Emily Park

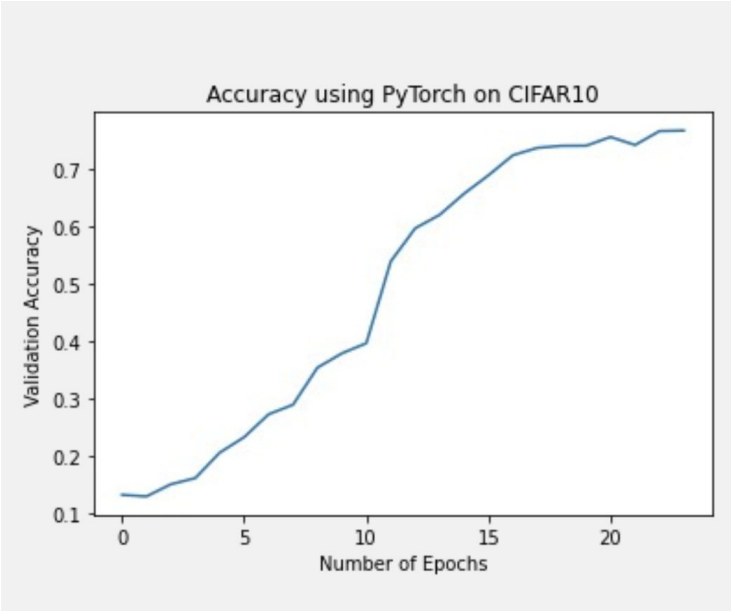
August 12, 2020

CRCV HSAP 2020 Report

As one of the two 2020 High School Apprentice Program interns, I was able to work at University of Central Florida Center of Research for Computer Vision (UCF CRCV). I came into this program with prior knowledge in Python but no experience in computer vision research. With not much exposure to computer vision and deep learning, I applied to hoping to understand more about what computer vision research entailed as well as its potential applications. With my mentors, Dr. Shah and Robert Browning, I was able to learn much more than what I had imagined coming into this program in the field of computer vision.

During the first two weeks, I worked with the previous REU lectures to learn about the materials prior to the actual research. In the first week, I ran Assignment 0, where I ran the given code using the *MNIST* dataset. I also worked with Assignment 1 and 2, where I used *Keras* and *Pytorch*, respectively, to construct a *Convolutional Neural Networks (CNN)* with the *CIFAR10 dataset*. I learned about the historical introductions of Computer Vision as well as its applications, *CNN* image processing, basic python algorithms used in training networks, and started learning about the deep learning Python framework *PyTorch* and *Deep Learning*.—

In the second week, I ran the same *CIFAR*10 Dataset with Pytorch instead of Keras using the *SGD* optimizer. I installed *PyCharm* with *Anaconda* and ran my code. With Keras, I obtained a final validation accuracy of 83.4% and a loss of 0.5 (19 sec/epoch). With PyTorch I obtained a validation accuracy of 76.6% and a loss of 0.416 (~28 sec/epoch; *SGD* optimizer).



In addition, I attended three presentations/lectures. One was with Dr. Aidean Shargi's presentation in *Video Textual Video Synopsys*, where I learned video summarization techniques and the hierarchical models based on query relevant events. Another was Aisha Urooj's presentation where she explained the visual understanding of image and video captioning and the steps of image information filtering and attaching an 'emotional significance'. The last presentation was Krishna Regmi's presentation, explaining

the steps for my new data collection project for ground and aerial videos. I learned to gather ground and aerial view videos from various cities and process them after a significant amount of collection. During the data collection, we recorded the YouTube ID, name of the publisher, and the date the video was published in an Excel sheet.

During week 3, I collected approximately 130 ground and aerial videos for the data collection project. During this week, I attended a live, private meeting with Dr. Aidean Shargi, where I was able to get a better understanding from the previous YouTube recorded lecture I had watched. I learned about the applications of synopsys and summarization and user-query information, as well as various overviews of models and its applications and shortcomings. I also started reading the research paper titled "Video Description: A Survey of Methods, Datasets, and Evaluation Metrics", where I learned about video description and its applications in real world, image and (dense) video captioning, as well as transformation methods of attention and sequence learning mechanisms.

