce a structural form known as a kissing loop. Our original formulation of the problem was for two RNAs, which led to the introduction of RNA-RNA interaction graphs. In these graphs, the optimal structure is given by a set of non-intersecting edges of maximum weight, an NP-hard problem for which we have a 2/3 factor approximation algorithm. Our notion of an "entangler" (recursive entanglers shown below) in the RNA-RNA interaction graph is instrumental in characterizing the sub-optimality of the predicted RNA complex.

We proved that every entangler-free solution for the RNA interaction problem is at most a 2/3 factor approximation of the optimal solution. We then pioneered a formulation for the multiple (two or more) RNA interaction problem. Multiple RNA interaction poses new computational problems which also fall under the realm of combinatorial optimization and NP-hardness. We developed approximation algorithms and, more generally, algorithms that belong to a class known as PTAS (Polynomial Time Approximation Scheme) in computer science. Below is a example output for four RNA sequences showing the interactions in yellow.
Hunter Speech Lab

The Hunter Speech Lab provides an environment for pursuing research in spoken language processing. Our research is aimed at detecting paralinguistic information from speech and language, including deception, trust, mental health, personality, and radicalization. Our research is at the intersection of computer science and linguistics and our lab includes students from both disciplines.

**Faculty**

![Sarah Ita Levitan](image)

Sarah Ita Levitan  
Hunter College

**Students**

![Yuwen Yu](image)

Yuwen Yu

**Automatic Detection and Analysis of Extremist Video Content on Social Media using Text, Speech, and Visual Features**

We are studying the influence of videos on social media in persuading viewers to adopt extremist beliefs. Our goal is to identify, collect, and analyze online radicalizing videos and develop machine learning models to predict the effects of such videos on their viewers. In addition to identifying aspects of videos which appear to lead to radicalization, we explore additional questions, including: What are the characteristics of radical video material? Can we measure the effectiveness of different materials? Can we track changes over time in the material that appears to influence viewers and automatically predict the effects of future videos? This project is currently funded by the Air Force Office of Scientific Research.

**Linguistic, Demographic, and Personality Factors in Trust Perception**

We are exploring the role of lexical content, acoustic-prosodic features, demographics and personality in trust perception. We are building a trust corpus based on a set of interview dialogue segments that people rate on different speaker traits, accompanied by demographics and personality profiles for each participant. We study the effects of modality (text vs. speech), conversational context, linguistic features, and rater characteristics on the perception of trustworthy and untrustworthy speech. trusted and mistrusted speech.

**Want more information?**

Visit Prof. Levitan's website.
Machine Learning, Data Science, & Bioinformatics

Data Science, Machine Learning & Bioinformatics Lab at Hunter College, CUNY develops new data mining and artificial intelligent algorithms for the analysis and predictive modeling of complex networks, high-dimensional data, 3-dimensional objects, and biological sequences.

In collaboration with experimental and clinical scientists, we apply state-of-the-art big data analytics, machine learning, and computational simulation to drug discovery and precision medicine with the focus on drug-resistant bacterial/virus infections, Alzheimer's disease, and cancers.

Our work is supported by NSF, NIH, and pharmaceutical and biotechnology companies.

Want more information? Visit Prof. Xie's website.

Faculty

Lei Xie
Hunter College

Students

Shuo Zhang  Yue Qiu  Li Xie  Ryan Tan  Tian Cai

Yoyo Wu  Joshua Rollins  Amitesh Badkul
Geometric deep learning

Multiple fields such as chemistry, biology, physics, semantic web, and computer graphic need to deal with data in non-Euclidean spaces (e.g. graph). Geometric deep learning is a fast-growing field that extends the deep learning techniques to the graph-structured data. Currently, we are developing robust and universal graph convolutional network for 3-dimensional object representation and classification, and applying it to designing new functional materials and therapeutic drugs.

Heterogeneous multi-layered network

A multitude of real-world problems such as infrastructure network and ecological system can be formulated as a heterogeneous multi-layered network (HMLN), in which each layer represents a kind of entities and their relations and the cross-layer linkage encodes the interaction or dependency between different entities. An interesting problem is to predict the cross-layer relationships such as drug-disease-side effect associations given a complex and large HMLN. We are developing scalable data mining methods to address this problem.

