ARIA

APPLIED RESEARCH IN ACTION

2021 INTERNSHIP PROJECTS

UNIVERSITY OF TORONTO

MScAC Master of Science in Applied Computing

THANK YOU TO OUR INDUSTRY PARTNERS

[24]7.ai
AMD
AppDirect
BlackBerry
Blees AI
CanDIG
Cohere
CPP Investments
Deloitte
Dewpoint Therapeutics
Geotab
Huawei Technologies China
Hypatia Systems
ICICI Bank
Kindred AI
Kraft Heinz
Layer 6
Loblaw Digital
Lunenfeld-Tanenbaum Research Institute, Sinai Health
Microsoft Turing
ModiFace
NVIDIA
New Vision Systems Canada
ODAIA
Pearson Canada
Samsung AI Centre
Scrawlr
Semantic Health
Sensibill
SickKids
SOTI
Surgical Safety Technologies
TealBook
The Vanguard Group, Inc.
TimePlay
Vidado
Winterlight Labs
Ziva Dynamics

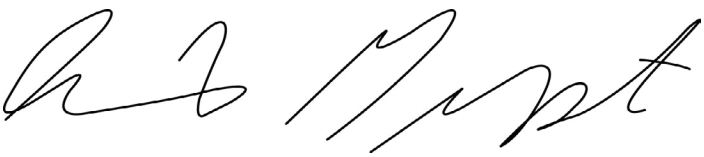
The Master of Science in Applied Computing program (MScAC) is proof that technological innovation happens best when we build effective industry-academic collaborations. Nowhere is this more apparent than at the annual Applied Research in Action (ARIA) showcase.

For the second year in a row, ARIA was a virtual event, welcoming participants from around the world. ARIA is an opportunity for our MScAC students to celebrate, by showcasing their eight-month long collaborations with their industry partners to conceptualize and implement new ideas that lead to the innovations fuelling technology.

For the first time, ARIA also included PhD students from the Department of Computer Science, whose original research acts as a springboard for technical innovation spanning a range of different industries. A number of companies also set up displays to showcase their internal R&D.

We were delighted to have so many people join us in celebrating this milestone achievement and the latest innovations in technology. We were particularly pleased to welcome so many students from many countries who came to learn about the MScAC program, which led to many of then subsequently choosing to apply for admission. Our sincerest thanks to everyone who made ARIA 2021 possible and who make the MScAC Program one of the world's best Applied Computing programs.

Congratulations to our 2020/21 cohort and welcome to our MScAC alumni family!



Arvind Gupta
Professor
Academic Director, Professional Programs
Department of Computer Science
University of Toronto

2021 AWARD WINNERS

Student Innovation Award

Alexander Cann

Nominated by an industry supervisor, this award is presented to a student in the internship phase of the MScAC program who has demonstrated innovation, dedication, and creativity in the completion of their internship project.

Alexander Cann wins this award for demonstrating a great mixture of dedication, initiative, innovation, as well as persistence, to both the company and his internship project at AMD.

Faculty Recognition Award

Marsha Chechik

The Faculty Recognition Award recognizes an academic supervisor who has shown exceptional commitment, time, and skilled attention to developing the talents of their MScAC students during the internship phase of the program.

Professor Marsha Chechik was described by her student as an enthusiastic, motivating and dedicated academic supervisor, who motivated her to go beyond responsibilities and learn skills that could be carried on in industry after she graduated.

Industry Icon Award

Jekaterina Novikova

The Industry Icon Award recognizes an industry partner who has demonstrated exceptional commitment, time, mentorship, and skilled attention to developing the talents of their MScAC student during the internship phase of the program.

Jekaterina Novikova provided constant mentorship to her student, reviewing project documents, presentations and papers right after they were done, and leaving constructive comments on how to improve them. She is described by her student as being the reason for "the best research experience," where invaluable things were learned about both academic research and industry.

Alumni of the Year Award

Tom LaMantia

The Alumni of the Year Award is nominated by the MScAC program team itself and is presented to one or more members of the alumni network, who have dedicated much of their own time to contribute to the program's development.

Tom LaMantia has been an extremely active founding member of the Alumni Association, and has been a willing volunteer to help with any professional training our students have required, providing students with guidance and advice on everything from resume building to interview training. Most recently, he was involved in the launch of MScACTalks – leading the inaugural discussion with our first guest speaker, Prof. Eugene Fiume.

ARIA Spotlight Award

Sam Toueg

The ARIA Spotlight Award is awarded to an individual who has had a tremendous impact on the MScAC program.

Professor Sam Toueg willingly held the Admissions Committee Chair position for four years (2016 – 2020). During his tenure he led the charge on reviewing almost 4000 applications and admitted over 50% of our current alumni network, as well as the MScAC students presenting at this year's ARIA.

2021 RESEARCH INTERNSHIP PROJECTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

Adapting Contextual Bandit Algorithms to Ad-Serving Platforms

Company: [24]7.ai

Student: Kaijian Zhong

Industry Supervisor: Abhishek Ghose, Emma Nguyen

Academic Supervisor: Leonard Wong

Contextual bandit algorithms are developed to gain maximum rewards from an unknown distribution in limited trials by exploiting contextual information. They can be adapted to yield solutions for ad-serving systems with a short campaign time. Several new contextual bandit algorithms have been proposed in recent years, but their performance is poorly understood in real-world scenarios. This study tests these algorithms, particularly the Adaptive-Greedy algorithm (David Cortes, 2019), in multiple simulated and real-world datasets. We found that the Adaptive-Greedy algorithm can achieve up to 5% click-through rate improvement compared to the widely used Bootstrapped Thompson Sampling algorithm (BTS). At the same time, the Adaptive-Greedy algorithm can be up to 5 times faster in training and predicting than the BTS algorithm. Off-policy evaluation (OPE), which allows us to evaluate our new models from the log data generated by other policies, provides vital backtesting results to inform new policy decisions. To this end, we study and compare different OPE methods and use them to evaluate our proposed algorithms.

Reference: David Cortes. Adapting multi-armed bandits policies to contextual bandits scenarios. 2019

Defending White Box Adversarial Examples with a Guided Denoiser

Company: AMD

Student: Alexander Cann

Industry Supervisor: Ihab Amer

Academic Supervisor: Gennady Pekhimenko

As deep neural networks become increasingly common in security conscious applications such as self-driving cars, or malware detection, protecting networks from manipulation by an attacker has become increasingly important. However, neural networks trained to classify images are vulnerable to mistakes caused by images imperceptibly altered by an adversary.

Existing defences against adversarial examples either require extensive retraining, as in adversarial training, or have been shown to be ineffective against an adversary that is aware of the defense (a white box adversary), as in the case of many pre-processing defenses. We show that a denoising neural network, can be trained to provide white box defense, extending previous work which showed effectiveness only against an adversary unaware of the defense. Our defense provides comparable clean and robust accuracy to adversarial training when applied to the model used to guide the denoiser during training. Additionally, the denoiser was transferred to other models trained on the same dataset where it provided a 35% increase in robust accuracy, on average, when compared to the same models with no defense.

Further investigation is required into how the defense transfers between distinct image classification tasks, and how the model used as a guide during training changes the efficacy and transferability of the defense.

Modern Gameplay Test Automation with Reinforcement Learning

Company: AMD

Student: Allen Bao

Industry Supervisor: Yonas Bedasso, Max Kiehn, Vinay Pandit

Academic Supervisor: Amir-massoud Farahmand

Gameplay testing on computer games to help ensure high performance for graphics cards is often performed manually, resulting in substantial time and cost expenditures. Though many reinforcement learning (RL) studies have been conducted on gameplay, most tend to focus on older game titles, and have access to internal game information through APIs. This project's objective is to develop a solution capable of automating a limited time (5-10 minutes) playthrough on a modern game title with RL, without relying on access to internal game information

.

The solution focuses on automating player navigation to make progress in singleplayer games. We collect screenshots during real-time gameplay as the RL agent traverses the game environment, which we then put through image analysis and computer vision tools for processing to serve as observations for the agent. A neural network then handles processing of policy updates during training timesteps, using the proximal policy optimization (PPO) method. The model is qualitatively evaluated by observing fully automated gameplay, which has shown the current solution to effectively maneuver through navigation-focused areas, as well as some instances when it is necessary for the agent to correctly react to in-game interaction prompts. Future work will involve training the model on different types of gameplay, and on transfer learning to new titles. The results from this project will ultimately demonstrate the ability of RL to perform well in modern gameplay settings and contribute to increasing efficiency in game testing pipelines.

Reducing Data Loading Time in Machine Learning by Reusing Minibatches

Company: AMD

Student: Jafeer Uddin

Industry Supervisor: Abhinav Vishnu

Academic Supervisor: Maryam Mehri Dehnavi

Training of models in machine learning is typically done by dividing a dataset into equally sized subsets called minibatches. At each step of training, a minibatch is loaded to the GPU, where it is used once to determine updates to parameters. If data loading is the bottleneck, then this process will lead to idle time for the GPU as it waits for the next minibatch, wasting precious GPU compute resources. Thus, we propose a change to the training procedure and suggest reusing minibatches that have already been loaded into the GPU to perform additional parameter updates while waiting for the next minibatch. If the number of times each minibatch is used is kept the same as regular training, then this idea will lead to a speed up in training time. An unintended consequence of reusing minibatches is that it can lead to overfitting as the model ends up memorizing the reused minibatch. To remedy this, we modify the technique to adaptively adjust the amount of reuse based on how the model performs on unseen data. We have evaluated our technique on some example models with data loading bottleneck and found that on average we get around 1.5× speed up with no loss in model accuracy. Some next steps include trying out more examples to see if the results still hold and perhaps attempting a more sophisticated algorithm for the overfitting issue to see if we can get more speed up.

Hybrid Recommender System for Applications Based on Heterogeneous Commercial Data

Company: AppDirect

Student: Yiwen Feng

Industry Supervisor: Dominic Lee

Academic Supervisor: Peter Marbach

Collaborative filtering is a widely used method in recommendation tasks. Incorporated with implicit feedback, collaborative filtering can recommend items to help users explore new interests. The currently developed recommender system in the application marketplaces focuses only on the surface interactions between users and items, which limits the use of information from various sources.

To further improve the existing recommendation engine, we use a wider range of user information, such as the browsing history in the website. Moreover, the model can take advantage of applications' information, including their properties and descriptions. The products' information is embedded by a pre-trained language model. The embeddings provide a better understanding for our model to recommend applications. Regarding our method, instead of using collaborative filtering, we introduce a factorization machine with the neural network to create an end-to-end model. This model can learn combined implicit features as well as corresponding low- and high-order interactions simultaneously.

Comprehensive experiments have been conducted on commercial data to illustrate the efficacy of the proposed model compared with existing baseline models. The results demonstrate that our proposed model outperforms the current collaborative filtering-based model with an increase of recommendation relevancy, reflected by recall growing to 0.36, in an offline environment. Furthermore, we surveyed our advisor team to provide an interpretation of our results, which also proved that our improved system is more effective from users' perspective. As we aim to continually enhance our recommender system, the next steps would involve researching how to represent and integrate more data, such as utilizing a knowledge graph or the sequential interactions, in our model.

