

INTERN FINAL REPORT
NIHERST/NASA INTERNATIONAL
INTERNSHIP PROGRAM 2019

KEANU NICHOLS

November 9, 2019

Contents

1	Introduction	2
2	Overview of the Internship Experience	2
3	Research Conducted	2
4	INTERN'S EXPERIENCE	4
5	FUTURE RESEARCH	5
6	OPENING NEW DOORS	5
7	REFERENCES	6

1 Introduction

I am a third year undergraduate student from the University of the West Indies studying Computer Science. In 2018 I took part in Google Summer of Code working under an organization known as CHAOSS (Community Health Analytics Open Source Software) a Linux Foundation Project. I am currently an undergraduate research student in TTLAB having published a paper titled 'Recommendations for Long-Profit Optimization' alongside Professor Patrick Hosein, Kieran Maharaj and Inzamam Rahaman who is a former NASA intern from the 2015 cycle. I am also the president of UWICS (UWI Computing Society) and one of my main goals is to try and expose other students to the type of opportunities that are available for them as students in the form of internships and also showing them the types of tools that make them a strong candidate.

2 Overview of the Internship Experience

Having an opportunity to be at the heart of Silicon Valley exposed me to an experience that has solidified in me the desire to continually focus on what I love. This internship allowed me to meet persons from around the world who are quite experienced in their field and are dedicated to work that they do. I learned numerous lessons about the importance of conducting research and the steps to conducting sound research. This has made me even more determined to want to go onto wanting to apply for a PhD and continue research in the field of Machine Learning with a focus in Computer Vision. My mentors Brian Coltin and Michael Furlong helped me in so many different ways and were very patient with me with the numerous questions that I had for them. Another big part of the internship experience was being away from home and being exposed to a different culture. I believe a major reason that this transition was easier for me were the other interns both from Trinidad and outside of Trinidad. My roommate Rahat Maini was definitely one of the persons that I believe helped me keep focused on my goal of attempting to obtain a research paper out of this experience. Even though I was not able to finish my research at the end of the internship it taught me some very important lessons I believe as a research scientist you will encounter and that is sometimes your outcome will not be what you expected it to be. I learnt many things from the other interns and one intern specifically Yana Charoenboonvivat reminded me of the importance of working hard. I was very lucky to have met these people and learn so much from them.

3 Research Conducted

For this internship I worked on two different projects under the project DELTA (Deep Earth Learning Training and Analysis). This project's goal is to be able to make it easier for Earth Scientist's to use state of the art Machine Learning models without having extensive knowledge of how these models work. My first

project was determining if buildings were affected by tsunamis. The aim of using this dataset was to be able to determine the type of accuracy an image classifier built on the after images from a tsunami would compare to another image classifier that was built on an image composing of a before and after image. This required me to first read a number of research papers and I came across a dataset [1] that was built after the earthquake that occurred in Japan in 2011. I first had to work on cleaning the data since the images were stored using 6 bands, where the first 3 bands stored the before image and the last 3 bands stored the after image. Hence I used the last 3 bands which was the after image, we then looked at implementing the CNN proposed in [1], this proved to not as straight forward since the exact hyper-parameters for the CNN were not exactly detailed in the paper. This required us to create a number of different models and calculate the accuracy of it. However after training the model a number of times we begun to realize that the model seemed to not actually learn anything and it appeared that it was performing a single classification. Meaning it assumed all the images were affected by the tsunami, or all the images were not affected by the tsunami. The method proposed in [1] also showed that it would typically take around 12000 epochs and as shown below:

Training Time

Number of epochs	Mode of computation	Time per epoch (run through whole dataset) (minutes)
12000	Intel i7 7th Gen (CPU)	90
12000	Intel i7 8th Gen (CPU)	30
12000	Pleiades Super Computer (GPU)	8

Hence by even using the Pleiades Super Computer which is in the top 50 fastest super-computers in the world and using the method proposed in [1] it would take 12000×8 minutes. We then came to the conclusion that something may have been wrong with this dataset but further research needs to be done to verify this.