Self-supervised and semi-supervised learning

The success of deep learning often relies on the availability of a large amount of high-quality labeled data. Taking advantage of a large amount of unlabeled data, we are developing new algorithms for self-supervised and semi-supervised learning of sequence, graph, and high-dimensional data, and applying them to drug discovery and precision medicines for neurogenerative diseases, cancers, and infectious diseases.

Multi-level predictive modeling and causal discovery

Multi-level hierarchy organization is the characteristics of many complex systems. For example, human trait (e.g. eye color) or disease (e.g. cancer) is the manifest of composite, distribution, and interaction of proteins, which depends on the expression level of RNAs and interplays between them. In turn, the expression of RNA is determined by the markup of DNAs. It is a fundamental and challenging problem to predict the top-level properties from the bottom-level features, and establish their causal relationships. We are developing and synthesizing a plethora of machine learning techniques such as transfer learning, multi-view learning, and contrastive learning to address this fascinating problem. An immediate application of this work is the development of individualized anti-cancer therapy.
Our research lab primarily focuses on biomedical imaging informatics, which involves development of algorithms for understanding of image contents, fusion of information extracted across multiple images, and large-scale analysis of such information. It spans the areas of computer vision, image processing, and machine learning. Not limited to just images, we also leverage other information modalities to include structural information in ensemble learning. We actively seek collaborations with experts in nature science to explore applications of big data in medicine and health care. Our research should generate greater scientific insights into hypotheses, not answered by traditional small data studies, while improving state-of-the-art machine learning models. Our PhD students work actively with collaborators from problem formulation to publication.

We currently conduct three research projects, and we welcome more PhD students to join our lab:

1. Weakly supervised segmentation of Optical Coherence Tomography (OCT) Imaging Biomarkers in Retinal Disease. This is a collaboration with NYU Langone Eye Center.

2. Machine learning on the neurobehavioral development of children affected by in-utero maternal stress due to natural disasters induced by climate change. This is a collaboration with Queens College, Psychology Department.

3. Machine learning on longitudinal data beyond newborn intensive care unit (NICU) for predicting developmental delay in pre-term infants. This is a collaboration with Infant Development Research Program, NYS Institute for Basic Research in Developmental Disabilities.

Want more information?
Email ctsai@gc.cuny.edu
Research Laboratory for Logic and Computation

We develop new logical systems that target foundations of Logic in Computer Science and applications in such fields as argumentation theory, artificial intelligence, belief revision, constructive semantics and provability semantics, constructive foundations of mathematics, cryptography, epistemic logic and applications, epistemology, evidence aggregation, evidence tracking and information reliability, hyperintensional logics, knowledge representation and truth maintenance systems, logic of proofs, provability and verification, logics of resources, paraconsistent reasoning, theory of verification, type theory and typed programming languages.

Faculty

Sergei Artemov
GC

Subash Shankar
Hunter College

Tudor Protopopescu
BMCC, Linguistics

Want more information?
Visit Prof. Armetov’s research page, or email him at sartemov@gc.cuny.edu.

Students

Eoin Moore

Vincent Peluce

Sreehari Kalloormana

Alexander Washburn
Justification Logic and Applications

Classical Logic is concerned, loosely, with the behavior of truth. Epistemic Logic shifts the emphasis to the intensional context and has to do with the behavior of known or believed truths. This field has been actively developed through the second half of the 20th century. Justification Logic is a logic of the 21st century. We cannot simply accept a claim that a given proposition is true. With an overabundance of easily accessible and yet unreliable information, we need checkable evidence that a given proposition is true. Justification Logic provides a well principled logical framework for the corresponding reasoning. It introduces to the logical language a long-anticipated general notion of justification with new atomic propositions “t is a justification for F.” We now have the capacity to reason about justifications, both simple and compound. We can compare different pieces of evidence pertaining to the same fact. We can measure the complexity of justifications, thus connecting the logic of knowledge to a rich complexity theory and to other fields of study.

