Example Projects

1. Use Machine Learning to enhance sketches or bloom filters. (<https://arxiv.org/abs/1802.00884> or <https://openreview.net/forum?id=r1lohoCqY7> )
   1. Either develop a new algorithm/idea and analyze it Or
   2. Implement at least four reasonable variants/algorithms and rigorously examine the accuracy-resource tradeoffs on two large datasets.
2. Use Machine Learning to create adversarial traffic by observing sketch latency: (Open-ended and untouched area with direct applications in security.) For more information contact the instructor.
3. Implement some of your favorite randomized algorithms from the course over any one of the following 1) single GPUs, 2) multiple GPUs and 3) Multi-node. Do a rigorous analysis of its performance with the non-parallel versions. Please discuss with the instructor if your choice of algorithm qualifies.

More involved Project (Can approve larger groups based on need)

1. Extend Slide Codebase C++ codes

( <https://github.com/RUSH-LAB/HashingDeepLearning> )

For any one of the following

* Convolution layers
* Attention
* Add Recurrence Function
* Any interesting feature you may imagine

1. Extend FLASH (<https://github.com/RUSH-LAB/D-FLASH> ) or D-FLASH (<https://github.com/RUSH-LAB/D-FLASH>)

For any of the following

* Text data
* Image Search
* Video Search
* Any other interesting thing

1. Find and implement novel applications of Capsule (<https://www.cs.rice.edu/~as143/Papers/CAPSULE.pdf>) or efficient image search on mobiles.

(Ask for code, it is similar to FLASH)

1. Code an efficient and parallel Search Engine using MACH (<https://github.com/RUSH-LAB/MACH> )

**Some titles from last few years projects.**

| Benefit distributed machine learning with LSH-enhanced batch normalization | Accelerate asynchronized distributed machine learning with straggle mitigation |
| --- | --- |
| Text similarity search with LSH |  |
| Deep Learning for the Enhancement of Bloom Filters |  |
| Using ML to enhance Bloom Filters |  |
| Plagiarism detection with LSH |  |
| Either see if LSH can improve over random sampling in the use of memory banks for unsupervised representation learning, or given a set of models and a sample we wish to query a subset of models for the best prediction (e.g. if we have a set of models which are good at different tasks) kind of like reverse LGD |  |
| Given efficient application CaPSuLe algorithm in image searching process, wewant to extend the idea of hashing into gif searching. |  |
| A hash-based approximate Top-K query processing on massively parallel hardware |  |
| See if LSH can improve IPFS search engine |  |
| Something about probabilistic encryption schemes and cryptographic algorithms |  |