My next project was related segmenting clouds in Landsat images. My mentors and I were able to discover a dataset proposed in [2] that would have been useful to the DELTA project since Earth Scientist's on numerous occasions have

to manually sift through Satellite images and it often occurs that these images are covered by clouds. Hence my task was to integrate the model proposed in [2] and be able to incorporate it into the DELTA pipeline. As with my previous project it required me to understand how to go about using Tensorflow and their dataset API which is used for large datasets. Due to the size of the Landsat images being around 7000*7000 pixels this meant that I would not be able to load these images easily. Hence the need for using the dataset api; due to the format of the DELTA pipeline I also needed to be able to chunk these images meaning I would take smaller sections of the images and use the center pixel of this section as my label for that section. In using this method one problem was that we tended to lose some data because of the chunk sizes we were using, hence when we reconstructed the image it was not the same image. Hence toward the end of my internship is when we worked on trying to determine how we could best reconstruct the image. But one of the major benefits is that we were able to have a general idea of the labels of the pixels so we could determine if an image is composed of mostly clouds.

4 INTERN'S EXPERIENCE

My time at NASA is an experience that I believe has shaped me into a more focused student and person overall. My time at the NASA Exchange Lodge is one of my favorite experiences of summer because I was able to meet so many fantastic individuals. I was able to learn about so many different cultures and got insights into many people's different perspectives on so many different views. One of my favorite yet challenging experiences was having to prepare meals for myself on a day to day basis. Hence this meant that I was in the kitchen constantly, some of the other interns even said that it seemed that I lived in the kitchen. This is due to the fact that my room was two door's down from the kitchen, I was able to meet a lot of people in the kitchen and learned about so many different types of meals. Also since I was in Mountain View California, the heart of Silicon Valley it felt like I was at one of the major cities that is at developing some of the major cutting edge technologies. Every time I spoke to interns and I would be so fascinated by the projects that they were working on. One of the experiences I believe was a favorite of mine was having to do our poster presentation, this required me to discuss what I worked on with people I had never met before. It was even more fruitful when I had to speak to Earth Scientists who would actually use the models that we created and they gave me more insight as to what things that they would find useful. Another concept I believe that was invaluable to this experience was learning about the importance of constantly reading research papers. This is because it helps you better understand what research has been conducted and the type of outcomes that have come out of it. It is also helpful when to research papers go into great detail as to how they went about implementing their solution or for example created their dataset. It makes it easier for persons to then be able to replicate the research that was conducted and be able to improve on it. Being in the

Intelligent Robotics Group meant that our lab was always bustling with robots moving around, it also meant that we had a few guest speaks at our weekly meetings to talk about cutting edge technologies that was being developed. I thought this was always a favorite of mine and also when IRG would have our monthly BBQ. I believe it allowed us to bond more as a group and I heard about other projects that people were working on and it always very fascinating.

5 FUTURE RESEARCH

Due to the research area that I was working on, that being image segmentation and image classification, this research has a profound relevance to a place like Trinidad and Tobago. After returning to Trinidad I was able to speak with a Professor Patrick Hosein about continuing the work that we did and we came to the idea of looking at the rate of deforestation in Trinidad. This is a major problem that would need to be constantly monitored and could better help policy makers make more informed decisions when it comes to the urbanization of areas in Trinidad. Since we are in the age of more prevalent satellite images it becomes easier to capture images of places like Trinidad and be able to look at how it changes overtime. Hence when we use techniques like semantic segmentation we are able to better understand how specific areas have changed overtime. This problem however is not something that will have a single solution, there are a number of different models that can be created to classify the individual pixels in images, hence when a solution is proposed more research can be conducted to try and improve on the models that are created. We can also look at creating datasets of the images of Trinidad but compare our models to other datasets. Another area of research could be looking at images of Trinidad during Natural Disaster, this can be used to determine which area are affected by disasters and potentially make it easier for emergency services to reach these places at a faster rate. It can also be used by for example insurance firms to determine if persons who claim that their homes were actually affected by Natural Disasters to avoid things like insurance fraud.

6 OPENING NEW DOORS

I am currently in my third year of university but I am now more confident that I want to obtain a PhD. However in my area of interest, that being Machine Learning applied to Computer Vision, both Calculus and Linear Algebra is fundamental to this area hence I am taking an extra year to focus on these areas. I will also continue to do undergraduate research and make me a stronger research candidate. I hope to continue working in the field of Machine Learning and being able to conduct research in Trinidad, one of the main problems with Machine Learning is that complex models tend to take up a lot of time and requires expensive equipment. However overtime computing power becomes cheaper hence the only constraint is having the people knowledgeable about