Epistemic Game Theory

We develop a knowledge-based rational decision model (KBR-model) which offers an approach to rational decision making in a non-probabilistic setting, e.g., in perfect information games with deterministic payoffs. The KBR-model is an epistemically explicit form of standard game-theoretical assumptions, such as Harsanyi’s Maximin Postulate. This model suggests following maximin strategy over all scenarios which the agent considers possible to the best of his knowledge. It can be shown that KBR is the only non-probabilistic decision-making method which is definitive, rational, and based exclusively on knowledge.

Other research

Other research activities of the Lab include:
- revising epistemic logic (modal language, Kripke-style semantics, proof theory)
- formal methods in artificial intelligence
Cities around the world are currently under quick transition towards a low carbon environment, high quality of living, and resource efficient economy. Urban performance depends not only on the city’s endowment of hard infrastructure, but also on the availability and quality of information and social infrastructure. There is a growing importance of Computing and Communication Technologies (CCTs), social and environmental capital in profiling the competitiveness of cities.

At the SCUC Lab we promote and conduct interdisciplinary research on challenging problems of sustainable urban development, such as environment pollutions, transportation and climate change/nature disasters (flooding and storm surge).

We have a team of professors and students from Computer Science, Environmental Science, Mathematics and Physics and CUNY Advanced Science Research Center (ASRC).

Faculty

Zhanyang Zhang
Staten Island

Michael Kress
Staten Island

Tobias Shaefer
Staten Island (Math, Physics)

Alan Benimoff
Staten Island (Engineering, Environmental Science)

Students

Eric Benedetto
Nan Jia

Want more information?
Email zhanyang.zhang@csi.cuny.edu.
Using the Lattice Boltzmann Method to Model and Simulate Ocean Oil Spills

The transport of oil spilled into the ocean is a complex process that depends on the current, wind, temperature and chemical composition of the oil and seawater. We apply Lattice Boltzmann Advection Diffusion Equation (LBM-ADE) to model and simulate ocean oil spill transport at the surface level with assimilation of real ocean current data from the Unified Wave Interface-Coupled Model (UWIN-CM). We validate our model in a benchmark study against GNOME, a tool developed and used by NOAA for ocean oil spill forecast. Our study shows the LBM-ADE model solutions are very close to the targeted analytical and GNOME solutions with the same initial oil spill and location.

Compound Flooding Modeling and Simulation: A Study of Staten Island Floorings

New York City is increasingly vulnerable to floods from extreme weather and rising sea levels. Many coastal neighborhoods suffer from high tide flooding on a monthly basis that is destroying infrastructure and property value. In 2021, hurricanes Henri and Ida struck NYC within weeks of each other with unprecedented levels of rainfall that caused deaths and damages. Additionally, these events flush detritus and contaminants - including oil and volatile organic compounds (VOCs) - polluting and degrading the city coasts and waterways. Staten Island (SI) is significantly affected by these events. As part of a citywide study. We focus on developing a local compound flooding (rains & storm surge) model and simulation with assimilation of multi-modal sensor data. As part of NYC FloodNet initiative, we have deployed a handful flood sensors and gateways on Staten Island. Our project integrates simulation models with wireless sensor networks, cloud-edge computing and machine learning technologies to take advantage of huge amount data come from multi-modal data sources to improve the models performance.
The City College Visual Computing Lab (CCVCL)

The City College Visual Computing Laboratory serves as an experimental environment for both research and education in advanced visual and other media computing. The research activities in the CCVCL primarily focus on the understanding of 3D natural scenes and the events in the scenes from multiple sensor modalities, including visible cameras, thermal sensors, and acoustic sensors. Our work is supported by NSF, DoD, DHS, ODNI and industry. Our graduate students are now working at Amazon, BAE Systems, Canon, CUNY, NASA, NAVAIR, Nvidia, Facebook Reality Labs, Lyft, etc.