APIFuzz: Introspection Assisted Fuzzer for QNX Resource Managers

Company: BlackBerry

Student: Congwei Chen

Industry Supervisor: Glenn Wurster

Academic Supervisor: David Lie

BlackBerry QNX is a microkernel real-time operating system built for the world's most critical embedded systems. Resource managers are widely used for local communication between processes in QNX system. These critical processes are usually stateful and accessed by other applications using portable operating system interface. Fuzzing is a dynamic analysis security testing technique that involves providing random or invalid data as inputs to programs. A greybox fuzzer leverages coverage-feedback to learn how to reach deeper execution states of the program. It often uses instrumentation to get runtime coverage for guiding the test case generation, which introduces a weak proof on whether the unmodified binary is safe and operates as designed in the presence of faults. Thus, to satisfy safety-specific audit criteria, the target binary should not be modified.

In this project, we replace instrumentation, which modifies the binary, with introspection, which examines the state of the binary without modifying it, by introducing a new introspection-based gray-box fuzzing tool targeted at resource managers in both kernel and user mode. The experiments show that the tool, using introspection, explores more program states than a corresponding black-box fuzzer.

The next step involves experimenting with more complex resource managers and the kernel.

Taking the First Steps Towards Building a Hardware Bill of Materials

Company: BlackBerry

Student: Divas Kapur

Industry Supervisor: Cris Sinnott, Billy McCourt, Paul Hirst

Academic Supervisor: Ishtiaque Ahmed, Courtney Gibson

Modern day software systems are powerful, efficient, and multi-tiered. However, this comes with the cost of increasing complexity, vastness, and reliance on a plethora of components and libraries, not all of which are known intimately to developers, testers, and consumers of these software systems, leaving them exposed to risks and vulnerabilities.

BlackBerry's Jarvis, a software compositional analysis platform, breaks down software binaries into a software bill of materials and highlights potential vulnerabilities by severity category that help customers ensure that their software supply chains are secure. In an effort to rise against competition and to lead the market in terms of capability and performance, our project attempted to take the first steps towards going one layer deeper and identifying and gathering information about the hardware and processor from the software binaries uploaded by customers, providing a more comprehensive picture of the potential threat landscape surrounding the software being used by them.

To start, we set out to find perceptible differences in the architecture families of the ARM processor lineage, closely inspecting various properties and features across generations to build a ladder of potential detection strategies for ARM families and processors. Having synthesized a working list of differences for most used ARM architectures, we disassembled sample binaries and wrote pattern matching rules to detect the variant versions, identify them and map them to known CVEs surrounding the processor and architecture.

Arming the Jarvis tool with this functionality will result in a broader and deeper understanding of the security and risk architecture of a customer's software and hardware infrastructure, and consequently, grant BlackBerry the ability to assess the security posture more holistically by mapping more relevant vulnerabilities and patches. Continuing this project and research will enable Jarvis to be a market leader in its class and empower users to know more about the software they interact with, use, and sell.

Automated Domain Specific Essay Scoring

Company: Blees AI

Student: Yuanhao Lou

Industry Supervisor: Sojin Lee, Alexandre Boulenger

Academic Supervisor: Linbo Wang

The problem of Automated Essay Scoring (AES) is complex and solving the problem will have significant impact on educators teaching and students learning experiences. Meanwhile, AES can significantly reduce cost of the education since it acquires less marking and tuition fees for students for their learning. The goal of this project is to provide a complete version of the software platform to the educators to assist their essay grading. To achieve this goal, the company automates the process of grading students' essays and instantly provides consistent and personalized feedback at scale. The main task is to analyze the performance of the machine learning AI models and create a road map to enhance the model. The AI models do not evaluate sentence structures, or grammar, format, but evaluate the context of the student responses and look for the key concepts that educators are looking for.

To solve this problem, the BERT model is used with data from Stack Exchange which primarily serves as a platform for users to ask and answer questions. Meanwhile, other methods such as LSTM and Word2Vec are used as baseline scores. Outcomes of the BERT model is very promising with 8 out of 10 tags can achieve accuracy higher than 93%, which are far better than other approaches. In the future, we are going to build automated personalized feedback modules for students to further assist their study.

Secure Cross-Service Genomic Data Federated Analysis with GraphQL

Company: CanDIG

Student: Siyue Wang

Industry Supervisor: Jonathan Dursi

Academic Supervisor: Yun William Yu

Health research is a complex field that requires access to the large number of datasets that contain extremely diverse data types such as genomics, transcriptomics, imaging, clinical records and more. Performing complex analysis across these health data in different data stores can be challenging. GraphQL as a new and mature technology for complex queries provides flexible queries to all users. However, its lack of built-in fine-grained authorization and the rich queries it allows can be a concern in the health data research. We propose that GraphQL can be used in the combination of a single decision policy point, to allow flexible, secure and complex cross-service queries. By connecting the underlying data services to the policy point, our GraphQL interface can only query the authorized data from different services.

We demonstrated that secure and private data analysis and machine learning model training can be achieved with GraphQL. Our results show that our GraphQL interface is able to return machine learning model without exposing sensitive data. In addition, our GraphQL interface provides a more efficient and convenient way to query. Here we illustrate that querying from our GraphQL interface transfer fewer bytes than querying from the traditional RESTful API interfaces to the client. Our GraphQL interface provides a sturdy ground for federated learning. In the future, we hope to enable different research institutions to perform federated learning by using our GraphQL interface's machine learning query.

Forecasting Future Fundamental Metrics from Earnings Call Transcripts Using GR4ML to Guide Machine Learning Solutions

Company: CPP Investments

Student: Aneek Das

Industry Supervisor: Brendon Freeman

Academic Supervisor: Eric Yu

An earnings call is a conference call between the management of a public company, analysts, investors, and the media to discuss the company's financial results during a given reporting period. In this report, we explore various NLP and Deep Learning algorithms/models that can be used to extract features from the earnings calls in our dataset and later use them to predict the company's future fundamentals. The noisy nature of our target and the extremely long length of the transcripts make training existing deep learning models challenging. Even though Bi-LSTM with attention is considered to be the best model for this task, we show that making certain changes to the data and model architecture, and, using lower complexity models like LDA and LSTMs with fewer parameters helps generalize well to out-of-sample data points. Finally, we use state-of-the-art transformer-based models designed to handle long sequence lengths for the same task and present our results.

Asset Vision: A Generalized Approach to Asset Tag Extraction

Company: Deloitte

Student: Brendan Kolisnik

Industry Supervisor: Jerod Wagman

Academic Supervisor: Radu Craiu

Organizations seeking to optimize their maintenance and procurement strategies have realized that the necessary contextual data for their core assets rest in physical piping and instrumentation diagrams (P&IDs). Hiring engineers to manually identify these assets, called asset tags, is time-consuming, of limited accuracy (humans are typically able to accurately find 70-80% of asset tags in this way), and prohibitively expensive.

Deloitte has developed a technology that can use machine learning and computer vision techniques to extract the asset tags present in a document. A portion of the technology involves a "merge model" to predict whether two texts identified via optical character recognition should be merged. We have enhanced the model to learn when to merge the appropriate texts from the data without the use of pre-specified text patterns. This resulted in a significant upgrade to the model's performance and reduced the need for client-specific feature engineering, streamlining time to delivery to new clients.

By leveraging a new merge candidate proposal algorithm and creating computer vision features using techniques such as Hough transforms, the "merge model" and downstream tag classifier achieve superior precision and recall scores on the test dataset compared to the corresponding models in the prototype pipeline. The number of asset tags accurately extracted on the test dataset by the new pipeline is similar to that of engineers while only taking a fraction of the time. Future work will focus on minimizing the number of optical character recognition errors.

Using a Temporal Convolutional Network for Demand Forecasting in Retail

Company: Deloitte

Student: Sasha Nanda

Industry Supervisor: Joe Roussy

Academic Supervisor: Arvind Gupta, Huaxiong Huang

Retailers face crucial replenishment decisions such as which products to restock and when. Stocking too few items will lead to stockouts and revenue loss, while stocking too many items will result in decreased working capital, higher costs, and stale inventory. To inform replenishment decisions, many retailers predict future product demand using a rolling yearly average of product sales. While the yearly average is a reasonable demand indicator for products that sell frequently, it fails to capture demand for infrequent purchases and does not account for seasonality or the effects of promotional price changes on demand. Current AI demand forecasting solutions improve upon industry standards by capturing seasonality and promotional changes. However, these solutions are not effective for products that sell infrequently. These products contribute significantly to overall revenue, so it is critical to produce accurate forecasts. Current AI solutions are also laborious, expensive, and time-consuming to implement due to the vast feature engineering required. Deep learning solutions are capable of circumventing lengthy feature engineering processes. Hence, a temporal convolutional network (TCN) was built to perform feature extraction. The TCN model architecture was designed and adapted to capture both short-term and long-term seasonal trends in the sales data. Forecasting one week into the future on products that frequently sell, the retail accuracy of the TCN model was 37.1%, compared to 30.2% for the yearly average, and 38.5% for the current AI solution (MLP). Forecasting one week into the future on items that do not sell frequently, the retail accuracy of the TCN model was -4.1%, compared to -45.2% accuracy for the yearly average, and -10.1% for the MLP. Although the TCN closely matches the performance of the MLP for frequently selling items, the model outperforms both the yearly average and the MLP for infrequently selling products. The results exhibit that the TCN, which requires far less engineering than the MLP does, shows much promise as a demand forecasting solution, particularly proving strong for sparse data where features that capture long-range seasonality are extracted effectively. Open research questions remain, such as whether the model is capable of learning the price elasticity of products given past demand and pricing information.

Evaluating and Debiasing Feature Embeddings for Condensates

Company: Dewpoint Therapeutics

Student: Lalit Lal

Industry Supervisor: Étienne Dumoulin, Grant Watson, Kshitij Gupta

Academic Supervisor: Alan Moses

Drug discovery is an involved task, including close collaboration between biologists, data scientists, and engineers. Unfortunately, this process can take several years to develop a single drug due to the manual requirement of screening cells. Recent advancements in artificial intelligence, specifically deep learning, has allowed for significant advancements to the classical pipeline of drug discovery.

We aim to see if deep convolutional neural networks (DCNNs) can be used to learn morphologically significant information about cells to shed light towards important cellular functions that may be relevant for drug discovery. Specifically, we aim to learn feature representations to detect and distinguish a biological phenomenon called condensates.

Condensates are liquid-like compartments of molecules for key biochemical processes. They play a crucial role in the life cycle of a cell. In biological images, specifically in high throughput microscopy screens (HTS), labels related to condensates are rare. Fortunately, condensates are uniquely identifiable based on other information, such as their localization. We investigate training DCNNs in a weakly supervised manner on these signals to develop weakly supervised embeddings (WSE).

We conduct our experiments on the Human Protein Atlas (HPA), an open-sourced dataset that contains signals weakly related to condensates. We show that a model can distinguish phenotypically different condensates, whilst also separating images that are phenotypically dissimilar. Finally, models that can recognize condensate embeddings well also scored generally higher in our quantitative metrics. In our next steps, we aim to evaluate on additional datasets, and employ modern machine learning techniques to the scarce label problem, such as self-supervised learning.

