Memory   
 Out of memory   
 Illegal Memory Access/ allocation   
 Memory Copies  
 Memory Usage

Environment   
 Driver issues   
 Version Issues   
 Devices Issue/ Environmental Variable

Synchronization (Streams, threads)

Compiler   
 Make Files   
 NVCC compiler   
 Compile   
  
multi-GPU   
Math Errors  
 Math Functions  
 Floating point   
API function/ Flagging Error

Library   
 NCCL  
 Thrust   
   
Visual/ Graphics issues

Other

Fixes   
Reduce batch sizes, block size, increase memory, limit GPU memory,

Change GPU type, correct package installation, update drivers, update versions,

Use pinned memory, improve multi-stream/ GPU

Matching NVCC versions, correct NVCC path, build make files correctly,

Fix code, correct dimensions, correct NCCL initialization, boundary checking,

METHODOLOGY

1. Sample Collection

In order to collect samples of GPU programming related issues, we retrieve commit samples from GitHub repositories. First, we obtain a list of GitHub issues mentioning keywords ‘GPU’ and ‘GPU programming’ using the GitHub search API dated from January 2020 to December 2020 to identify 118,438 issues posted. From this set of issues, we download all of them for further evaluation. Next, we search the issue types for patterns related to ‘GPU\*’, ‘memory’, ‘threads’, ‘streams’, ‘environment’, ‘synchronization’, ‘concurrent’, ‘multi-GPU’ and ‘compile\*’ to find issues related to GPU programming. This step reduces the number of commits to 1,581. In order to confirm which of these issue links were related to GPU programming related, manual inspection was performed on the issues in the list. After manual inspection, and removing duplicate issues, the number of verified GPU programming bugs is 241.

1. Portion of GPU Programming bugs to other bugs
2. Sample Inspection

After collecting these issues of GPU Programming bugs from GitHub, we manually inspect the collected samples to identify information relevant to our research questions. In particular, we analyze the collected issue reports by inspecting the issues for following traits: the root cause of the bug, how the bug is manifested, and how the bug was fixed. For the issue reports, we inspect the developer comments and the linked commits. Through the inspection we obtained the sample set of 241 GPU programming bug issues.

1. Dataset Composition

Our dataset consists of a diverse of GPU programming bugs. The languages of the GPU programming bugs analyzed are CUDA, OpenCL, and GPU-accelerated computing with Julia, python.

|  |  |  |
| --- | --- | --- |
| Root Cause  Categories | Root Cause  Subcategories | Total |
| Memory | Out of memory  Illegal Memory Access/ allocation  Memory Copies Memory Usage | 35 19 9 5 |
| Environment | Driver issues  Version Issues  Devices Issue/ Environmental Variable | 5  27 14 |
| Synchronization | - | 36 |
| Compiler | Make Files  NVCC compiler  Compile | 5 11 7 |
| Math Errors | Math Functions Floating point | 8 5 |
| API function/ Flagging Error |  | 9 |
| Library | NCCL Thrust | 11 4 |
| Visual/ Graphics issues | - | 6 |
| Multi-GPU implementation | - | 17 |
| Other | - | 8 |

Memory   
 Out of memory

"https://github.com/oreilly-japan/deep-learning-from-scratch-2/issues/16"

"https://github.com/PlasmaControl/DESC/issues/1"

Illegal Memory Access/ allocation

"https://github.com/JuliaGPU/CUDA.jl/issues/558"

"https://github.com/ptillet/torch-blocksparse/issues/15"

"https://github.com/microsoft/onnxruntime/issues/5555"

Open Issue but solve the problem:  
"https://github.com/cupy/cupy/issues/3452"

Memory Copies

"https://github.com/rapidsai/dask-cuda/issues/438"

"https://github.com/JuliaGPU/CUDA.jl/issues/105"

Memory Usage

Environment   
 Driver issues   
 Version Issues   
 Devices Issue/ Environmental Variable

Synchronization

Compiler   
 Make Files   
 NVCC compiler   
 Compile

multi-GPU   
Math Errors  
 Math Functions  
 Floating point   
API function/ Flagging Error

Library   
 NCCL  
 Thrust   
   
Visual/ Graphics issues

Other