Annex D2

Index No: 200398D

Summary of the work experience

(Duration: From 27th November 2023 to 5th of January 2024)

Week 01:

Upon arrival in Singapore, my first week was dedicated to sorting out paperwork, fulfilling essential requirements to secure entry to the laboratory, and officially enrolling as a research intern at SUTD (Singapore University of Technology and Design).

Week 02:

After that, I received my project guidelines, ongoing tasks and existing works, which focuses on **Energy-efficient, Scalable, and Reliable Distributed Green Streaming Machine Learning for Edges.** In my second week of the internship, I virtually meet my project professor in charge due to his travel on abroad. I also met other contributors who already in this project. From those meetups I got the big picture of the project as well as my part. This project is about comes under Real time Data-Streaming and Real time Machine learning model training on the edge device in an energy efficient manner. Since it deals with the real time data, model training also needs to adapt the online setting to keep the model update according to the time. So, one direction address towards the Online continual Learning on the ML model. IN this period, I read related papers of the team about the previous works.

Week 03:

In my third week as a research intern, I was allocated a task which related to Data selection for the online model training. When we need to train the model in the online setting, we need to give the data as stream. For that I learn and play around with Apache-Kafka beaming which stream the data. So now we can train the model based on the stream's arrival. Models need to be trained on the arriving streams. I researched about the cons and pros in the online training and what are limitations we have in the present state.

Week 04:

In my fourth week, I explored existing model training approaches. For that I read the research papers which involved in similar tasks in the past. From that I learned the significance of the data selection process on the data streams. Because latency and model complexity are key factors affecting the training. High latency leads to the missing of the streams which downgrade the model. Most of the data streams have redundancy and unwanted element so we need to filter them in a efficient as well as quick manner. Again, I research about the existing Data selection approaches used for the online learning.

Week 05:

During my fifth week, I dedicated my time to develop a own implementation of the data selection process for the model training and estimated the time consumed for the training based on the data selection process. For the data selection, I initially adopt the core set selection approach where Select the representatives of the data samples arrived on the stream. For the selection process I extract the distinct data from the stream and then classify them based on the classes. Then select the representatives from each of the classes to avoid the unbalance across the classes. But I got high latency and get quite low accuracy in the model performances. So, I discussed about this with my supervisor. He suggested to use the libraries to perform some tasks rather than develop the own one.

Week 06:

In the sixth week, I further researched on how to improve the data selection process in terms of accuracy and latency. Through the research I find out the major problem in the updating of the model which is catastrophic forgetting which is the reason for the low performance of the model in the online setting. So, try to modify the approaches. I get to know about the coreset selection process and selection via proxy model which are used for the data selection and usage of buffer management which reduce the effect of catastrophic forgetting. Meanwhile I have a discussion with my professor. He gave a Paper called CAMEL which implement both in their paper for stream learning. I went through that paper. I also search for the improvements in their approach.

My initial six-week internship proved to be a rejuvenating and insightful experience, focusing primarily on a research project. Throughout this period, my emphasis was on assimilating knowledge and applying it practically. The fundamental comprehension I gained in my first five semesters at university proved to be extremely valuable, particularly in the design phase of the control board. The educational perspectives and skills acquired from previous academic projects played a crucial role in successfully overcoming various challenges. Starting to explore the new and popular area of StreamML and Edge ML, I didn't know much about the theories at that area. But My supervisor and PHD research assistant help me to understand the vast topic and gave valuable insights on my works and progress.

In summary, my internship has been a mix of acquiring knowledge, working together, and personal development. I eagerly anticipate further learning and exploration in the upcoming weeks.