Weak/Self Supervised Learning for Drug Discovery

Company: Dewpoint Therapeutics

Student: Zhihuan Yu

Industry Supervisor: Étienne Dumoulin, Kshitij Gupta, Grant Watson

Academic Supervisor: Igor Jurisica

Weakly supervised learning is a relatively new technique used in drug discovery High Content Screening to find relationships between compounds or drugs. Models are trained on some auxiliary tasks hence "weak supervision" helping to discover novel mechanisms of action and localizations patterns within cells etc. While these characteristics are not presented to the model during training, they are extremely important for understanding the core biology and function of a drug and greatly accelerate the drug discovery process. In our study, we used several variations of self-supervised models which leverage the contrastive learning framework to learn appropriate representations from single-cell fluorescent microscopy images for the task of Mechanism-of-Action classification.

The proposed work is evaluated on the annotated BBBC021 dataset, which is a publicly available benchmark for mechanism-of-action classification where cells in this image collection have been treated with chemical compounds of known activities. We observed that, without using any labels or post-processing, the self-supervised model achieved similar performance as the weakly supervised model.

Assessing Driving Risk with Enhanced Telematics Data

Company: Geotab

Student: Jiawei Yu

Industry Supervisor: Willem Petersen

Academic Supervisor: Andrei Badescu, Sheldon Lin

With the increasing adoption of telematics data in fields like fleet management and vehicle insurance, understanding and assessing driving behavior and drivers' risks are drawing more and more attention. Most telematics-based evaluation metrics reduce millions of rows of data into a few scalar features, losing useful information in the process. In this project, we adopt the "speed-acceleration" heatmap introduced by Wüthrich (2017), and propose modifications to it, while introducing a similar speed-acceleration heatmap for sideways acceleration, and a contextual-speeding heatmap. Supervised binary classification was done on over 18,000 rental car drivers, a small percentage of whom had incurred damages to the car. We show that compared to models with the usual scalar features, the introduced heatmaps allow models to achieve higher prediction power (0.016 average increase in Cross-Validation AUC-ROC, with standard deviation 0.012).
In addition, the random forest models we trained provide valuable insights into which specific maneuvers are riskier than others. The outputs of the model can also be used to produce "driver safety scores" with very reasonable distribution. Finally, we look forward to improving this work in the future, by utilizing more data and more sophisticated model structures such as self-attention modules.

Reference Wüthrich, Mario. (2017). Covariate selection from telematics car driving data. European Actuarial Journal. 7. 10.1007/s13385-017-0149-z.

Collision Detection with High-Rate Telematics Data

Company: Geotab

Student: Jack Ellis

Industry Supervisor: Willem Petersen

Academic Supervisor: Andrei Badescu, Sheldon Lin

Road safety affects all drivers, passengers, and pedestrians; not just Geotab customers. Telematics data is typically heavily compressed, however Geotab has a data capture logic that can capture higher frequency sampled data that will have the granularity to find interesting acceleration-based events. The objective is to use this to detect vehicular collisions. The methodology behind this project consisted of 5 major stages. First, a complete background research on telematics data and collision detection was completed. This includes learning from domain experts what the telematics data for potential collisions may appear as. Next, features were extracted from the raw telematics acceleration and speed. These features were then used to create a heuristic classification model. Afterwards, machine learning approaches were experimented with. The final stage was evaluation. The heuristic classification model was found to have an accuracy of 92% and a recall of 91%. This was calculated by using hundreds of hand-labeled and known collisions/near misses. The best performing machine learning model was a Deep Autoencoder with K-Means clustering on the encoded space. Clear separation between types of events were formed from the encoding. With higher sampling frequency on the telematics data, it is easier to accurately classify both collisions and near misses. This is evident in the ability to capture low g events. In terms of next steps, more labels would be necessary to have promising models using supervised learning on time-series data, which may have potential to outperform the Deep Autoencoder.

Evaluation of IFIT as an OAM Tool

Company: Huawei Technologies China

Student: Tianyu Wang

Industry Supervisor: Fang Xin

Academic Supervisor: Baochun Li

Operations, administration, and maintenance (OAM) methods are necessary for Internet Service Providers to ensure network quality. In-situ flow information telemetry (IFIT) is one of the OAM performance measurement protocols developed by Huawei to monitor network quality. Unlike traditional performance measurement tools such as ping test that make indirect measurements by sending extra packets. IFIT uses a new type of performance measurement mechanism. It tracks the actual service flow directly by adding an extension header to packets in real traffic which enables it to make measurements with higher precision and smaller granularity.

However, the effectiveness of IFIT and its impact on network traffic haven't been systematically studied. In this project, we built a small test environment to evaluate IFIT and compared it with other commonly used measurement protocols in measurement precisions and impacts on network throughput. Current results show IFIT can reach a constant 0% deviation from the true value when measuring frame loss, but it causes almost 12% of throughput degradation while other protocols nearly have no effect. These results show directions for future improvement of IFIT.

Online Learning and Teaching Math Editor

Company: Hypatia Systems

Student: Ram Gurram

Industry Supervisor: Ladislav Stacho, Jan Manuch

Academic Supervisor: Fanny Chevalier

Creating effective tools to enhance learning online has been a goal many are attempting to enhance. Hypatia System is one of those company's whose mission is to create an innovative platform that enhances the online learning experience for students and for instructors. Hypatia Learn is a tool for teachers to design problems for students, and for students to work through the problem with the ability to check their work along the way. During the COVID-19 pandemic, academic institutions have had to quickly move to online instruction delivery to comply with public health guidelines. As a result, the learning experience has dramatically changed increasing the need for effective tools for online learning.

The goal of our project is to extend a powerful math editor that allows teachers to design elaborate/complex/rich mathematical problems, to expand the capabilities of the system to allow teachers to create problems in calculus. Furthermore, we expand the software to support the correctness checking and student feedback for various calculus computation and provide hints to students when they make mistakes.

In order to expand parsing, we make use of data structures such as abstract syntax trees and design a context free grammar and use an open-source framework which can perform various mathematical computation that is used to perform mathematical validation. Our approach enables the creation of a rich set of calculus problems, such as related rates problems and implicit differentiation. We also enhanced the error/hint reporting system to provide students with warning on their computations, for example we provide warning when students use variables that are not instantiated or unnecessary for solving the problem. This feedback motivates students to write clean and precise solutions to problems.

For future work, we look to investigate how to provide students with insight on how to solve problems by providing hints on next steps they can take to obtain the solution. This would better mimic how a teacher would approach helping a student in class. One way to approach this is to examine the correct solutions and identify where the student's solution deviated from the correct one and provide appropriate guidance targeted at the error specifically.

Designing a Multimodal Prediction Model through Voice Data Mining for Retail Banking

Company: ICICI Bank

Student: Saarthak Sangamnerkar

Industry Supervisor: Sandipan Ray

Academic Supervisor: Mark Chignell

A large majority of organization cutting across industries and domains rely heavily on customer care agents to help them pitch their products to the customers and drive sales. At ICICI Bank, these agents are referred to as Virtual Relationship Managers (VRMs). Currently, every VRM is scored based on various quantitative metrics like number of calls every day, number of tickets generated etc. The primary objective of this project is to develop a machine learning model which incorporates voice call transcripts and the associated Key Performance Indicator (KPI) metadata to evaluate the performance of the VRMs. Some of the KPIs are – a) the conversational pace, b) sentiment tagging, and c) greetings at the beginning/end of the call. This model trains over the free-form call transcripts and the categorical and numerical KPI metadata and gives a binary rating to the VRMs. The addition of the KPI metadata in the learning process has shown an increase in Gini coefficient by 9 percentage points in predicting the performance of the VRM. We intend to keep adding onto the KPI metadata to further improve the prediction model and also, increase the number of prediction classes in the future.

DampingBot: Learning to Stabilize Objects In-hand During High-Speed Manipulation

Company: Kindred AI

Student: Yip Sang Leung

Industry Supervisor: Bryan Chan

Academic Supervisor: Animesh Garg

In Kindred, non-rigid items, such as clothing apparels, are being handled frequently by the robotic arms. The problem of dangling can happen when they are being manipulated at high speed. This can tremendously affect any downstream task like scanning and placement. In this work, we present a reinforcement learning approach to learn an active damping policy. The policy controls a relatively slow robotic arm to stabilize a rapidly moving and underactuated object with a target position. Training is carried out in a simulation that is domain randomized with multiple parallel environments. We randomize relevant parameters in a specific way that facilitates the policy to adapt to the more difficult environments which were otherwise unsolvable. With a highly optimized simulator, we are able to train a policy network using thousands of concurrent environments in one consumer-level GPU within an hour. We demonstrate that the policy can be transferred to reality without any fine-tuning. The policy can stabilize a swinging pendulum 8x faster than without any intervention. Currently, we are planning to replace the pendulum with a polybag in a gripper, so that we can design some downstream tasks closer to practical use cases such as placement. We will then measure the success rate in those tasks to evaluate the performance of the algorithm.

Machine Learning Techniques to Optimize the Performance of an Automated Robotic Arm

Company: Kindred AI

Student: Borys Bryndak

Industry Supervisor: Jan Rudy

Academic Supervisor: Animesh Garg

The main theme of this project is improving a system that controls a robotic arm working in an automated warehouse, primarily trying to resolve some common failure modes like double picks. Previous approaches to this problem involved expensive and error-prone hand-labeling, leading to unsatisfying performance of the supervised methods. We have shown that a weakly supervised training method for a double pick detector is a viable alternative that doesn't require any labeled data. The offline evaluation has shown promising performance of the developed method, justifying the real-world tests to be carried out. Also, we worked on an average reward reinforcement learning approach that would improve the coordination of various systems involved in the operation of the robotic arm, which is under development.

Multi-Object Tracking in Production Environments

Company: Kindred AI

Student: William Ngo

Industry Supervisor: Lavi Shpigelman

Academic Supervisor: Florian Shkurti

Today, the problem of tracking multiple objects is typically solved using deep networks that require frame-by-frame annotations during the training step. This project explores an approach to perform multi-object tracking applied to robotic picking, in cases where real-world frame-by-frame annotations are unavailable.

Instead of training a deep tracker, tracking is done by the paradigm of tracking-by-detection. Tracking-by-detection is performed by having a pre-trained object detector predict objects at each frame and the detections are matched with detections in the subsequent frame. To perform matching, similarity measures between detections are obtained from object characteristics. In this project, an instance segmentation network was used instead of an object detector. As instance segmentation provides a mask for each object, detected objects can be matched based not only on the characteristics of their bounding boxes, but also on the characteristics of their masks.

Although this tracker cannot be directly evaluated using standard tracking metrics as frame-by-frame annotations are unavailable, evaluations of downstream applications of this tracker can be obtained. A downstream application of this tracker is to better infer item classification of grasped items by aggregating tracking information. By tracking objects, classification accuracy is improved to 92.81% (+1.32%) compared to the classification predictions from the segmentation network.

