

*February 12, 2018*

**Student Projects**

Fast Cloud Image Segmentation with Superpixel Analysis Based Convolutional Networks

Siyuan Liu Master May 2019 Electrical & Computer Engineering

Yunzhen Zou Master May 2020 Electrical & Computer Engineering

Due to the data complexity of various noises, the cloud image segmentation becomes a big challenge for atmosphere prediction. Convolutional neural networks (CNN) are capable of learning discriminative features from complex data, but time consuming in pixel-level segmentation problems. We proposed superpixel analysis based CNN (SP-CNN) for high efficient cloud image segmentation.

First, applied local clustering algorithm to divide all pixels in the image into groups based on pixels’ characteristic and these pixel groups are called super pixels. Next, SP-CNN selects the representative pixels uniformly from the eroded superpixels to avoid the ambiguity from representative pixels on superpixel boundaries. And then, CNN took image patches centered at the representative pixels as input to determine the label of every representative pixel. Finally, all superpixels are classified as cloud or non-cloud part by voting of the representative pixels. This method achieved over 99% accuracy and reduced time cost from hours to 20 seconds.

We designed the cloud image segmentation system and GUI based on Caffe and MATLAB. Local clustering algorithm, CNN prediction, pixel mapping, artificial supervision and display/save segmented images are integrated into a standalone program, which allows users to obtain the result and revise the mistakes of the result.

Design Automation and Microfluidic Resource Optimization of Reconfigurable Flow-Based Biochips

Aditya Sridhar Bachelor May 2020 Electrical & Computer Engineering/Computer Science

As a research undergraduate student of Duke ECE Chair and Professor Krishnendu Chakrabarty, I will present a project that will demonstrate design automation of digital-microfluidic biochips. It will focus on microfluidic applications that improve chip performance. Specifically, I will discuss architectural-level design and the impact of chip parameters. I will also present an approach for determining the control sequences for scalable routing of droplets in complex-valve networks. Finally, I will motivate a solution for pin-count minimization of valve actuation in these same microfluidic systems.

Predictive Model Development for Cisco Network Element Failure Detection

*Prabhakar Nanduri* Master December 2018 Engineering Management (MEMP)

*Manan Vasani*  Master December 2018 Engineering Management (MEMP)

The predictive model developed helps determine whether any disturbances in the network-element traffic would lead to the element failure causing network disruption. Two independent models were developed as a part of the Data Science class. The best features of the two models were merged together to make the current on-display model. Models were primarily built around the open-source training set data provided by Cisco. Tableau was the primary source of data-visualizations in conjunction with visualization libraries in Python. Python was the primary programming language used for the model development.

**A Network Game Designed Using Unreal Engine 4**

Xu Lu Master May 2019 Electrical & Computer Engineering

This is a 2D network game in which players can control their characters to fight against each other. They can melee or pickup weapons to attack (only booms are available now). The first one who kills 20 characters wins.

Players can play this game via LAN or using Steam Server. If the player uses Steam Server, their Stream ID will automatically be used as their nickname in this game, otherwise, their computer names will be used (the nickname can be set in options). To start a game session, a player can establish a lobby, and wait for other players to join. Other players can find all available lobbies on their LAN or on the Internet. In the lobby, players can choose their characters and chat. When all the players are ready, they can start the game.

In the game, the characters can move, jump, chat and fight. Weapon boxes will be generated at places randomly. Players can pick them up and get weapons. Players can pause and quit the game.

Duke Sakai iOS App

Chengzhang (Michael) Ma Master May 2018 Electrical & Computer Engineering

Duke Sakai is a web application for students to check information related to their courses, including assignments, announcements, gradebook and so on. To make students life convenient, we developed an iOS app version of Duke Sakai with a clear and simple UI layout to make it easy for Duke students check Sakai every time and everywhere. It can get student’s information from Duke Sakai APIs after the user log in and display the parsed data in webview and tableview. We also parsed the course sources using DFS algorithm and built an in-app file system to store these data. What’s more, we associated the user’s personal events in Duke Sakai calendar with an in-app calendar view so that users can look them up locally.

Stock Data Analyzer Pipeline

Amber Wang Master May 2018 Electrical & Computer Engineering

We are at the age of scaling and speeding our software to handle much larger quantities of requests from users and handle them efficiently. Using big data technology and tools is the way to go. From this perspective and out of interest, I independently built a data infrastructure pipeline application which was developed to process stock data using big data technology such as Aoache Kafka, spark streaming, Cassandra and Redis. It contains a complete flow of data ingestion, data storage, data computation processing and simple data visualization in a real-time fashion.

Duke Gallery

Si (Ethan) Zhang Master May 2019 Electrical & Computer Engineering

Designed and implemented a fully functioning web application from scratch that allow students view campus life pictures with short descriptions. Logged-in users can post pictures and are able to leave comment or edit their posts. Implemented front-end with HTML5/CSS3/JavaScript/JQuery/Bootstrap for responsive layout great for mobile. Deployed back-end with Node.js as environment for sever side code, Express for handling routes, MongoDB as datastore, and Mongoose for configuring MongoDB models.