Faculty:
Zhigang Zhu, Kayser Chair Professor of Computer Science, CCNY and Graduate Center, zzhu@ccny.cuny.edu
Hao Tang, Professor at BMCC and Graduate Center, htang@bmcc.cuny.edu
Arber Ruci, Director of the NYC Innovation Hot Spot, Entrepreneur-in-Residence and National Faculty for the NYC Regional Innovation Node, Arber.Ruci@cuny.edu
Jin Chen (DSE Graduate at CCNY, CTO at Nearabl, a CUNY Spin-Off)

Research Assistants:
Xuan Wang (CS PhD Program at CUNY Graduate Center)
Yaocheng Wu (CS PhD Program at CUNY Graduate Center, with Prof. Huy Vo)
Bilal Abdulrahman (CS PhD Program at the CUNY Graduate Center)
Jianing Qi (CS PhD Program at the CUNY Graduate Center, with Prof. Hao Tang)
Lizheng Zhou (CS PhD Program at the CUNY Graduate Center, with Prof. Hao Tang)
Jiawei Liu (DSE at CCNY, with Prof. Hao Tang)
Some Recent Projects

**ContrastNet: Unsupervised feature learning for point cloud understanding**
To alleviate the cost of collecting and annotating large-scale point cloud datasets, we propose an unsupervised learning approach to learn features from unlabeled point cloud "3D object" dataset by using part contrasting and object clustering with deep graph convolutional networks (GCNs). The contrasting learning step forces a ContrastNet to learn high-level semantic features of objects but probably ignores low-level features, while the ClusterNet in the clustering step improves the quality of learned features by being trained to discover objects that probably belong to the same semantic categories by using cluster IDs.

**SR-GAN: Using CNNs for crowd analysis with little labeled data**
In this work, we have used a generative adversarial network (GAN) to train crowd counting networks using minimal data. We studied how GAN objectives can be modified to allow for the use of unlabeled data to benefit inference training in semi-supervised learning. More generally, we explained how these same methods can be used in more generic multiple regression target semi-supervised learning, with crowd counting being a demonstrative example. Given a convolutional neural network (CNN) with capabilities equivalent to the discriminator in the GAN, we provided experimental results which show that our GAN is able to outperform the CNN even when the CNN has access to significantly more labeled data.

**Emotion analysis using audio/video, EMG & EEG**
This is a study on automated emotion recognition using four different modalities – audio, video, electromyography (EMG), and electroencephalography (EEG). We collected a dataset using the 4 modalities as 12 human subjects expressed six different emotions or maintained a neutral expression. Three different aspects of emotion recognition were investigated: model selection, feature selection, and data selection. This is a collaboration with Prof. Tony Ro at CUNY Graduate Center.

**MultiCLU: Multi-stage Context Learning and Utilization**
In this work, a storefront accessibility image dataset is collected from Google street view and is labeled with three main objects for storefront accessibility: doors (for store entrances), doorknobs (for accessing the entrances) and stairs (for leading to the entrances). Then MultiCLU, a new multi-stage context learning and utilization approach, is proposed with the following four stages: Context in Labeling (CIL), Context in Training (CIT), Context in Detection (CID) and Context in Evaluation (CIE). Our experiment results show that the proposed MultiCLU framework can achieve significantly better performance than the baseline detector using Faster R-CNN.
Computer & Network Security Lab
CUNY Hunter College

The Computer & Network Security Lab at CUNY Hunter College examines the security concerns that impact computing today. Computers and systems play a huge part in our everyday lives, opening up opportunities for criminals and hackers to try to access information and interfere with our lives. Our research is funded by CUNY and others. Our students and graduates work or have worked at LinkedIn, American Express, Accenture, Deloitte, ViacomCBS, and Raytheon.

We partner with multiple universities worldwide, including Korea University, University of Bonn, and more.

Get in touch: spock@ieee.org
www.cs.hunter.cuny.edu/~spock/

PhD Students

Current

Shoufu Luo
Jeremy D. Seideman
Isa Jafarov

Past

Patrick Duessel
(Graduated 2018, now Senior Director at Raytheon)

Why computer and network security?

From data breaches to software vulnerabilities, network intrusions and malware, there are many ways that bad actors can try to disrupt people's lives.

Since smartphones, home automation, and cloud services play a large part of our lives, we have more of our lives in the digital world. This means that there are that many more opportunities for systems and data to be captured, changed, and leveraged for nefarious purposes.

