Yao Jianyu

Residence
Beijing, China

Email
yaojianyu89@gmail.com

My name is Yaojianyu.

Education

Masters degree. Institute of Computing Technology, Chinese Academy of Sciences [2019 - now]

Bachelors degree in Software Engineering. Shandong University [2015 - 2019]

Experience

IAAT: A Input-Aware Adaptive Tuning framework for Small GEMM

November 2020 - May 2021. The paper is under review in ICPADS'21

GEMM with the small size of input matrices is becoming widely used in many fields like HPC and machine learning. Although many famous BLAS libraries already supported small GEMM, they cannot achieve near-optimal performance. This is because the costs of pack operations are high and frequent boundary processing cannot be neglected. We proposes an input-aware adaptive tuning framework(IAAT) for small GEMM to overcome the performance bottlenecks in state-of-the-art implementations. IAAT consists of two stages, the install-time stage and the run-time stage. In the run-time stage, IAAT tiles matrices into blocks to alleviate boundary processing. This stage utilizes an input-aware adaptive tile algorithm and plays the role of runtime tuning. In the install-time stage, IAAT auto-generates hundreds of kernels of different sizes to remove pack operations. Finally, IAAT finishes the computation of small GEMM by invoking different kernels, which corresponds to the size of blocks. The experimental results show that IAAT gains better performance than other BLAS libraries on ARMv8 platform.

IAAT for different platforms

June 2021 - now

Our previous work only focused on ARMv8 platform, and was not suitable for other platforms. Therefore, we reaserch the common method and implementation of small GEMM for different platforms. Besides, we plan to utilize JIT(just-in-time) to optimize small GEMM

OpenVML: Vector Math Library

August 2020 - April 2021

As a basic mathematical operation, the high-performance implementation of trigonometric functions is of great significance to the construction of the basic software ecology of the processor. Especially, the current processors have adopted the SIMD architecture, and the implementation of high-performance trigonometric functions based on SIMD has important research significance and application value. In this regard, the project uses numerical analysis method to implement and optimize the five commonly used trigonometric functions sin, cos, tan, atan, atan2 with high performance. Based on the analysis of floating-point IEEE754 standard, an efficient trigonometric function algorithm is designed. Then, the algorithm accuracy is further improved by the application of Taylor formula, Pade approximation and Remez algorithm in polynomial approximation algorithm. Finally, the performance of the algorithm is further improved by using instruction pipeline and SIMD optimization. The experimental results show that, on the premise of satisfying the accuracy, the trigonometric function implemented is compared with libm algorithm library and ARM*M algorithm library, on the ARM V8 computing platform, has achieved great performance improvement, which is 1.77-6.26 times higher than libm algorithm library, compared with ARM*M The library of M method is 1.34-1.5 times higher

C	++++	C++	++++	Assembly	+++	SIMD	++++	CUDA	+++	MPI	+++
pthread	+++	Makefile	+++	CMake	+++	Latex	++++	Linux	++++	Git	++++
Python	+++	Java	+++	JavaFX	++++	Android	++++	SQL	++		