Augmented Latent Dirichlet Allocation (LDA) Topic Model with Gaussian Mixture Topics

Kedar Prabhudesai PhD May 2019 Electrical & Computer Engineering

Latent Dirichlet allocation (LDA) is a popular statistical model to discover topics or themes in a large collection of documents. In the LDA model, topics are modelled as discrete distributions over a fixed vocabulary of words. Due to its popularity with text data, the LDA has been used to model other datasets spanning a discrete domain. However, in order to model data spanning a continuous domain with LDA, discrete approximations of the data need to be made, which lead to loss of information and may not represent the true structure of underlying data. We present an augmented version of the LDA topic model, where topics are represented as a mixture of Gaussians. Gaussian mixture models (GMMs) are distributions spanning a continuous domain and are multimodal in nature. We denote the augmented LDA topic model with Gaussian mixture topics as the GMM-LDA model. We provide a collapsed Gibbs sampling scheme to infer model parameters and apply the model to the problem of clustering sleep stage EEG data. We demonstrate superior clustering performance with our GMM-LDA algorithm compared to K-means, standard GMMs, standard LDA, and LDA with unimodal Gaussian topics.

Hardware and Algorithms for an Earthquake Early Warning System for Nepal

Pratiksha Sharma Bachelor May 2018 Electrical & Computer Engineering/Computer

Science

An earthquake early warning algorithm was developed by detecting the p-waves in an earthquake.

A stand-alone hardware was assimilated and the algorithm was implemented in the hardware.

The system is to be tested and implemented in Nepal in early March.

mHealth for Better Routine Immunization Data in Honduras (mVax)

Robert Steilberg Bachelor May 2018 Computer Science

Despite a national routine immunization program in Honduras, children in Roatán remain susceptible to preventable infections due to delays in, refusal to accept, or lack of access to vaccinations. The current vaccination record system in Honduras is entirely paper-based, preventing the rapid identification of, and outreach to, under-immunized or unimmunized children. Our project hopes to streamline the process of recording immunization data, enabling clinics to focus more time on care delivery. We plan to achieve our goal by developing an Android application, mVax, that will be used for recording immunization events for patients of Clinica Esperanza in Roatán, Honduras.

Electromyographic Signal (EMG)-Controlled Prosthetic Arm

Yuan Feng Master May 2019 Electrical & Computer Engineering

EMG-Controlled Prosthetic Arm is designed for people who lost their forearms. It could perform certain actions based on the user’s intended motions. This project was the first-ever case in Asia to combine a Targeted Muscle Reinnervation surgery with an EMG-controlled facility.

In this project, an online training system was designed using C which analyzed the user’s intended motion via Linear Discriminant Analysis (LDA), and controlled the motion of the prosthetic arm accordingly. In this way, the prosthesis first records specific Electromyography (EMG) signal patterns of the user’s various movements, and is then able to recognize the user’s intended motion and make corresponding responses.

Empirical Study of Topic Modeling in Movie Recommendation

Xingu Yan Master May 2018 Statistical Science

Xilin Cecilia Shi Master June 2018 Statistical Science

Xialingzi (Kim) Jin Master May 2018 Statistical Science

Huang Huang Master May 2018 Statistical Science

This project aims to provide movie recommendation based on similarity of movies. Given a movie selected by a user, our goal is to pick a list of movies that share the most similar topics with the target movie. The data is parsed from IMDb movie reviews and Wikipedia descriptions of thousands of movies. Latent Dirichlet Allocation (LDA) is applied to discover latent topics and their distributions within each movie. Various similarity metrics are used to measure distances among movies.

***Peer Konnect: IOS App for Tutoring Matching and Scheduling***

Siyang Chen Master May 2018 Electrical & Computer Engineering

Peer Konnect aims to help students do peer tutoring in a fast and flexible way. In this app, tutors can post a tutoring session and tutee can request it. Tutees can also look for tutors and contact them if no suitable sessions found (e-mail can also be automatically built and sent to the tutors). Besides, users can choose to be a tutor, a tutee or both, and they can also find the available sessions in the map either.

In this project, TableView and CollectionView are customized with friendly features and asynchronization queue to deal with displaying the picture in a fast and smooth way. Logic of interacting with tutor and tutee is along with the session posting to completion. Firebase database is implemented to store user information on the cloud side. Google Map API is integrated to display the nearby events in a clear way. Last but not least, user friendly advertisements are shown by utilizing Firebase Admob.

A Novel 3D Selection Technique – EyeSQUAD

Yunhan Wang Master May 2018 Mechanical Engineering and Materials Science

Traditional ray-casting technique is the most commonly used selection technique in virtual reality. However, it performs baddly in a highly clutterred virtual environment. A novel 3d selection technique - EyeSQUAD (i.e. Eye-controlled sphere-casting refined by QUAD-menu selection technique) - is developed to address this problem. EyeSQUAD allows users to have a hands-free experience during selection with eye tracking. With this technique, user will first select a bunch of objects in the virtual environment which are within a "selection bubble" controlled by eyes. Then a QUAD menu selection scene will be displayed to the user to perform progressive refinement until finally gets a single object.

