

OPERATOR COMPILATION METHOD AND APPARATUS

ABSTRACT

An operator compilation method and apparatus are disclosed, to implement tensor data processing in a compute-in-memory architecture, and improve processing parallelism of computing tasks in the compute-in-memory architecture. The method includes: obtaining a compilation operator, where the compilation operator indicates a compilation rule for source code; compiling the source code based on the compilation operator, to obtain one or more tensor instructions, where the tensor instruction is used to execute a computing task in a compute-in-memory system, and the tensor instruction includes one or more of the following instruction segments: an operation code, a tensor address, a data type, a data bit width, and a tensor length; and storing the tensor instruction into an instruction storage list, where the instruction storage list is used to store a compiled tensor instruction.

Fig. 2