Exploring Visualization and Analysis Methods of Manufacturing Data

Company: Kraft Heinz

Student: Peifeng Zhu

Industry Supervisor: Humberto Consolo Holanda

Academic Supervisor: Fanny Chevalier

The Kraft Heinz company generates massive volumes of rich and complex data through the process of manufacturing food products. It usually takes the management team a great amount of time to collect and analyze these data in order to improve their management. This project aims to design a visualization dashboard which conveys key information to the management team, helping them to compare the manufacturing cost at different times and analyze the sources of cost increase. We first retrieved raw data from the company's database and computed measures that allow to capture the factors associated with total cost increase. Informed by user requirements that we gathered from members of the management team, we designed an initial static dashboard that contains all the necessary tables and charts to support their analytical needs. Through informal, regular usability testing with our target users, we identified limitations and opportunities for improvement. We integrated users' feedback in a revised interactive and dynamic dashboard design that includes additional features that better support users' analytical needs. The user experience has also improved according to users' feedback. As an immediate next step, we plan to conduct a formal evaluation of our prototype through a controlled experiment or deployment study. Possible future directions to expand our work include supporting prediction by machine learning, using additional data to enrich the analysis, or leveraging computation to recommend strategies automatically.

Bag of Little Bootstraps and its Subsample Construction

Company: Layer 6

Student: Taewoo Kim

Industry Supervisor: Kin Kwan Leung, Barum Rho

Academic Supervisor: Rahul G. Krishnan

The bootstrap is a commonly used resampling method to compute quality measures (such as bias, variance, and confidence intervals) of estimators. This method can be easily applied to machine learning models where the quality measure is often hard to compute in closed forms. For example, to measure the uncertainty of a model's prediction, first draw bootstrap samples by randomly sampling with replacement and train ML models for each of the bootstrap samples. Then, the predictions from these models estimate the sampling distribution of the prediction. A major drawback of this method is that it is too time-consuming to train multiple ML models for those resamples when the sample size grows large. The Bag of Little Bootstraps [BLB] is a recently introduced technique to reduce the model's training time by using significantly smaller subsets during resampling. However, the BLB draws the subsets completely randomly and it often requires many resamples to converge to the ground truth.

In this work, we first studied the application of BLB to XGBoost, a popular family of gradient boosted decision trees. Next, we studied the scenario where one has access to a reliable metric that characterizes similarities and distinctions between datapoints. We found out the number of distinct data points in the subsets of BLB is proportional to the accuracy of BLB. To sample more distinct data points in the subset, we use k-means to group similar data points together in a cluster and draw a single point as a representative for that cluster. We empirically show that the confidence intervals acquired by our method achieve a 5% higher intersection-Over-Union rate on binary classification tasks.

Knowledge Modulated Multimodal Sentiment Analysis

Company: Layer 6

Student: Sourav Bhattacharjee

Industry Supervisor: Felipe Perez

Academic Supervisor: Frank Rudzicz

Our world is inherently multimodal and recent work has highlighted the importance of machine learning models leveraging multiple streams of information in making decisions. Multimodal sentiment analysis has been an active area of research that requires models to take advantage of the linguistic, acoustic and visual signals available in an utterance. However, most current models do not consider any social common-sense knowledge which is crucial in how we perceive sentiment in a conversation. To address that, in this project, we modulate modality representations with common-sense knowledge obtained from a generative social common-sense knowledge base. We provide a novel way to modulate the linguistic, acoustic and visual features corresponding to an utterance by scaling and shifting these representations. We use the knowledge base to obtain knowledge latent representations for an utterance corresponding to different states of the speaker such as the intent and the reaction and use it to shift and scale the three modalities. Our experiments on popular multimodal sentiment analysis benchmark datasets show that our proposed method is on par and often surpasses the current state-of-the-art models.

Non-Parametric Options in Offline Reinforcement Learning

Company: Layer 6

Student: Valentin Villecroze

Industry Supervisor: Harry Braviner, Gabriel Loaiza

Academic Supervisor: Chris Maddison

We introduce a non-parametric model for offline learning of options in reinforcement learning. Our method allows us to modify the number of options in a data-driven manner, by using Bayesian non-parametric methods with a GEM prior, and variational methods. We demonstrate empirically that our method can perform on par with state-of-the-art offline option learning algorithms while not having to fine-tune the number of options as an hyperparameter.

Text-to-Video Cross-Modal Attention in Retrieval

Company: Layer 6

Student: Noel Vouitsis

Industry Supervisor: Guangwei Yu

Academic Supervisor: Animesh Garg

In text-video retrieval, the objective is to learn a cross-modal similarity function between a text and a video. The purpose is to use similarity to rank videos given text queries or to rank texts given video queries. However, videos inherently express a much wider gamut of information than texts, so a text generally cannot fully capture the entire contents of a video. Instead, texts are most semantically similar to subsets of videos, which means there are multiple equally valid texts that can plausibly match a particular video. To mimic this behaviour in retrieval, a model should only examine the most similar video regions described in a text to extract the most relevant information. Yet, most existing works aggregate video frames in time without directly considering text, such as through mean-pooling or self-attention. As such, we have designed a cross-modal attention model to learn to reason between a text and video frames. We adapted a scaled dot product attention for a text to attend to its most similar frames and then we generated an aggregated video representation conditioned on those frames. Our method was built on top of a pre-trained CLIP model designed to match texts and images, following the success of recent works. We evaluated our method on the benchmark retrieval datasets of MSR-VTT, MSVD and LSMDC, and obtained state-of-the-art results across datasets. Our findings thereby highlight the importance of joint text-video reasoning to extract important visual cues according to text. Our next steps are to examine the robustness of our model to noisy and irrelevant frames through experimental analysis. As part of future work, we plan on applying our method to other text-video tasks such as visual question answering.

Reducing Popularity Bias in Recommender System

Company: Loblaw Digital

Student: Ruijian An

Industry Supervisor: Richard Downe

Academic Supervisor: Scott Sanner

In online shopping, recommender systems are algorithms that recommend relevant products to customers based on their purchase history. Many works of literature have pointed out that a recommender system tends to recommend popular products. However, an ideal system should not only consider popularity bias but also provide personalized recommendations to customers. To provide better email recommendations at Loblaws, we aim to de-popularize our recommendations and help customers discover new products.

In this project, we implemented a new recommender system called NCE-PLRec (Noise Contrastive Estimation-Projected Linear Recommendation). The model has two components: noise contrastive estimation and projected linear recommendation. To deal with popularity bias, noise contrastive estimation reweights the user-item interaction matrix and explicitly de-popularize item embeddings. To scale the recommender system to millions of users, projected linear recommendation is implemented, which provides a highly scalable closed form solution.

We compared the quality of the Top-200 items recommended by NCE-PLRec and the current model in production in terms of mAP (Mean Average Precision), precision, recall, and NDCG (Normalized Discounted Cumulative Gain) in offline evaluation. This is followed by online A/B testing measuring CTR (click-through rate) and weekly GMV (Gross merchandise volume) between the two models. The A/B testing showed a 12% increase in CTR and an increase in weekly GMV. As a result, NCE-PLRec provides a better recommendation experience for customers to discover new products and handles the cold-start problem for products with little historical data. In the next step, we'll enhance NCE-PLRec by incorporating side information of products and replenishment behavior of customers.

Application of Deep Learning Algorithm for Medical Images Based on Low-Dose Computed Tomography for Lung Cancer

Company: Lunenfeld-Tanenbaum Research Institute, Sinai Health

Student: Chenyang Lu

Industry Supervisor: Rayjean J. Hung

Academic Supervisor: Babak Taati

Low-dose computed tomography (LDCT) screening of lung cancer has demonstrated 20% to 40% lung cancer mortality reduction. Hence, there are many management protocols for these pulmonary nodules with indeterminate malignant potential from watchful waiting, biopsy to surgical resection. We developed a model to provide a highly robust and efficient platform to automatically detect the pulmonary nodules on the LDCT scans and conduct segmentations to facilitate the downstream analysis that includes patient factors. For nodule detection, we have the data marked by professional radiologists and propose a 3D nodule detector framework using 3D U-net as the backbone. Next, we trained another classifier with 3D CNN to determine the false-positive cases. For the future, we want to build a platform for radiologists to quickly apply the model in real use and apply new models to the task.

Enterprise Knowledge Mining with Multi-Task Self-Supervised Graph Neural Networks

Company: Microsoft Turing

Student: Hao Zhang

Industry Supervisor: James Vuckovic

Academic Supervisor: Sushant Sachdeva

Graph neural networks (GNNs) have demonstrated the ability to leverage the structural information in graph relational data for both self-supervised and semi-supervised learning; access to this structural information often yields better performance on tasks such as node classification, graph classification, and edge prediction. This report explores GNN methods for combining structural and semantic information to produce representations for documents, users, and topics that occur in enterprise data. These representations are trained in a multi-task, self-supervised setting to capture general-purpose information about each node for use in various downstream applications. Using a custom automatic evaluation setup, this report shows that these representations yield higher related topic, document, and user retrieval metrics than a strong baseline averaging algorithm. Since the self-supervised learning scheme is not specific to related node retrieval, these results imply the representations can be leveraged for additional downstream tasks.

Model Compression for Transformer-based NLG Models

Company: Microsoft Turing

Student: Zhu-Mu Chen

Industry Supervisor: William Buchwalter, Eric Lin

Academic Supervisor: Eyal de Lara

The recent success of large-scale transformer-based language models has exacerbated the need for neural network model compression. Most proposed methods for NLP are not examined on natural language generation (NLG) models. Despite the structural similarity, transformer-based NLG models tends to be more sensitive to compression than their variants for natural language understanding (NLU) due to the higher complexity of the task. In this project, we discover several challenges while applying quantization and pruning. For quantization, naïve quantization-aware-training introduce excessive noise to the model and prevent convergence. For sparsity, unstructured pruning suffers from huge storage overhead while structured pruning results in severely degraded performance. We developed a one-shot compression procedure, integrating variational quantization, stratified embedding for sparsity, and distillation. Furthermore, for quantization, the application of range reduction and several noise control techniques significantly reduces quality loss; for sparsity, introducing the information of token frequency improves the quality of sparsity pattern. We evaluated our method on an internal large-scale NLG model in Microsoft Turing with a structure like GPT-2. The proposed method greatly enhances the compressed performance by approximately 35%, nearly lossless, compared to the previous best baseline model with dynamic QAT. Our on-going works focus on automatizing the design of sparsity pattern.

Production-Ready Federated Learning

Company: Microsoft Turing

Student: Grigory Dorodnov

Industry Supervisor: Justin Harris

Academic Supervisor: Aleksandar Nikolov

Federated Learning (FL) is a new and rapidly developing area of private data analysis taking advantage of distributed optimization without any centralized data collection. Even though the current state-of-the-art has already achieved some level of maturity, it is still far from being universally applicable in the production setting. The major causes of that are high unpredictability of the overall optimization process and of the user behaviour. Specifically, data scientists currently have limited ways to ensure high-quality result and robustness against outlying (possibly, malicious) model updates.

To reduce the uncertainty in the FL process, we have developed a flexible simulation framework, allowing data scientists to experiment with the federated environment and the training process by varying the number of available participants, their behaviour, and the heterogeneity of their local data. The framework can assess the trade-offs of introducing privacy-preserving techniques, such as Differential Privacy and Secure Multi-Party Computation as well. Making use of our framework and the Microsoft News Dataset (MIND), we have explored the feasibility of training an article recommendation model in a federated paradigm. Based on our findings, we have developed a production-ready prototype and integrated it into one of the company's products.