Security is an arms race - both sides are constantly innovating and mitigating, in order to bypass the other. By understanding the threats that exist, we look for new ways to protect against them, and look for ways to protect against future threats.

Last modified: September 2022
Selected current projects at the CNS Lab

Source code and binary similarity
Software projects move towards community-based development (using tools such as SourceForge and GitHub) and create more opportunities for developers to reuse and share code. While this helps advance the field and allows for new and varied software, it also opens up the possibility of software vulnerabilities - defects in software that can lead to data integrity issues or attack possibilities. Being able to detect similar chunks of code in both source and binary files helps detect these vulnerabilities, see how widespread they are, and determine the optimal strategy to fix them, protecting the entire software landscape.


Software Vulnerability Origin Discovery
Vulnerabilities are a prevalent problem - bots, trojans, and ransomware exploit them computers, steal data, and interrupt systems, the Internet, and access. Being able to determine the origin of that vulnerability - where it came from and how it evolved - helps researchers determine how to protect the software supply chain and where it originated. In this way, both software and security engineers can help track down the source responsible for the software flaw. Cf. NTIA’s Software Bill of Materials (SBOM) in industry.


Insider Threat
Insiders present a special problem in computer security. They are “inside the system” and can cause severe damage. We aim to find behavioral patterns to identify abnormal behavior in enterprise networks and host log files using machine learning techniques.


Cybersecurity Experimentation Frameworks
We are building a highly configurable cybersecurity experimentation testbed. This DETER-style testbed allows for proper testing of attack and defense mechanisms in a safe environment away from the production networks. It also allows for federation with existing experimentation testbed networks. Our multi-node testbed allows for experiments to be run on a variety of hardware platforms and at network speeds up to 10 Gbps. The next generation of testbeds is already in the making.

DDoS, botnets, and next-generation networks
We are investigating new attack and defense techniques for the current and next-generation Internet architectures, including Software Defined Networks.


CUNY PONDER
The ProGramming laNguaGeS anD software EnGineering ReseaRch Lab

CUNY PONDER, situated at CUNY Hunter College, works at the intersection of software engineering, programming languages, and (Machine Learning) systems. Our lab develops and assesses techniques for automating critical software engineering tasks, including (static and dynamic) program analysis, software evolution and maintenance, and empirical software engineering. Our work results in algorithms and associated tools that analyze and transform (e.g., refactor) large and complex programs for improved modularity, comprehension, maintainability, safety, security, and performance. The techniques span multiple subfields of theory and application, such as programming languages, type theory, (front-end) compilers, data science, informational retrieval, and mathematical logic. We collaborate with the DAIR lab.

We are funded by the National Science Foundation (NSF), the Japan Society for the Promotion of Science (JSPS), Amazon Web Services (AWS), the Verizon Foundation, and the CUNY Diversity Projects Development Fund (DPDF). Our graduates have obtained research positions at prestigious universities and work at Google, New York Times, Squarespace, TD Ameritrade, New York Foundling, J.P. Morgan, and AD/FIN. Our publications have appeared in top software engineering and programming language conferences and have won distinguished paper awards.

Get in touch at ponder@hunter.cuny.edu and find out more at http://ponder-lab.github.io.

Professor
Raffi Khatchadourian

Students
Tatiana Castro Vélez
Yiming Tang
Ye Paing
Allan Spektor
Oren Friedman
David Morant
Walter Rada
Olivia Moore
Md. Arefin
Sample Projects

Migrating Imperative Deep Learning Programs to Graph Execution

Efficiency is essential to support responsiveness w.r.t. ever-growing datasets, especially for Deep Learning (DL) systems. DL frameworks have traditionally embraced deferred execution-style DL code that supports symbolic, graph-based Deep Neural Network (DNN) computation. While hybrid approaches aim for the "best of both worlds," the challenges in applying them in the real world are largely unknown. We conduct a data-driven analysis of challenges—and resultant bugs—involved in writing reliable yet performant imperative DL code by studying 250 open-source projects, consisting of 19.7 MLOC, along with 470 and 446 manually examined code patches and bug reports, respectively. The results indicate that hybridization: (i) is prone to API misuse, (ii) can result in performance degradation—the opposite of its intention, and (iii) has limited application due to execution mode incompatibility. We put forth several recommendations, best practices, and anti-patterns for effectively hybridizing imperative DL code, potentially benefiting DL practitioners, API designers, tool developers, and educators.

