Efficient implementation and acceleration of DIP-NMF-MM algorithm for high-precision 4D PET image reconstruction

Yoshinao Yuasa† Kaito Matsumura‡ Tatsuya Yokota‡ Satoshi Ohshima†3 Hidekata Hontani‡
Toru Nagai†4 Yuichi Kimura†5 Muneyuki Sakata†6 Takahiro Katagiri†4

Graduate School of Informatics, Nagoya University† Graduate School of Engineering, Nagoya Institute of Technology‡ Research Institute for Information Technology, Kyushu University†3 Information Technology Center, Nagoya University†4 Faculty of Informatics, Cyber Informatics Research Institute, Kindai University†5 Tokyo Metropolitan Institute of Geriatrics and Gerontology†6

1 Introduction

Positron emission tomography (PET) is a technique to observe biomolecular mechanisms such as metabolic processes by measuring the distribution of radioactive tracers in a human body. PET images are reconstructed from observed data called sinograms, and there are high expectations for the use of GPUs to handle large data and computationally intensive algorithms. Yokota et al. [1] proposed a new method using non-negative matrix factorization (NMF) and Deep Image Prior (DIP). The method achieved more accurate reconstruction than conventional methods. Furthermore, Matsumura et al [2] used the MM algorithm instead of gradient descent in the reconstruction process to achieve more stable PET image reconstruction than conventional methods. In this study, we evaluate the performance of a 4D PET image reconstruction code that implements the DIP-NMF-MM algorithm in a 2X higher-resolution execution and four acceleration methods.

2 4D PET image reconstruction

4D PET images are obtained by repeating the following four processes using the observed data, called "sinograms," as input :

Step1. Adding noise to inpute data
Step2. Update U-Net prameters
Step3. Normalize each spatial basis
Step4. Update time basis using MM algorithm

3 Implementation

3.1 UVM

When the input data size is doubled, the GPU memory is insufficient for normal execution. In this study, we use NVIDIA's Unified Virtual Memory (UVM) as a countermeasure to the problem of insufficient GPU memory. This feature is a memory management system that simplifies GPU development by providing a single memory space accessible by all CPUs and GPUs in the system. This feature is available on NVIDIA GPUs for datacenter and HPC applications from the Kepler generation onward.

3.2 4 Accelerated implementations

For the optimization of execution time, we implemented and evaluated the performance of the following four acceleration methods.

1. Optimization of noise additive processing
2. Computational graphing of the entire process
3. Embedding input data into computed graphs
4. Distributed execution of U-Net

4 Experiment

Table 1 shows the execution time per step for the 4D PET image reconstruction code when the four acceleration methods shown in Section 3.2 are implemented in sequence in addition to UVM shown in Section 3.1.

Table 1: Execution time for each optimization

	original	1.	2.	3.	4.
Run time [sec/step]	29.26	20.21	18.36	18.35	2.86
Rate of acceleration		1.45	1.59	1.60	10.23

5 Conclusion

In this study, we succeeded in outputting images with twice higher resolution by using UVM with respect to the 4D PET image reconstruction code using the DIP-NMF-MM algorithm. In addition, a maximum speed-up of 10x was achieved by applying four speed-up methods.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP19H05662.

REFERENCES

[1] Tatsuya Yokota et al., "Dynamic PET Image Reconstruction Using Nonnegative Matrix Factorization Incorporated with Deep Image Prior," Proceeding of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[2] Kaito Matsumura et al., "Dynamic PET Image Reconstruction Stable Optimization of Nonnegative Matrix Factorization," Proceeding of Signal Proceeding Symposium (SIP), December 2022.