We believe data poisoning prevention to be the next step in increasing applicability of FL. We are interested in preventing model performance degradation resulting from users' malicious activity using computational techniques and incentivization.

Simultuning: Simultaneous Finetuning for Contradiction Classification and Conditional Dialogue Generation

Company: Microsoft Turing

Student: Kyle Oppenheimer

Industry Supervisor: Adam Atkinson

Academic Supervisor: Frank Rudzicz

In recent years, we have seen tremendous progress in open-domain dialogue systems through the use of large pretrained transformers. However, state-of-the-art natural language generation (NLG) models used for dialogue systems still struggle to remain logically consistent. On the other hand, natural language understanding (NLU) models fare quite well at identifying contradiction. We bridge the gap between NLU and NLG models via simultaneous finetuning (simultuning) for a) contradiction classification and b) dialogue response generation conditioned upon an in-text label indicating whether the model's response should contradict the context. In this way, the model can leverage its understanding of contradiction at the time of generation to reduce the occurrence of contradictory responses. We demonstrate that simultuning is an effective approach for both NLU and NLG, as we achieve state-of-the-art results on the dialogue contradiction detection task (DECODE), while maintaining the ability to generate high quality responses. We also demonstrate that this method can be used to prompt models to generate responses in a certain style as we were able to generate contradictory responses with up to 50% accuracy (judged by humans). Work is ongoing to determine whether this method can successfully reduce the occurrence of contradiction at the time of generation.

Design Choices for Fast and Accurate 3D Face Reconstruction

Company: ModiFace

Student: Saad Saleem

Industry Supervisor: Brendan Duke

Academic Supervisor: Sven Dickinson

3D face reconstruction from a single image is an ill-posed problem due to the depth ambiguity in the face shape. State of the art methods reduce this ambiguity by using priors on the face geometry, such as 3D Morphable Models (3DMMs).

Lightweight regression models can be trained to predict the 3DMM parameters of a face from an image in real time. However, they require data labelled with facial landmarks or 3DMM parameters for training, which is not easily available nor reliably accurate.

Other models encode the image to the latent space of face and image features then reconstruct the input image using differentiable renderers and minimize a formulation of photometric loss on the reconstructed image. These models can be trained with more unlabelled data and achieve better reconstruction accuracy, but they can't provide real-time inference. We aim to train a lightweight model using these self-supervision methods while also providing real-time inference.

We have observed that state of the art accuracy is achievable with the regression models. Next, we will compare the accuracy and speed of photo reconstruction models to these results. Also, we will study the effect of training with synthetic data on the performance of these models.

Hardware Aware Neural Architecture Search for Semantic Image Segmentation

Company: ModiFace

Student: Shamitra Rohan

Industry Supervisor: Vicky Yu

Academic Supervisor: Lueder Kahrs

Hair color simulation is an important Augmented Reality task at ModiFace. It is fundamentally based on semantic segmentation, the pixel-level identification of classes in an image. Since our models are deployed on platforms with limited computing capabilities, inference speed and performance are key considerations during the model design process. Current methods involve tediously handcrafting Convolutional Neural Networks to achieve desired metrics. In this project, we use Neural Architecture Search (NAS) to find faster or more accurate networks on the semantic segmentation task.

We adopt a One-Shot NAS method, which trains a supermodel while fine-tuning the parameters of smaller networks and maintaining performance. This contrasts with early NAS methods, which require repeated training of sampled models for different scenarios, thus being computationally prohibitive. Model performance is measured using the mean Intersection over Union (mIoU) metric. In our experiments, a baseline model architecture consists of a preset backbone and the same segmentation head as the supermodel. On our in-house hair dataset, a searched model with 33% fewer FLOPS and 23.7% fewer parameters than our deployed model almost matches its accuracy: 90.5% versus 91.4% mIoU. Cityscapes is a public dataset comprising street scenes from various cities and a popular benchmark for segmentation tasks. On Cityscapes, the supermodel achieves 60.4% mIoU while the closest baseline to it is EfficientNetV2 with 58.9% mIoU.

The next steps involve continuing the evolutionary search process to explore and profile candidate models with better capacity, accuracy, and efficiency. Through this project, we demonstrate the usefulness of One-Shot NAS in generating segmentation models, which will enable the efficient design of specialized models for different platforms at ModiFace.

Learning Controllable Directions of GAN Space

Company: ModiFace

Student: Zikun (Shelly) Chen

Industry Supervisor: Irene Jiang

Academic Supervisor: Babak Taati

Generative Adversarial Networks (GANs) are capable of generating highly realistic images given random inputs from a latent space, where meaningful directions and channels have been shown to exist, and linearly interpolating in these directions or changing individual channel values result in interpretable transformations, e.g., adding eyeglasses or smile to a person's face. However, controls obtained by supervised approaches require large amounts of labelled data and are sometimes entangled, yet unsupervised settings involve manual examination of many different manipulation directions and identification of meaningful controls.

In this work, we propose a gradient-based approach that learns controllable directions of GAN latent space. Our approach utilizes the gradient directions of auxiliary networks to control semantics in the latent codes, which requires minimal amounts of labelled data with sizes as small as 30 samples that can be obtained quickly with human supervision. Grad-CAM is a form of post-hoc attention, which visualizes where a convolutional neural network model is looking. We propose to select important latent code channels with Grad-CAM based masks during manipulation, resulting in more disentangled controls.

Flickr-Faces-HQ (FFHQ) is a benchmark dataset for GANs, consisting of 70000 high-quality PNG images of human faces at 1024x1024 resolution. Our qualitative experimental result for manipulating outputs of StyleGAN trained on the FFHQ dataset shows that our approach achieves an overall better performance and disentanglement than the state-of-the-art supervised method using support vector machines (SVMs). In the remaining terms of research, we will focus on quantitative evaluations, manipulation of real images using GAN inversion and application of our method to a wider range of datasets.

ATPM: Evaluation Platform and Advanced Solutions for Autonomous Transportation Leveraging 5G Networks

Company: New Vision Systems Canada

Student: Tao Wu

Industry Supervisor: Hossam Gaber

Academic Supervisor: Bo Wang

Autonomous driving has long been favored by most manufacturers and researchers in the automotive industry. On top of traditional self-driving cars which rely on on-board detecting and computing, the technology of Connected and Autonomous Vehicles (CAV) has recently been proposed with the assumption of ultra-reliable low-latency communications. However, whereas the upcoming 5G networks have set the stage for cellular Vehicle-to-Everything (Cellular-V2X) connections, CAV and autonomous transportation has not yet been widely investigated. Therefore in this work, we build the system named Autonomous Transportation Planning and Management (ATPM). By adopting the Cellular-V2X standard to connect vehicles, pedestrians and roadside infrastructures, we develop advanced solutions for a variety of traffic scenarios to modernize the mobility management taking advantage of quick data transfer. To demonstrate the efficiency and safety our system brings, we also created an evaluation platform based on virtual street simulators, which helps test and optimize the transportation planning at low experiment cost. The result of a case study we did in Toronto showed that our transportation planning increased the highway capacity by 9.0% with a reduced collision rate. In future work, we plan to develop smart apps for mobile devices of drivers, pedestrians and all other road users, to further enhance mobility and safety.

GANVerse3D: Single Image to 3D Object Reconstruction

Company: NVIDIA

Student: Ao Tang

Industry Supervisor: Masha Shugrina, Clement Fuji Tsang

Academic Supervisor: Shurui Zhou

To infer 3D properties such as geometry, texture, and 3D segmentation map from one single 2D image is a difficult inverse graphics task. Most of the current approaches focus on training such networks using multi-view synthetic datasets, while it is infeasible to find such large-scale-real-image datasets in practice. Therefore, these approaches fail to perform well when testing on real images due to the noticeable domain gap between synthetic and real images.

We utilize the Generative Adversarial Networks (GANs) to generate realistic-looking images for training. We use the official pretrained StyleGAN2 model provided by another research team in NVIDIA. By analyzing and disentangling the latent space, we successfully transform StyleGAN2 into a multi-view image generator. We train the 3D prediction network using a differentiable renderer, and the entire architecture is iteratively trained using cycle consistency losses.

We further improve our approach by collecting a new dataset for training StyleGAN2, resulting in a significant visual improvement in the quality of the generated image. We also encode viewpoint information into latent space and adapt the architecture of StyleGAN2, which allows the StyleGAN2 network to generate controllable multi-view images.

We show that our model can perform well on both real images and synthetic images, leveraging the 3Dreconstructing network to another level. As an ongoing project, we aim to further leverage the prediction fidelity of our current model by improving the multi-view consistency of generated images from StyleGAN2. We also aim to train the pipeline on multiple categories.

eXplainable AI Applied to Pharmaceutical ML Model

Company: ODAIA

Student: Cheng Han Hsieh

Industry Supervisor: Zhou Fang

Academic Supervisor: Dehan Kong

ODAIA is a startup dedicated to understanding customer journey from health care professions and enabling pharma sales to dynamically target the next best audiences. Using the cutting edge machine learning technique, ODAIA makes a trend prediction and thus provides broader customer insights. To better understand which customer features (e.g., age, region) take leading components in the prediction model, we implemented Explainable artificial intelligence (XAI), which aims to disassemble the black-box inside machine learning models, on our business scenario.

The whole project has two parts: (1) Feature Importance measures the relative contribution of each customer feature. For example, does region feature have greater influence than age feature in terms of drug prescription volume prediction? Our experimental approaches are trying to answer such question and based on two different methods. Local feature importance describes how customer features affect the prediction in individual level and gives us a better understanding to a specific customer; while global feature importance does the same thing in high-level interpretation on the whole prediction model. We compare those two methods to generalize a conclusion for most important features. (2) Counterfactual Explanations (CF) explores how the small change in the customer features can achieve desired prediction outcome. For example, how much income increase for this customer would lead to double market share prediction? By constraining changeable and non-changeable features, we have successfully generated some useful CF examples with different feature settings.

In the next step, we hope to utilize more robust feature importance methods to further generalize top critical features. Also, we will keep exploring potential business scenario under powerful support from XAI.

Speech Enhancement and Recognition with Generative Adversarial Network

Company: Pearson Canada

Student: Zibin Yang

Industry Supervisor: Richard Wang

Academic Supervisor: Gerald Penn

We measure the effectiveness of the Generative Adversarial Network (GAN) for noisy speech enhancement and recognition. GAN can work end-to-end, with the raw audio waveform as inputs and targets. Different from the prior work [1] by evaluating the quality and intelligibility of enhanced audios, we investigate the recovery of the Word Error Rate (WER) lost in the noisy speech recognition context. Besides, most of the automatic speech recognition (ASR) systems are trained on 16 kHz sampling rate audios and above. We conduct the study on an 8 kHz Pearson internal audio dataset from non-native English speakers. First, we train SEGAN [1] by directly adding the real-world noise from the DEMAND dataset [2] to clean audios with various speech-to-noise ratios (SNR) as input features and clean audios as targets. Inspired by [4], two PyTorch-Kaldi ASR systems [3] are trained (Model-1, Model-2). The Model-1 uses both the clean and noisy audios as input. For Model-2, we use the same audios from Model-1 and pass the noisy audios through SEGAN to generate their enhanced version to train the ASR system. Model-2 reduces the WER from 34.9% to 34.2% compared to the Model-1 on the testing noisy speech. The ASR model will help us to generate more accurate transcripts for the noisy inputs and benefit the downstream tasks like scoring. At the current stage, the model is built on a relatively small dataset (25h hours of audio). More clean and noisy speech together with clipped and low volume audios will be introduced to further improve the ASR system performance for bad quality speech in the future.