Refactorings and Technical Debt in Machine Learning Systems

Machine Learning (ML), including Deep Learning (DL), systems, are pervasive in today’s data-driven society. Such systems are complex; they are comprised of ML models and many subsystems that support learning processes. Unfortunately, there is a gap in knowledge about how ML systems actually evolve and are maintained. In this project, we fill this gap by studying refactorings, i.e., source-to-source semantics-preserving program transformations, performed in real-world, open-source software, and the technical debt issues they alleviate. We analyzed 26 projects, consisting of 4.2 MLOC, along with 327 manually examined code patches. The results indicate that developers refactor these systems for a variety of reasons, both specific and tangential to ML, some refactorings correspond to established technical debt categories, while others do not, and code duplication is a major cross-cutting theme that particularly involved ML configuration and model code, which was also the most refactored. The results can potentially assist practitioners, tool developers, and educators in facilitating long-term ML system usefulness.

Safe Automated Refactoring for Intelligent Parallelization of Java 8 Streams

The Java 8 Stream API sets forth a promising new programming model that incorporates functional-like, MapReduce-style features into a mainstream programming language. However, using streams efficiently may involve subtle considerations. For example, although streams enable developers to run their code in parallel with little alteration, it is often not obvious if such code runs more efficiently this way. In fact, under certain conditions, running stream code in parallel can be less efficient than running it sequentially. Moreover, it can be unclear if running sequential stream code in parallel is safe and interference-free due to possible lambda expression side effects. This project involves an automated refactoring approach that assists developers in writing optimal stream client code in a semantics-preserving fashion. Based on a novel data ordering and typestate analysis, the approach consists of refactorings that include preconditions and transformations for automatically determining when it is safe and possibly advantageous to convert a sequential stream to parallel and improve upon already parallel streams. The approach is implemented as a plug-in to the popular Eclipse IDE utilizing both WALA and SAFE.
**RESEARCH TEAM**

**NAME:** Minh Nguyen  
**ROLE:** PhD candidate  
**AFFILIATION:** The Graduate Center, CUNY  
**JOINED:** Fall 2016  
**INTEREST:** Cloud security, software-defined networking

**NAME:** Xiaojie Zhang  
**ROLE:** PhD candidate  
**AFFILIATION:** The Graduate Center, CUNY  
**JOINED:** Fall 2017  
**INTEREST:** Edge computing, machine learning

**NAME:** Motahare Mounesan  
**ROLE:** PhD candidate  
**AFFILIATION:** The Graduate Center, CUNY  
**JOINED:** Fall 2022  
**INTEREST:** Big data, trustworthy computing

**NAME:** Manal Zneit  
**ROLE:** PhD student  
**AFFILIATION:** The Graduate Center, CUNY  
**JOINED:** Fall 2022  
**INTEREST:** Federated learning, visual computing

**NAME:** Shima Yousefi  
**ROLE:** PhD student  
**AFFILIATION:** The Graduate Center, CUNY  
**JOINING:** Spring 2023  
**INTEREST:** Edge computing, machine learning

---

**MISSION STATEMENT**

**PROF. SAPTARSHI DEBROY**

The broader research vision of DiSsect lab is focused on solving ‘system management and security problems of future distributed and smart systems that are uniquely challenged by their components, purpose, and operating principles and have direct societal interactions and impact’.

---

**RESEARCH INTERESTS**

**Distributed Computing Data Science:** Data-driven resilient management of distributed systems, application of AI/ML towards resource provisioning for ‘data-at-rest’ vs. ‘data-in-motion’, federated system measurement and monitoring, hosting complex applications on IoT systems and cyber physical systems.

**Usable Cyber Security:** Trustworthy edge and cloud computing, end-to-end security assurance, Science DMZ and securing campus cyber-infrastructure, distributed reputation and trust management, SDN vulnerability analysis and threat mitigation.