Movie Recommendation System

Jiankun Chen Master May 2018 Electrical & Computer Engineering

Junbo Li Master May 2019 Computer Science

Shihao Li Master May 2018 Mechanical Engineering

This project is to design a movie recommendation system with MovieLens dataset on AWS EC2. The original data is preprocessed and transformed with Spark. Then, the model is trained with ALS algorithm and evaluated. Finally, the project can show 5 recommendation movies for each user.

Age Identification System Based on Children Speech

Xiaoyu Li Master May 2019 Electrical & Computer Engineering

The objective of this project is to establish a system that can identify the age of a child aged between 5 and 15 and speaking any language using a short segment of his or her speech. First, the speech waveform is converted into a sequence of feature vectors. Most commonly these feature vectors are mel frequency cepstral coefficients (MFCCs). The feature vectors are used to update the parameters of a “Universal Background” Gaussian Mixture Model (GMM), and then we use the GMM models to transform the (arbitrary number of) feature vectors into a fixed dimensional “super vector”.

Classification is done in the super vector space using a support vector machine (SVM) or a Neural Network, or simply by comparing the supervector with average supervectors for each age. The project used the data from PF-STAR corpora of children’s speech: British, German, Italian and Swedish children speaking their native language plus English.

Count Me In

Parker Rice Master May 2018 Electrical & Computer Engineering

Count Me In is and iOS application designed for Duke students and staff that allows users to manage attendance lists for events such as classes, club meetings, etc.  Once the user creates an event with a title, description, location, and time, attendees can scan their IDs using the camera in the app to record their attendance.  After the event, the user can view or email a list of the attendees.  Count Me In uses OAuth 2.0 authentication for user login and backend databases/APIs provided by the Duke Colab for information storage.

Predicting Neonatal Encephalopathy From Maternal Data in Electronic Medical Records

Thomas Li Bachelor May 2018 Computer Science

Neonatal encephalopathy (NE) is a leading cause of neonatal mortality and lifetime neurological disability. The earlier the risk of NE can be assessed, the more effective interventions can be in preventing adverse outcomes. Existing studies that focus on intrapartum risk factors do not provide the early prognostic forecasting necessary to prepare healthcare professionals to intervene early in a high-risk NE case. This work uses maternal data in a supervised machine learning framework to predict NE events. Specifically, we 1) collected the electronic medical records (EMRs) for 104 NE newborns and 31,054 non-NE newborns and their mothers, 2) trained and tested a regularized logistic regression on imbalanced and high-dimensional EMR data, and 3) discerned important features that could be possible risk factors. The learned model offers prenatal predictions of NE cases with an average area under the receiving operator characteristic curve (AUC) of 87% and identified the most important predictors.

Course Exchange During Drop/Add Periods

Tai-Lin Wu Master May 2018 Economics & Computation

Selecting courses is a dynamic process for students. Students tend to change their mind after attending the first few sessions or discussing with classmates. This project studies how to make use of students' revealed preferences to reallocate spots of courses during the Drop/Add period. The main problem the project aims to solve is that students tend not to give up classes they do not need because they cannot get the courses they preferred. We tackle the problem by introducing an exchange market. Exchanges of courses among students could benefit all parties involved. We devise an algorithm that is inspired by the famous Top Trading Cycles algorithm (Shapley and Scarf, 1974) to find the maximum number of courses exchanged in a static setting.

Can We Save the Taxi Industry?

*Ashka* Stephen Bachelor May 2019 Computer Science

With the advent and rise of ride-sharing services and applications, such as Uber and Lyft, it has become increasingly hard for traditional ride-sharing industries (primarily the taxi industry) to keep their first-mover advantages. Previous research done on the subject has shown that ride-sharing applications have significantly impacted taxi industries across multiple cities since 2012 (Wallsten). These ride-sharing applications are able to dynamically match supply and demand according to time and location. This is one of the factors which enable such companies to thrive, displacing traditional ride-sharing transportation institutions. Thus it comes as no surprise that the aggregate number of taxi rides has decreased over time since 2012 (Wallsten). Thus it is an important time to understand the changes occurring and what the taxi industry could do to turn the situation around. In this project, we analyze three core ride-matching algorithms based on data-driven simulations to understand where the taxi industry could improve.

Cloud Cache Server

*Shangxing* Sun Master May 2018 Electrical & Computer Engineering

Cloud storage is a useful and cheap storage, however the reading speed is often limited by internet latency and bandwidth. This project, the cloud cache server, will automatically cache hot data into SSD from cloud service to greatly improve data reading speed. The project is developed in C++ with the use of FUSE (File System in Userspace).

Bundled

*Chenqi* Zhao Master May 2018 Electrical & Computer Engineering

Bundled is a mobile app working on Apple iOS system. The goal is to integrate the food recommandation, grocery shopping list, cooking instructions into one app.The App can recommend meal plan options according to users input constraints, as "Bundles", and users can make selection. Then the system will give shopping list, preparation guide and cooking recipe.