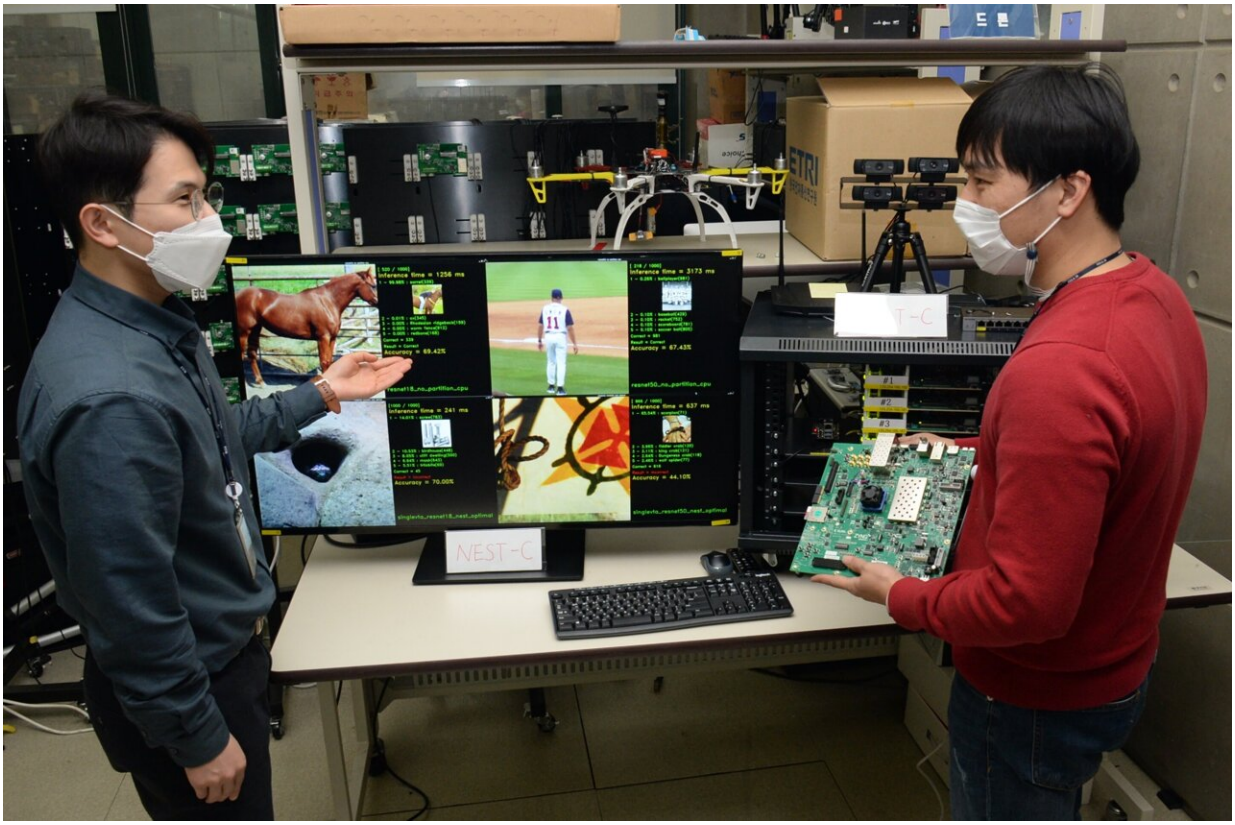
AI core software 'Deep Learning Compiler' developed

March 8 2022



Researchers from ETRI are giving on a demonstration that image classification deep learning technology operates on NPU through Nest Compiler. Credit: Electronics and Telecommunications Research Institute (ETRI)

A Korean research team has unveiled a core technology which reduces the time and cost invested in developing artificial intelligence (AI) chips

in small and medium-sized enterprises and startups. Thanks to the development of system software that solved the problems of compatibility and scalability between hardware and software which have been obstacles in the past, the development of AI chips is also expected to accelerate.

The Electronics and Telecommunications Research Institute (ETRI) has developed an AI core software, Deep Learning Compiler 'NEST-C'. It was released on the web along with the internally developed AI chip hardware so that developers can use it easily.

As AI technology develops, deep learning application services are expanding into various fields. As the AI algorithm to implement this service becomes more complex, the necessity for better and more efficient arithmetic processing has increased.

ETRI resolved this problem by defining a common intermediate representation suitable for AI applications to apply it to the nest compiler. By resolving the heterogeneity between AI applications and AI chips, AI chip development becomes easier. This technology was also established as a standard of the Telecommunications Technology Association (TTA).