---

**SPONSORS**

[NSF logo]  
[Verizon Foundation logo]  
[Other logos]
Machine Learning Lab @ Hunter-CUNY

The Machine Learning Lab at Hunter College builds computational agents for dynamic, real-world environments, with a strong focus on knowledge representation, reasoning, and learning. We are founded by the National Science Foundation and Google. Our graduates are employed at Google, Amazon, and in academia.

Collaborative intelligence allocates subtasks to people and computers in pursuit of human goals. Collaborative intelligence provides advice and information, and is particularly useful in response to change. A shared world model and communication are essential.

Synergistic Agriculture harnesses AI and robots to design, create, and maintain develop new climate-sensitive, innovative urban food garden

Situation-based reasoning recognizes and adapts to situations through a properly balanced, diverse set of rationales.

Online learning and planning is essential for crowd-aware service robot navigation.

Additional research thrusts: the development of expertise, cognitive architectures, cognitive robotics, and computational cognitive neuroscience.

cs.hunter.cuny.edu/~epstein/
Topological Data Analysis

PROFESSOR MIKAEL VEJDEMO-JOHANSSON (MVJ)

Topological Data Analysis is a new research field that draws on algebraic topology in pure mathematics to construct applied algorithms, statistical tools and data science approaches.

At CUNY, MVJ is currently the only faculty member primarily interested and active in the TDA community.

Get in touch: mvejdemojohansson@gc.cuny.edu

WEEKLY SEMINAR

The Data Science and Applied Topology Seminar will start up again as soon as possible. More details: http://cunygc.appliedtopology.nyc

PRIMARY RESEARCH INTERESTS

Persistent Homology
From a point cloud, we can construct a simplicial complex (approx. mesh) that captures the underlying geometry. Topological invariants of this complex help us understand the data.

Persistent Cohomology
Starting the same way as above, Persistent Cohomology can produce new coordinate functions taking values on the circle. These help us understand periodicity in time series as well as directional data.

Mapper
The Mapper algorithm produces a simplicial complex model from arbitrary data, with facilities to mark aspects of the data as important for distinguishing groups.

Software
We maintain the JavaPlex library. It is the only major library available for TDA for Java and Matlab. We contribute to implementations of Mapper.

Multiple Hypotheses
Applications to genetics call for very many simultaneous instances of statistical testing with topological tools. We have developed Multiple Hypothesis correction methods in topology.

Coverage and Evasion
How well does a collection of sensor cover a region? Are there gaps and holes? How about moving sensors and moving targets? Topology has methods for this, and we study extensions to sensors that could fail.

BUT ALSO

Color Naming Semantics
Systems for naming colors in different languages seem to follow specific patterns. We analyze the World Color Survey at a finer grain using machine learning and data science tools.

Mathematical Art
Using 3d printers and laser cutters we produce and exhibit mathematical art depicting concepts and objects in topology and geometry.

[your area of interest]
We are keen on new areas to work in – Bayesian Statistics; Spatial Statistics; Applied Category Theory… If you want to work in the interface of pure and applied mathematics – see if we can find a project idea!
Statistics in Topological Data Analysis

Hypothesis testing with persistent homology can be done using repeated simulations following a theoretical null model for the data source. If the observed topology is much more extreme than in the null model, the test can be seen as indicating topological structure in the data.

To introduce multiple hypothesis correction we are forced to go back to basics and investigate the behavior of the null model itself, producing novel hypothesis tests for a topological setting and constructing multiple hypothesis corrections that do not scale the computational burden too much.

Path Distances and Cylindrical Coordinates

Some data can have both periodic (seasonal) and linear (trend) components, and would best be analyzed with coordinate functions to spaces like \((S^1)^n \times \mathbb{R}^m\). We explore notions of distances between paths and methods for computing such hypercylindrical coordinates to find novel applications and algorithms for mixtures of coordinate spaces.

Pursuit Strategies and the Topology of Evasion

The Topology of Redundant Coverage

Given a pattern for \(N\) agents moving in a region and pursuing an evader, topology can determine whether or not an evasion path exists. From this determination, we can evaluate strategies for pursuit and build new heuristics to direct groups searching for an entity.

