

EE/CprE/SE 492

HAML: Heterogeneous and Accelerated Computing for Machine Learning

Semester 2 Week 3-4 Report

9/6/24 - 9/19/24

Faculty Advisor Phillip Jones
Client JR Spidell

Team Members:

Jonathan Tan	- Memory Affinity, Kria Board Manager
Josh Czarniak	- DPU Control Developer
Justin Wenzel	- Multi-threaded Developer
Kai Heng Gan	- Image Processing/Semantic Segmentation Developer
Santiago Campoverde	- Model Analytics

Summary for Progress These Two Weeks

The past two weeks, our focus was to start working on our individual tasks to accomplish “milestone 3”, which is to have a multi-threaded program capable of running all three algorithms on the Kria KV260 board in parallel, successfully profile the performance of the system, and calculate system accuracy.

These Two Weeks’ Individual Contributions

- Justin
 - Improve synchronization issues occurring in multi-threaded program
 - Testing using semaphores to synchronize threads
 - Begin implementing how model outputs will be handled in the multi-threaded program based on the models’ specific output.
 - This includes handling all three models’ different outputs and using the outputs to influence the multi-threaded programs output and feedback to the user.
- Jonathan
 - Debug error when using vaitrace, currently, profiling is not working due to a divide by zero error, which I believe is caused by the profiling program (vaitrace) failing to detect events.
 - Things tried:
 - Update XRT version (XRT’s version was older than Vitis AI’s version), didn’t work.
 - Tried different kernel tracer settings, in-progress.
- Josh
 - Led week four’s meeting
 - Looked into DPU API libraries
 - Was provided two articles by the client to help research
 - Researched multithreading SSD for DPU cores

- Kai
 - Validate the input data of the image preprocessing xmodel
 - Create a custom “transforms” function in cpp that aligns with the torchvision “transforms” function’s definition.
 - Examine the output of the segmentation xmodel and got different value from the original Pytorch model.
 - I tired modify the cpp code to be the exact functionality as Python script that would load the semantic segmentation model and output a segmented image.
- Santiago
 - I got assigned a new role for the project, which involved testing the various model's accuracy.
 - Began researching model accuracy tools and the specifications for testing each model.
 - Met up with Kai and Justin to understand Semantic Segmentation output and ways to validate the output.
 - Got output files and ground truth file from the original frames to analyze data.
 - Set up a plan to start testing Semantic Segmentation data using Vitis AI library tools.
 - Test model on Vitis ai docker container and determine the feasibility of using these tools instead of creating new ones.

Team Member	These Two Weeks’ Task	Completion Date	Hours Took	These Two Weeks’ Hours	Total Project Hours
Justin Wenzel	Attended meetings	NA	1	5	108.5
	Improve synchronization using semaphores	9/12	2		
	Begin implementing model output handling	9/16	2		
Jonathan Tan	Attended meetings	NA	3	8	127.5
	Debug error when running profiler on the board (vairtrace)	On-going	3		
	Update XRT version on the board	9/12	2		
Josh Czarniak	Attended meetings	NA	1	5	95
	Looked into DPU API libraries	9/12	3		
	Researched multithreading SSD for DPU cores	9/17	2		
	Attended meetings	NA	3	9	121.5

Kai Heng Gan	Validated the input data of the image preprocessing xmodel	9/14	3		
	Examined the output of the segmentation xmodel	On-going	3		
Santiago Campoverde	Attended meetings	NA	3	8	88
	Began researching model accuracy tools and the specifications for testing each model	9/13	3		
	Met up with Kai and Justin to understand Semantic Segmentation output and ways to validate the output.	9/16	1		
	Set up a plan to start testing Semantic Segmentation data using Vitis AI library tools.	9/19	1		

Note: 1. This is per week hours, Σ "hours taken" = "week hours". 2. Due to multiple meeting times, meetings' "completion date" are "NA".

Plans for Coming Two Weeks

Team Member	Plans for Coming Week	Planned Completion	Planned Hours Required
Justin Wenzel	Transfer multi-thread managing program from docker container environment to the Kria board (pushed back from previous week while fixing synchronization issues and output handling)	9/23	5
Jonathan Tan	Continue debug error with vaitrace	9/25	5
	Start looking into implementing memory affinity on the Kria board.	10/10	12
Josh Czarniak	Continue working on DPU management and research multithreading SSD for DPU cores	ongoing	5
Kai Heng Gan	Continue working and testing on preprocessing cpp code that will run on the Kria KV260. Resolve the invalid output from the xmodel.	ongoing	10
	Re-quantize and re-compile the semantic segmentation xmodel.	ongoing	2

Santiago Campoverde	Test model on Vitis ai docker container and determine the feasibility of using library tools instead of creating new ones.	9/22	3
------------------------	--	------	---