References:

[1] S. Pascual, A. Bonafonte, and J. Serr`a, “SEGAN: Speech Enhancement Generative Adversarial Network,” in Interspeech, 2017.

[2] Thiemann, Joachim, Ito, Nobutaka, & Vincent, Emmanuel. (2013). DEMAND: a collection of multi-channel recordings of acoustic noise in diverse environments (1.0) [Data set]. 21st International Congress on Acoustics (ICA 2013), Montreal, Canada. Zenodo. https://doi.org/10.5281/zenodo.1227121

[3] Ravanelli, Mirco et al. “The PyTorch-Kaldi Speech Recognition Toolkit.” ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019): 6465-6469.

[4] K. Kinoshita, T. Ochiai, M. Delcroix, and T. Nakatani, “Improving noise robust automatic speech recognition with single-channel time-domain enhancement network,” in 2020 ICASSP, 2020, pp. 7009–7013.

Learning Tone Curves for Local Image Enhancement

Company: Samsung AI Centre

Student: Luxi Zhao

Industry Supervisor: Michael S. Brown, Abdelrahman Abdelhamed

Academic Supervisor: Kyros Kutulakos

Contrast and colors are important factors influencing an image's perceptual quality. Image enhancement techniques aim to adjust these factors to achieve more aesthetically pleasing results. Many data-driven approaches formulate image enhancement as an image-to-image translation task. These methods directly output an edited image, providing little insight on what transformations were performed. We propose a new method to perform automatic local image enhancement for smartphone cameras. Instead of learning the output image itself, our neural network predicts a grid of transfer functions, specifying the mapping from each possible input pixel value to its corresponding output value. These transfer functions are known as "tone curves". The predicted tone curves are then applied on corresponding patches of the input image with tile-based interpolation to produce smooth and artifact-free results. Compared with image-to-image translation models, our model is advantageous in a smartphone camera setting because tone curves are commonly used within image signal processors (ISPs), thus can be easily integrated with them. Experiment results demonstrate that the proposed approach outperforms global image enhancement methods and reaches similar performance as the state-of-the-art local enhancement methods in terms of perceptual metrics.

Learning-Based Obstruction Removal

Company: Samsung AI Centre

Student: Youheng Ge

Industry Supervisor: Alex Levinshtein

Academic Supervisor: Kyros Kutulakos

Images captured by cameras may contain unwanted obstructions, such as fences, reflections and raindrops. Modern approaches often exploit multiple image frames, taken with a slightly moving camera (e.g., cell phone camera), to reconstruct a clear, obstruction-free version of the scene in the background. Compared to traditional multi-frame approaches, learning-based methods require less computationally expensive optimization, but they often fail in real-world sequences. Our fence-removal work builds on a recent learning-based approach that alternates between estimating dense optical flow fields of background layers and reconstructing background from the flow-warped images via a deep convolutional neural network, in a coarse-to-fine manner. To improve both quality and speed, we first port this work from TensorFlow to PyTorch and re-implement the synthetic data generation process. We further improve the data synthesis pipeline by applying random color and out-of-focus blur data augmentation. Our experimental results show that our model reconstructs clearer background and yields fewer artifacts compared to the pre-trained baseline. Besides, our implementation operates at 1.6x the speed of the baseline. In the future, we plan to extend the method to reflection and raindrop removal, and integrate the best of single- and multi-frame obstruction removal methods.

Referring Expression Understanding in Navigation

Company: Samsung AI Centre

Student: Guanxiong Liu

Industry Supervisor: Afsaneh F., Isma H., Ran Z., Federico F.

Academic Supervisor: Frank Rudzicz

Applications of robotics continue to grow rapidly, especially with the advancement of machine learning. Consider giving the robot a set of commands: “Face the table. Second, find a deck of books on the table. My glasses are placed above the books.” It is hard for an AI robot to understand this sequence of instructions because the system needs to ground the words with visual perception. The robot requires the locations of referenced items and the field of views of the human. It also requires several intermediate instructions to describe a path in a 360° panorama. In this work, our goal is to find the final location of an object given a sequence of referring expressions. To solve this problem, we treat it as a sequence of image retrieval problems. We split a panorama into several image regions and use referring expressions to retrieve each corresponding target region. We test two models and their variants with a panorama dataset to obtain the performance of the models. Our models perform reasonably well given the difficulty of the problem. The next step would be to further improve the success rate in order to build a fully functional robot assistant that can retrieve objects as requested by a human.

Web Page Contextual Subcomponents Recognition and Identification

Company: Scrawlr

Student: Kexin Yan

Industry Supervisor: Matija Boban, Jonathan Shahen

Academic Supervisor: Shurui Zhou

A web page typically contains many subcomponents such as navigation bar, advertisement, and paragraphs. However, not all subcomponents provide useful information. Thus, the definition of the informative contextual subcomponent is proposed, representing a subcomponent that provides useful information to the user, such as sentences or paragraphs in the main content. The goal of the project is to implement a browser extension, which can identify informative contextual subcomponents on the webpage to improve the efficiency of interaction between users and web pages. The designed extension contains two main features: manually highlight and provide recommended highlights. There are already some tools such as Hypothesis and LINER that can help user create highlights manually, but there is no existing extension that can recognize informative subcomponents according to the content of the webpage to create recommended highlights to users. Currently, our extension is tested on five major categories of web pages, including news articles, documents, recipes, Wikipedia articles, and educational articles. The test result shows that the user can select any sentences on these web pages to create highlights, and all recommended highlights created by heuristic rules provide helpful information to the user. Furthermore, the proposed approach will be continuously improved when more and more web pages are involved.

Structured Information Extraction in Clinical Texts Using Graph Model

Company: Semantic Health

Student: Kai Wang

Industry Supervisor: Michal Malyska

Academic Supervisor: Michael Brudno

Medical coding is a process of identifying diagnoses and procedures in medical notes and transforming them into universally recognizable alphanumeric codes. With highly unstructured medical notes and thousands of possible codes, the training of any deep learning algorithms will be noisy. In this project, we used a knowledge graph approach to improve the recall in this extreme multi-label classification problem without restricting our label space. We first used named-entity recognition to pick up medical terms from the unstructured text. The identified terms were then linked with terms in an existing medical ontology to populate the patient graph with relationships provided by the ontology. In the end, the ontology codes were mapped to disease codes using the ontology native code conversion schema. Our model achieved 32% recall on full disease codes and 57% recall for the root interventions[1] without restricting the label space and was similar to the BERT-based baseline. However, the model did not achieve production-level precision, leaving room for improvement.

Enriched Understanding of Retail Receipts for Personalized Financial Insights

Company: Sensibill

Student: Elisa Du

Industry Supervisor: Daniel Wagner

Academic Supervisor: Michael Guerzhoy, Rohan Alexander

Sensibill is offering an agile spend management platform that digitizes retail receipts and invoices through its proprietary Optical Character Recognition (OCR) technology. The extracted attributes from receipts can then be used to generate personalized financial recommendations and spend insights for customers of banks, helping increase customer engagement and improve financial wellness.

To improve personalization, there needs to be granular categorization of receipt line items. The unique challenge lies in the fact that receipt line items are short, abbreviated, and prone to OCR errors, making the derivation of financial insights difficult. To overcome this challenge, item normalization – or standardizing line items so that their most comprehensible form is displayed – is necessary.

The project focuses on optimizing and developing an evaluation flow for the existing line item embedding-based model tuned for both enrichment objectives: item normalization and categorization. In this unsupervised approach, the model extracts the most normalized form of an item by grouping historical line items from our database into clusters through various pruning criteria that reduce the number of clusters and improve computational efficiency at inference time. In training, we found that pruning clusters using an embedding-based approach in place of text similarity resulted in clusters that encode more relevant item-item context, as well as more efficient search and retrieval of the normalized item during inference time. We also assigned confidence thresholds tuned for each enrichment objective during inference time, with optimal values determined through a series of classifier experiments. To measure model performance, we developed a set of custom quantitative and qualitative metrics. The optimized model demonstrated a boost in item normalization accuracy, with no decrease in item categorization accuracy on the benchmark dataset.

The optimized model has been deployed in the production environment. We are continuing to explore optimization strategies for the line item embedding model to increase the coverage of items enriched.

Object Detection and Image Segmentation for Receipt Images

Company: Sensibill

Student: Danting Dong

Industry Supervisor: Daniel Wagner

Academic Supervisor: Lueder Kahrs, Michael Guerzhoy

The goal of the project is to extract high-quality receipt images from user-uploaded photos to increase the accuracy of Sensibill's process for extracting purchase information from receipts. Sensibill provides machine-learning-powered financial tools for receipt images that help banks and credit unions turn customer data into actionable insights in order to know and serve their customers better.
In this project, a new dataset was collected, and several neural-network-based methods for receipt extraction were implemented and evaluated. We improved the performance over the baseline and evaluated object detection frameworks based on Tensorflow Lite[1], and image segmentation frameworks based on Mask R-CNN[2], U- Net[3], and Deeplab v3[4]. The evaluation results of fine-tuned receipts extraction methods were based on our test set. The metric was the proportion of the photos for which the bounding boxes or image masks were correct. The U-Net model successfully detected receipt objects in 86% of testing photos. The evaluation results for Mask R-CNN and Deeplab v3 are 45.6% and 67.2% respectively. Besides that, a future evaluation would be done for AWS Textract[5], a commercial service that can be used for extracting text from receipts.

[1] Li, S. (2020). Tensorflow lite: On-device machine learning framework. Journal of Computer Research and Development, 57(9), 1839.
[2] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
[3] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[4] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.
[5] Announcing specialized support for extracting data from invoices and receipts using Amazon Textract | Amazon Web Services. Amazon Web Services. (2021) https://aws.amazon.com/blogs/machine-learning/announcingexpanded- support-for-extracting-data-from-invoices-and-receipts-using-amazon-textract/.

Achieving Clinical Automation in Pediatric Emergency Medicine with Machine Learning Medical Directives

Company: SickKids

Student: Xinqi Shen

Industry Supervisor: Devin Singh, Erik Drysdale

Academic Supervisor: Yun William Yu

In emergency departments, increased patient volume, wait times and long stays impose bad hospital-visiting experiences to the patients, as well as common medical challenges to the hospital. In order to provide guidelines for nurses, medical directives can request certain diagnostic tests at triage, which expedite the process by providing test results to the physician at the initial encounter. The goal of this project is to propose a machine learning medical directive approach to autonomously ordering downstream tests such as ultrasounds, X-rays, etc.