When sensors can fail, we may need redundant coverage to keep holes or evasion paths from emerging. This calls for novel constructions – the \(n\)-fold Čech Complex, the \(n\)-fold Delaunay Complex etc., should be able to do the job in this novel setting, but we need to prove that it works and uncover the complications we might meet.

Applications of Circular Coordinates

We, and other research groups, have been applying circular coordinates to time series data – of motion capture time series, on financial data – and on directional data for HAM radio fox hunting.

We are looking for more interesting areas of application. Both for new time series questions we have not considered, but also for completely different areas of application.
Underwater Network and Intelligent System (UNIS) Lab

The Earth is a water planet. For decades, there has been significant interest in monitoring aquatic environments for scientific exploration, commercial exploitation and coastline protection. Highly precise, real-time, and temporal-spatial continuous aquatic environment monitoring systems are extremely important for various applications, such as oceanographic data collection, pollution detection, and marine surveillance.

Faculty

Zheng Peng, CCNY
zpeng@ccny.cuny.edu

In the UNIS lab at the City College of New York, we conduct research on fundamental issues in the underwater wireless network and systems, including efficient acoustic communications, medium access control, data routing and forwarding, reliable data transfer, security and robustness, data storage and management, localization and synchronization, and data fusion, dissemination, and tracking, etc. We also explore various design issues of integrated underwater sensor nodes (which are composed of sensing, computing, communication, and mechanical control components), such as efficient power and resource management. In the long run, we aim to make underwater wireless technology practical and useful for scientific research and commercial applications.

Here are some ongoing research work at the UNIS lab.

Underwater Mobile Networks

The past few decades have witnessed a drastic increase of interest in developing autonomous underwater vehicles (AUVs) and other tetherless underwater robots. They have been considered as powerful tools for the exploration and utilization of the planet’s aquatic environments. We envision the future underwater application would employ a fleet of submersible robots to perform collaborative operations, hence forming an underwater mobile network. This integration of marine robots and underwater networks would revolutionize the way we conduct ocean sampling and exploration.

Software Defined Underwater Acoustic Modems

To better support underwater wireless networks, the design of acoustic modems needs to have upper-layer protocols and modules in mind. Communications
between software and hardware developers are critical to delivering a high performance underwater communication and network system. In this research direction, we focus on the flexibility, extendibility and configurability of acoustic modems. The objective is to design highly customizable underwater acoustic communication devices based on novel hardware and software platforms.

Efficient Underwater Medium Access Control

Medium access control (MAC) protocols determine how nodes in the network share the underlying communication channel and are quite important to network performance. In this effort, we investigate different MAC protocols, including parallel reservation based MAC and multi-channel MAC protocols for underwater networks. To fully utilize the available bandwidth, we designed a number of novel MAC protocols to support multiple users in more efficient ways.

Cognitive Underwater Networks for Sustainable Ocean Monitoring and Exploration

In this research effort, we focus on the environmental awareness, scalability and sustainability of underwater wireless network systems. We introduced the concept of cognitive underwater networks, which are aware of the coexistence of multiple underwater acoustic systems. The project goals include investigating acoustic spectrum usage via data collection and analysis, designing algorithms, software and hardware for acoustic spectrum sensing and spectrum management, and evaluating the feasibility and performance of the proposed work via simulations, and lab and field experiments.
Network Simulators for Underwater Wireless Networks

In this effort, we are developing network simulators that are dedicated to underwater systems. We have previously developed, Aqua-Sim, a NS-2 based simulator that can accurately model the underwater acoustic channel, the three-dimensional topology and the mobility pattern of underwater objects. In a recent development, an updated version of the simulator, called Aqua-Sim-NG, is implemented based on NS-3 with enhanced channel support, realistic modem features, better efficiency and memory management.

Security Issues in Underwater Wireless Networks

Similar to terrestrial wireless networks, underwater wireless networks are inherently vulnerable to attacks due to their broadcast nature. However, underwater wireless networks have their special characteristics that require new designs in encryption, attack detection, self-protection, defending and recovering. Preliminary work has been done to study the effectiveness of different types of attacks, as well as the method to protect communications. This is an area we are keenly interested in exploring more.