

INTISARI

Jaringan syaraf tiruan (JST) banyak diimplementasikan pada perangkat *mobile* yang menggunakan sumber daya terbatas. Keterbatasan sumber daya mengakibatkan penurunan kecepatan komputasi sistem dan mengakibatkan JST tidak optimal diimplementasikan pada perangkat *mobile*. Masalah ini dapat diselesaikan dengan penggunaan akselerator berbasis FPGA. Metode yang diusulkan untuk menyelesaikan masalah ini adalah mempercepat proses konvolusi dengan aritmetika terdistribusi yang memiliki karakteristik penggunaan sumber daya yang lebih hemat dan hanya menggunakan komponen memori. Penggunaan konvolusi berbasis memori membuat sistem mudah diimplementasikan di FPGA kelas apa pun karena tidak memerlukan komponen seperti DSP yang hanya terdapat pada FPGA kelas atas. Ketersediaan komponen memori yang melimpah pada FPGA juga memungkinkannya untuk melakukan eksplorasi berupa lebar data sistem dan jumlah elemen proses pada kernel konvolusi. Hasil penelitian merupakan metode akselerasi konvolusi berbasis memori beserta peningkatan penggunaan sumber daya pada FPGA ketika lebar data operasi ditingkatkan dan jumlah kernel direduksi. Hasil implementasi konvolusi 8 bit beserta fungsi aktivasi pada FPGA Xilinx Artix-7 Nexys-4 seri XC7A100T1CSG324C memanfaatkan sumber daya berupa: LUT sebanyak 647 *slice* dari 63.400 (1,02%), FF sebanyak 474 *slice* dari 126.800 (0,37%), IO sebanyak 8 blok dari 210 (3,81%) dan BUFG sebanyak 2 blok dari 32 (6,25%). Frekuensi maksimum yang dapat diterima oleh sistem adalah 68,96MHZ.

Kata kunci—FPGA, akselerasi proses, aritmatika terdistribusi, jaringan syaraf tiruan

ABSTRACT

Recently using of neural network (NN) in mobile device is increased. The using of mobile device which has limited resource is slowing down computation speed and made NN not optimize when implemented in mobile device. This problem can be done by using FPGA based accelerator. The method we proposed is using FPGA based accelerator on convolutional process by using distributed arithmetic algorithm. Distributed arithmetic convolves with only using memory as main resource, made this method easy to be implemented in any class of FPGA. Furthermore, using only memory means that we can explore data width of system and reduction of processing element for comparing used resource in FPGA. The result of the research contains about implementation of FPGA based accelerator for one layer CNN and corresponding of increased data width and reduction of processing element by used resource in FPGA. The implemented system in FPGA Xilinx Artix-7 Nexys-4 XC7a100T100T1CSG324C series is using resource: 647 slice LUT (1,02%), 474 slice FF (0,37%), 8 block IO (3,81%) and 2 block BUFG (6,52%). The maximum acceptable frequency in system is 68,96MHZ.

Keywords—FPGA, process acceleration, distributed arithmetic, neural network