During week 4 and week 5, I collected a total of 110 new videos. In the beginning of week 5, I attended a Graduate School Workshop where I learned about various fellowships and opportunities as well as grants. I was also able to get insight on graduate school programs and its benefits from various speakers. Lastly, I continued on reading the research paper on Video Description, where I learned machine translation and image captioning and its datasets with *BLEU*, *ROUGE*, *METEOR*, and *CIDEr*. I also learned about the statistical methods which is used for large datasets (i.e. Rohrbach) and how deep learning works, involving *CNN*, *LSTM*, and *RNN*, which currently is dominating sequence modeling in visual content extraction and text generation.

During week 6, I continued on data collection on ground/aerial videos and research paper reading, but also attended the REU 90 seconds spotlight presentation. For the data collection, I collected around 60 new videos for both ground and aerial videos, making it about 120 new videos and 320 total videos. The 90 second spotlight presentations showcased short overviews of REU students' computer vision projects. As I continued on with the research paper, I learned about *CNN* and the encoding process, which is structured into fixed-size and variable-size. I also learned Variable Visual Representation Models, which directly maps input videos that have different number of frames to variable length. Moreover, the paper addressed *Deep Reinforcement Learning Models*, popularized by *Google DeepMind,* which is considerably harder to devise compared to traditional supervised techniques, since it doesn't have full access to the function being optimized and its interaction with the environment is entirely state-based. Lastly, I was assigned an additional project called the "UTRAP Project". I installed *BULKR* and *Adobe AIR* and received the keywords spreadsheet and the data collected. My task was to

download relevant photos from the keywords with *BULKR*, which became about 2000 photos per term (only about 500 photos collected before I started).

During my last week, I finished collecting ground and aerial videos for Krishna Regmi, reading the research paper and collecting data for the UTRAP project. For the data collection on Krishna Regmi's project, I recorded around 20 additional videos for both ground and aerial videos, making a total of 40 new videos and an overall total of around 360 videos (including both ground and aerial videos). I also finished the research paper on "Video Description", where I learned four main classes of datasets: cooking, movies, videos in the wild, and social media. I learned about *TACoS*, which is constructed by filtering through *MP-2* composites, providing alignment of sentences describing activities, obtaining approximate time stamps for each start and end times in each activity. For movies, I learned about how it contains transcribed audio descriptions from movies, and how *M-VAD* data split consists of video clips for training, validation, and testing. Furthermore, I learned about manual filtering and how it's carried out to ensure each video meets the prescribed criteria on different datasets such as *MSR-VTT*, *Charades*, and *ANet (Activity Net)*. Social Media also contains a multi-sentence description dataset, aimed to address the short narration of long videos that can't fit within a single sentence. Here, *ANet-Entities* are used to build on the training and validation splits, but with different captions. I also got a brief overview on Evaluation Metrics, such as Automatic Evaluations. There are four metrics: *BLEU* (popular metric used to quantify the quality of machine-generated outputs), *ROUGE* (metric that evaluates text summaries and calculates recall score of generated sentences), *METEOR* (metric proposed to address the shortcomings of *BLEU*, introduced in semantic matching, and is based on the score computation of how well the generated sentences are aligned), and *CIDEr* (recently introduced evaluation metric for image captioning task and converts texts into root forms). Lastly, I worked on the UTRAP data collection for Rohit and Xiaoyu Zhang. The terms were *Military Uniform, Assault Rifle, Aircraft Carrier,* and *Cannon*. I created about 20 key terms for each term, and collected 1000 new data for each term.

Coming in with little to no experience aside from some knowledge in Python, the lectures were difficult and challenging. However, in retrospect, I have learned so much and gained many amazing connections and experiences. Looking back, this was one of the most successful internships that changed how I look at my everyday surroundings, constantly looking for applications of deep learning and computer vision. Although the obligated summer weeks have passed, after a short break, I am planning to come and help with more projects and data collection with Dr. Shah and Robert Browning in the coming fall.