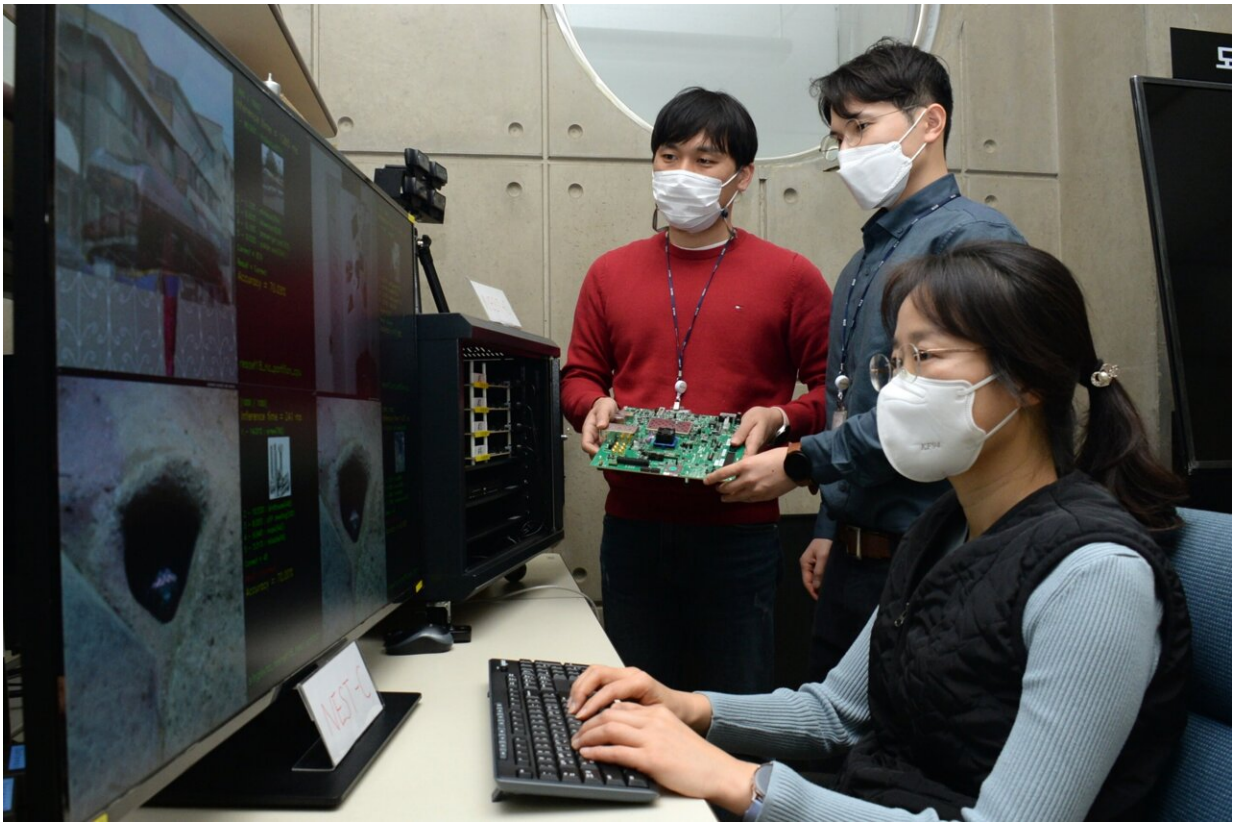
Instead of CPU or GPU, Neural Processing Unit (NPU) AI chip specializing in AI computation processing is drawing more attention. In order to run applications such as autonomous driving, Internet of Things (IoT), and sensors, optimized AI chips to each application must be designed.

At the same time, an optimized compiler must produce accurate execution code for each application to achieve optimal performance. It is because Deep Learning Compiler, core system software that guarantees the inference performance of deep learning models, is

important. It acts as a bridge between hardware and application software.

In general, manufacturers develop this tool along with AI chip and system software for sale. So far, SMEs and Startups have difficulty in focusing their capabilities on chip design. It is because a considerable amount of time needs to be devoted to developing and optimizing system software and application. System software supplied by a large manufacturer is optimized to its own chip, and there is a limitation to its application as it is developed privately.

Especially, there is an inconvenience of developing different compilers depending on the chip type and AI application. There has been a limitation on inter-compatibility and extensibility into new areas.



Researchers from ETRI are giving on a demonstration that image classification deep learning technology operates on NPU through Nest Compiler. Credit: Electronics and Telecommunications Research Institute (ETRI)

This development will make it possible to shorten the time for developing applications and their optimization. It is also related to reducing the cost of chip production and sales. It is compatible with NPU processors as well as CPU and GPU.

Its difference becomes even more significant when it supports more types of AI applications and chips. It was necessary to develop as many compilers as 'deep learning platform type and chip type' in the past, but it became possible to replace the compilers with one highly versatile nest compiler now.

While releasing the Nest Compiler as an open-source, ETRI also revealed a reference model where the Nest Compiler was applied to an AI chip developed by ETRI internally to revitalize the related industry ecosystem. This is the first time when both software and hardware for AI chip development have been disclosed.

The research team announced that this disclosure is significant in that Nest Compiler can play a pivotal role in the current AI chip ecosystem where development is fragmented.

ETRI also applied Nest Compiler to the high-performance AI chips which were internally developed by a Korean AI chip startup. ETRI plans to expand the scope of supporting deep learning compilers by collaborating with related companies, and it is also promoting the commercialization of AI chip application services through specialized software companies. Moreover, ETRI plans to contribute to creating new

services by improving the performance and convenience required for developing AI application services.

Taeho Kim, Assistant Vice President of AI SoC Research Division, said, "The release of the standard deep learning compiler open source is to revitalize the Korean AI chip ecosystem. Technology cooperation is underway to apply the technology to various AI chip companies."

More information: Deep Learning Compiler on Github: github.com/etri/nest-compiler

Provided by National Research Council of Science & Technology

Citation: AI core software 'Deep Learning Compiler' developed (2022, March 8) retrieved 17 April 2024 from https://techxplore.com/news/2022-03-ai-core-software-deep.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.