Our solution used the triage, testing and diagnosis data from patients' electronic health records to train the logistic regression, random forest, gradient boosting and deep neural network models, then make predictions. We also constructed a retrospective model-training pipeline to validate the safety of model performance by performing error analysis and biases assessments. In doing so, we found the best hyperparameter settings for each downstream test. Models obtained high area-under-operator-curve(0.9-0.99) across each of the use cases. The result demonstrated the feasibility of using our approach to achieve clinical automation. The next step of the project is to deploy models into a real-time prospective pipeline and propose a self-auditing function to monitor performance metrics.

HostSeq COVID-19 Data Visualization

Company: SickKids

Student: Shikai Liu

Industry Supervisor: Lisa Strug

Academic Supervisor: Michael Brudno

The CGEn HostSeq project is funded by the Federal Government to sequence approximately 10,000 Canadians infected with SARS-Cov-2. The goal of this project is to build a data visualization tool for CGEn HostSeq databank. Data visualization is an active research area, which focuses on methodologies and tools (e.g., Tableau) to deliver the information from the data. The visualization tools referred to as data portals will deliver information to researchers (and the public) to facilitate understanding of what is available in the HostSeq databank and enabling them to design and conduct future research studies. The HostSeq databank is comprised of both the clinical data and the genetic sequence of each participant. Therefore, the team has implemented two data portals. The portals are written using Python and Flask frame. As of October 2021, the clinical data portal has been developed, with approximately seventy features available to be visualized and summarized. The genetic data portal is still under development, but will be comprised of displays and summary statistics from genome-wide analysis of SARS-CoV-2 infection severity and other related outcomes. Displays will include Manhattan plots of association statistics, Quantile-Quantile displays, and others. In the current project, the backbone of the visualization tool has been developed. As more data becomes available, we will present more features and include larger sample sizes in the data portals. Lastly, linking the data portals with the multi-omics statistical integration tool developed by our team, “LocusFocus” will enable an improved user experience for visualizing and interpreting the genetic results.

Continuous Mixout: Preventing Forgetting Problem of Large-Scale Language Models in Continual Learning

Company: SOTI

Student: Varun Pandya

Industry Supervisor: Hossein Taghinejad

Academic Supervisor: Scott Sanner

Just like the human brain, Large Scale Language Models also suffer from catastrophic forgetting. These Language Models (LMs), when fine-tuned on smaller domain specific data for downstream tasks, often forget their pretrained knowledge due to the smaller number of training instances available. The combination of this scenario with continual learning, where the model learns new knowledge iteratively on a small number of training instances as they become available, causes further degradation in performance. To mitigate this catastrophic forgetting problem in LMs, this paper proposes a new hybrid method, where the combination of stochastically mixing the parameters of two models (Mixout Regularization) with geometric decay of learning rate through the layers of the model with the AdamW optimizer mitigates the forgetting problem. The results are shown for the DistilBERT model across three datasets which suggests that compared to a simple retraining method which results in an accuracy drop anywhere between 2% to 25% across various iterations, the hybrid method not only prevents these accuracy drops, but also increases the baseline accuracy in some cases. The mitigation of the forgetting problem by this hybrid method facilitates the future use of Large LMs for Multi-Task Learning as well as Continual Learning setups.

Creating and Analyzing the No-Code Platform for Generating RESTful APIs for Database Manipulation

Company: SOTI

Student: Yuan Chen

Industry Supervisor: Sheldon Davis

Academic Supervisor: Marsha Chechik

Low-code and no-code application development platforms (LCAPs and NCAPs) improve the ease with which business applications can be delivered, requiring smaller and less specialized teams of developers than traditional code-based application development. For example, SOTI SNAP, a NCAP for cross-platform mobile application development, provides a drag-and-drop interface to reduce the accessibility barrier for users without prior programming experience. One of the issues in building a general purpose NCAP is the integration of user-friendly data entry and retrieval with corporate databases. In this project, we created an NCAP CRUD API tool to create and manage RESTful API services that allow other platforms to perform Create, Read, Update and Delete (CRUD) operations on a wide variety of relational databases. Prior internal analysis estimated that integrating NCAP CRUD API tool into SOTI SNAP will save roughly 15% of a typical application development time.

Throughout the project, we have done multiple rounds of usability testing. Prior to the implementation, several rounds of focus groups have been conducted within the organization to ensure usability. Among others, participants emphasized the need for documentation, ease of use, functionality, security, consistency with related products, simple navigation, familiar terminology for a non-technical audience, effective on-boarding, error diagnosis and prevention, and visibility of system status. Then, a moderated remote usability study with 10 participants was conducted on a preliminary version of the user interface, aimed to measure the above metrics. The study had three phases: explorative testing, usability testing with simulated tasks, and a post-study questionnaire. The analysis of the results is still in progress, and we aim to use them to assess the usability of the tool and to suggest areas for improvement.

Mislabeled Data Detection Pipeline for Snap Analytics

Company: SOTI

Student: Xinming Ye

Industry Supervisor: Hossein Taghinejad

Academic Supervisor: Scott Sanner

Snap Analytics allows users to perform database queries via natural language inputs. To convert natural language to SQL queries, multiple BERT-based models are used to learn the relationship between natural language questions and their corresponding SQL queries. However, the presence of incorrectly-labeled data can significantly degrade the models' performance since they cannot learn the correct text-to-SQL mapping. Therefore, as an important step towards good model performance, my project focuses on ensuring data quality by detecting and removing mislabeled data. For this purpose, the pipeline detects mislabeled data by combining different metrics based on the model's training dynamics as well as performing a semantic match between input and label. First, the pipeline trains the language model on the data to calculate the confidence and AUM score of each data sample. Additionally, a hybrid approach of Computational Linguistics and Universal Sentence Encoder is applied to look for any semantic match between the data and the label. As a result, the recall of this approach is 1.0 and the precision is 0.5 - 1.0 depending on the ratio of mislabeled data. This means that the pipeline captures all the mislabeled data, and the model can maintain its performance instead of suffering an accuracy drop of 7 - 15%. Moreover, a conversation pipeline is also integrated so that users can interact with the Snap Analytics system. A point for future improvement is finding other metrics that can better identify whether a data is correct or not.

UAV Onboard Real-Time Scene Classification

Company: SOTI

Student: Yunze (Peter) Pan

Industry Supervisor: Dmitry Shesterin

Academic Supervisor: Anthony Bonner

SOTI is in researching advanced intelligent visual systems for unmanned aerial vehicles (UAV). UAVs require indoor scene classification for navigation since it provides the drone with the capability of distinguishing different scenes during navigation. Recently, convolutional neural network (CNN) based methods show a significant advantage of image classification tasks, whereby the deployment of the methods requires large computing to complete. However, to deploy on a UAV, a real-time method that is based on a compact model is desired.

In this project, a new CNN-based scene classification approach is being proposed for SOTI's Aerospace UAV. Our method is based on the EfficientNetV2 backbone, which contains a customized classifier trained with large-scale datasets, including ImageNet and Place2. To acquire a robust use case, the preparation and undergo training has been upgraded to a customized data augmentation pipeline. Subsequently, the new method achieved a better testing result compared to the current real-time models. Reaching 85% accuracy on the MIT67 benchmark and 15ms latency when deployed on the NVIDIA Jetson Xavier NX board.

The next step will be optimizing and integrate this method into SOTI's current UAV vision system. To further consolidate the models, one potential approach is by using a transfer learning-based method. Combining the scene classification model with the existing object detection model.

Workflow Recognition of Surgical Videos

Company: Surgical Safety Technologies

Student: Yi Xiang

Industry Supervisor: Frank Rudzicz

Academic Supervisor: Maryam Mehri Dehnavi

Computer-assisted surgery (CAS) systems have been growing in impact in modern surgery. Such systems provide automated service to surgical education, surgeon performance evaluation, and operating room workflow optimization. One critical component of CAS is surgical workflow recognition, which identifies the operational phases through surgery. Such task was often done manually in the past which is time-consuming. People tried to automate the process, but the proposed methods did not investigate the generalization over different surgery types. To address the issue, we propose a deep learning method for automatic workflow recognition from surgical videos, which learns the mapping from the surgical videos to the corresponding surgical steps. Furthermore, we utilized the domain knowledge to further improve the prediction. Our method not only yields state-of-the-art performance on a common surgical dataset known as cholec80 but also shows great generalization over multiple surgical types. We deployed this model to Amazon Web Services (AWS) using Docker containers, which automate the process from uploading the videos to visualizing the prediction. Our current pipeline took a long time to process the whole video. Thus, as a next step, we will explore techniques to improve processing efficiency while maintaining accuracy.

Anomaly Detection and Data Quality of Supplier Profiles Using Snorkel and Auto-Labeling

Company: TealBook

Student: Marshall Ho

Industry Supervisor: Arnold Liwanag

Academic Supervisor: Mark Chignell

The main goal of our research at TealBook is to improve the quality of our supplier profile data. Our customers rely on our data to help them discover new suppliers, and to meet diversity and ESG (environmental, social, governance) goals.

Most of our data is sourced from unstructured text through crawling the internet. Because of the volume of data involved, human verification or editing is not typically feasible financially. Deep learning is an excellent tool for web-scale information extraction and anomaly detection, but typically requires large amounts of well-labeled training data. The cost of obtaining sizable human-labeled datasets is frequently the biggest barrier to adoption of machine learning methods in industries.

Our research addressed significant issues in this problem domain. We successfully created a training pipeline using Snorkel, an auto-labeling framework developed at Stanford. Snorkel helps to drastically lower the cost of labeling training data, through the use of labeling functions and a learned Label Model. In our experiments, we used Snorkel to create an auto-labeled dataset that is 35x bigger than our hand-labeled training dataset. In pairing a large auto-labeled training dataset with a pre-trained language model such as BERT, our work achieved significant improvements in categorizing unstructured address text, and reduced the error rate of country prediction from 4.6% to 1.6%.

AlphaPortfolio: Large-Scale Automatic Portfolio Construction Using Reinforcement Learning

Company: The Vanguard Group, Inc.

Student: Ruiwen (Tina) Wang

Industry Supervisor: Jithin Pradeep, Jerry Chen

Academic Supervisor: Sebastian Jaimungal

Portfolio management driven by machine learning has regained huge interest in the recent years, followed by a plethora of success in many areas where machine learning excels (e.g., AlphaGo, GPT-3, etc.). Following the latest trend in both academia and industry, we propose a general framework for dynamic portfolio construction based on deep reinforcement learning (DRL). Our approach contrasts the traditional way of constructing a portfolio, consolidating the two-step process that involves both estimation of assets returns and optimization of assets weights into a single, end-to-end pipeline. Our DRL framework handles different portfolio construction needs by leveraging three functionality modules: 1) a module for achieving excess alpha by maximizing the information ratio; 2) a module for mitigating downside risk through maximizing maximum drawdown-adjusted returns; and finally 3) a module for reducing transaction cost via minimizing turnover rate. The experiments are conducted on US stock data from 1971 to 2016, which includes more than 14,000 unique stocks and over 1.8 million records, and the model is able to auto-select stocks from the enormous pool and outperforms standard benchmarks like S&P500 index in various financial metrics. For each module, we also perform perturbation analysis and feature selection using Lasso and Random Forest, which reveal the driving factors that contribute to each portfolio construction goal. Our analysis improves the interpretability of the DRL models and provides novel insights to traditional factor investing.

