

УДК 004.31

РАЗРАБОТКА ПРОТОТИПА ВЫЧИСЛИТЕЛЬНОЙ ПОДСИСТЕМЫ НЕЙРОСЕТЕВОГО ПРОЦЕССОРА НА ПРОГРАММИРУЕМОЙ ЛОГИЧЕСКОЙ ИНТЕГРАЛЬНОЙ СХЕМЕ

Табунщик С.М. (Университет ИТМО)

Научный руководитель – к.т.н., Быковский С.В.
(Университет ИТМО)

Введение. Сверточные нейронные сети являются одним из основных алгоритмов решения задач в области дополненной реальности. Одной из таких задач является расчет карты глубины. Поэтому одной из актуальных задач в процессе разработки системы дополненной реальности является создание нейросетевого процессора для ускорения выполнения сверточных нейронных сетей.

Так как для работы систем дополненной реальности требуется выполнять вычисления в реальном времени и использовать автономные источники питания, то разрабатываемый процессор должен соответствовать требованиям по энергопотреблению и производительности.

Существующие нейросетевые процессоры, которые позволяют ускорять вычисление сверточных нейронных сетей, являются закрытыми продуктами коммерческих компаний. Некоторые компании предоставляют онлайн-доступ к вычислительным ресурсам кластеров из нейросетевых процессоров, но это не позволяет получить готовый продукт и использовать его для внедрения в систему дополненной реальности.

Нейросетевые процессоры, которые могут быть приобретены в виде готового продукта, также не подходят для создания автономной системы дополненной реальности из-за больших габаритов и большого энергопотребления [1].

Целью работы является создание прототипа нейросетевого процессора на программируемой логической интегральной схеме (ПЛИС) для аппаратного ускорения работы сверточных нейронных сетей и дальнейшего тестирования, улучшения и внедрения в системы дополненной реальности.

Основная часть. В работе описываются технические решения, выбранные для создания прототипа вычислительной подсистемы нейросетевого процессора на программируемой логической интегральной схеме (ПЛИС). В качестве микроархитектуры вычислительной подсистемы нейросетевого процессора используется разработанная тензорная микроархитектура. Основным вычислительным узлом в разработанной микроархитектуре является систолический массив. В исходном коде описания вычислительной подсистемы предусмотрена возможность настраивать параметры внутренних компонентов на этапе синтеза аппаратуры.

По результатам разработки прототипа вычислительной подсистемы нейросетевого процессора получены следующие результаты:

- 1) Исходный код описания вычислительной подсистемы на языке описания аппаратуры SystemVerilog,
- 2) Схематическое представление и описание внутреннего строения всех блоков вычислительной подсистемы на уровне RTL,
- 3) Потактовый план работы всех блоков вычислительной подсистемы,
- 4) Описание протоколов взаимодействия вычислительной подсистемы с другими подсистемами,
- 5) Описание формата входных и выходных данных.

Выводы. Разработанный прототип вычислительной подсистемы нейросетевого процессора на ПЛИС может быть использован для тестирования возможностей и эффективности выполнения сверточных нейронных сетей для решения задач по расчету карты глубины в реальном

времени, а также для дальнейшего улучшения и расширения поддерживаемых нейронных сетей.

Список использованных источников:

1. Norman P. Jouppi, Cliff Young et al. In-Datacenter Performance Analysis of a Tensor Processing Unit // arXiv:1704.04760 [Электронный ресурс] – Режим доступа: <https://doi.org/10.48550/arXiv.1704.04760> (дата обращения: 15.02.2023).