Controllable Text Generation with Reweighed Diverse Beam Search

Company: The Vanguard Group, Inc.

Student: David Landsman

Industry Supervisor: Hussain Zaidi

Academic Supervisor: Gerald Penn

Recent advances in natural language processing have led to the availability of large pre-trained language models (LM) with rich generative capabilities. Although these models are able to produce fluent text, it is very difficult to control the sentiment, emotion, or other attributes during generation. A majority of current approaches in controllable text generation attempt to select the best input to a LM via prompt tuning or steer the generation towards the desired attribute by fine-tuning a subset of the LM's parameters. These methods suffer from a lack of interpretability, inconsistency in generations, and a lack of generalization to low-resource domains.

We hypothesize that desirable control and comparable fluency can be achieved by controlling the output distribution of a LM, with no changes to the original model. We propose a new extension of beam search, whereby an auxiliary attribute model is used to reweigh hypotheses provided by the search. We evaluate our approach in terms of desired attribute presence, fluency, and content relevance in the resulting generations, and compare to a state-of-the-art fine-tuning approach. Our results support the hypothesis that a similar or better level of control can be achieved, without sacrificing fluency. Our approach allows the flexibility of selecting both the generation model and the auxiliary attribute model, improving the generalizability to low-resource domains. Moreover, our approach offers an interpretable and transparent method for controllable generation, as it operates directly on natural language hypotheses. Future work can explore multi-dimensional attribute control and further enhancements to improve fluency and content-preservation.

SET-C: Sentence and Entity Type Clustering for Unsupervised Open Relation Extraction

Company: The Vanguard Group, Inc.

Student: Stephen Brock

Industry Supervisor: Jerrod Parker, Shi Yu

Academic Supervisor: En-Shiun Annie Lee

With the abundance of unlabelled text available, there is a need for techniques to extract meaningful information from these texts. Unsupervised open relation extraction (UORE) is the process of extracting open-domain relations between named entities from a corpus of text without the use of labelled data or knowledge bases. Current state-of-the-art unsupervised open relation extraction techniques either do not use entity types as input explicitly or fail to provide meaningful contextual embeddings that can supplement entity types.

In this paper, we propose an unsupervised open relation extraction technique, Sentence and Entity Type Clustering (SET-C) that uses k-means clustering on the embeddings produced by our sentence and entity type (SET) encoder. Our method utilizes both entity type and entity context information. Results on a large benchmark dataset show that SET-C outperforms other state-of-the-art approaches in terms of clustering metrics (B^3, V-measure, and ARI) and classification metrics (precision, recall, F1) using the majority relation of each cluster.

For extracting meaningful information from unlabelled text, SET-C is an efficient and novel approach that outperforms state-of-the-art models while also utilizing both entity type and entity context information, and without the need for any fine-tuning. Due to the performance and speed of SET-C, its applications extend beyond unsupervised open relation extraction to other NLP tasks such as information retrieval, knowledge base construction, and natural language understanding.

Real-Time Big Data Visualization Pipeline with Apache Flink and Real-Time Data Warehouse

Company: TimePlay

Student: Linzhi Feng

Industry Supervisor: Francois-Michel Schumaker

Academic Supervisor: Nick Koudas

Data had become the most valuable resource of commercial businesses. As a company with a growing business, Timeplay was struggling on processing large amounts of user data in real-time to ensure the quality of service of its software products. Initially, Timeplay only had an offline data processing tool that required more than 30 minutes to process the data generated by a show. The goal of this research project was to help Timeplay replace this tool with a real-time data visualization pipeline by reviewing and comparing existing methods. This project compared a traditional data-warehouse-based approach with an experimental real-time visualization approach called I2 as well as their associated data middleware, including Apache Flink, ClickHouse, Amazon Timestream, and TimescaleDB. The results showed that I2 performed a little better than the traditional approach in terms of event-time latency, but that the traditional approach could provide users with a better of ecosystem and ease of use. For Timeplay, the traditional approach appeared to be a better solution. Further investigation and additional software infrastructure were needed for applying I2 and other similar approaches in a production environment.

Data-Driven Approach to Defining Symptoms of Mental Disorders and Cognitive Impairment from Noisy Data

Company: Winterlight Labs

Student: Malikeh Ehghaghi

Industry Supervisor: Jekaterina Novikova

Academic Supervisor: Frank Rudzicz, Andrei Badescu

Acoustic and linguistic features are prominent markers of both mental disorders and cognitive impairment, but there is an unfilled gap between these features and the real-world clinical symptoms. The goal of this project is to characterize these diseases by automatically identifying their symptom components as an aggregation of features derived from spoken language and associating those with the clinical symptoms.

To retrieve the data-driven symptom components, we make use of unsupervised learning methods that include both dimensionality reduction and clustering. To explore linear approaches, we initially applied PCA (principal component analysis) over an aggregated dataset given by picture-description task consisting of subjects with AD (Alzheimer's disease), MCI (mild cognitive impairment), depression, and HC (healthy control) diagnosis labels. Additionally, we used non-linear dimensionality reduction techniques including UMAP (uniform manifold approximation and projection) and t-SNE (t-distributed stochastic neighbor embedding), as well as another linear method of LDA (linear discriminant analysis) over the same dataset to compare their performance in terms of characterizing the diseases as well as separability level of distinct diagnosis labels.

After dimensionality reduction with PCA, 8 major symptom components appeared with a 47.8% total explained variance ratio. In addition, among all dimension reduction methods, UMAP indicated the best performance according to its Silhouette score after K-Means clustering as well as the optimal number of clusters which was the same as the number of disease labels(4).

Next steps are to (1) Work on the interpretability of the components by LIME across different dimension reduction techniques to extract highly relevant features aggregated in each component. (2) Augment the dataset with various audio augmentation techniques to upsample minor classes and re-run the dimension reduction techniques.

Fast Biomechanical Simulation with an Iterative Solver

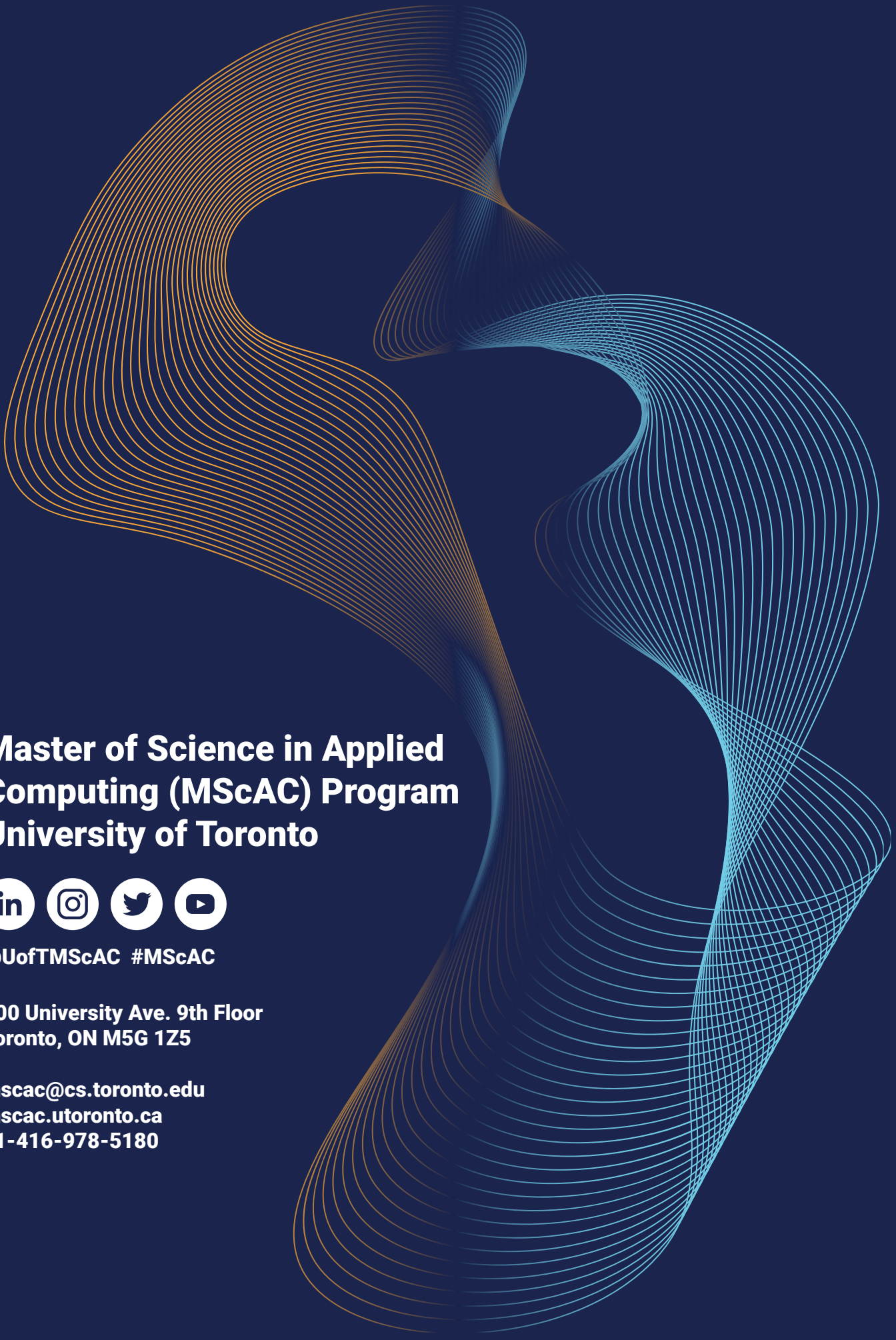
Company: Ziva Dynamics

Student: Yue Li

Industry Supervisor: Crawford Doran

Academic Supervisor: David Levin

Anatomical tissue simulation is one of the most challenging areas within the scope of computer graphics research. It has been receiving increasing attention due to interest from the film industry, video game development, and biomedical research. Ziva's current tissue simulation algorithm is robust and accurate, but not particularly fast. The primary goal of this research project is to improve the speed of the simulator, which will accelerate the iteration cycle for character artists to turn their designs into reality. Experiments reveal that the most significant performance bottleneck comes from the direct solver because inverting the system or matrix decomposition is computationally expensive. The proposed method is to switch from the direct solver into iterative solvers. The new algorithm starts with an initial guess and then iteratively moves towards the optimal until reaching a solution that satisfies the tolerance. The experiment results show the iterative solver is more than fifty percent faster than the direct solver in many internal biomechanical simulation benchmarks; meanwhile achieving the accurate solution thus preserving the original animation effects. Since sparse matrix-vector multiplication is the most needed operation for iterative solvers, the next goal is to speed up this computation by efficient parallelization. Another future work is to overcome ill-conditioned input systems as the number of iterations required for convergence highly depends on the system condition number.



Master of Science in Applied Computing (MScAC) Program University of Toronto

@UofTMScAC #MScAC

700 University Ave. 9th Floor
Toronto, ON M5G 1Z5

mscac@cs.toronto.edu
mscac.utoronto.ca
+1-416-978-5180