

РАЗРАБОТКА АРХИТЕКТУРЫ НЕЙРОСЕТЕВОГО ПРОЦЕССОРА ДЛЯ АППАРАТНОГО УСКОРЕНИЯ РАБОТЫ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

*С.М. Табуницкий, Университет ИТМО, Санкт-Петербург.
Научный руководитель – доцент, к.т.н, С.В. Быковский, Университет ИТМО, Санкт-Петербург.*

Аннотация

В работе рассматривается повышение эффективности аппаратного ускорения работы сверточных нейронных сетей с помощью разработки собственной архитектуры нейросетевого процессора. Предлагается архитектура нейросетевого процессора, ее описание и варианты ее параметризации.

Введение.

Одной из актуальных задач в области дополненной реальности является создание ускорителя выполнения нейронных сетей. Данный ускоритель должен быть эффективным по энергопотреблению и производительности, чтобы производить расчет карты глубины или выполнять другие схожие задачи в реальном времени.

На данный момент существуют готовые решения позволяющие ускорять вычисление нейронных сетей, но они не подходят для использования в системах дополненной реальности либо из-за своих больших габаритов и большого энергопотребления, либо из-за невозможности получить готовый продукт для внедрения в системы дополненной реальности.

Целью работы является разработка архитектуры нейросетевого процессора для аппаратного ускорения работы сверточных нейронных сетей и дальнейшего использования в системах смешанной реальности для расчета карты глубины.

Основная часть.

В работе для аппаратного ускорения работы сверточных нейронной сети предлагается разработать собственную тензорную архитектуру нейронного процессора, которая позволит снизить энергопотребление и повысить производительность. Выбор архитектуры нейросетевого процессора производится на основе результатов сравнительного анализа архитектур существующих нейросетевых процессоров.

По результатам разработки архитектуры нейросетевого процессора получены следующие результаты:

- 1) Схематическое представление и описание архитектуры нейросетевого процессора
- 2) Описание вариантов параметризации архитектуры нейросетевого процессора
- 3) Описание принципов взаимодействия блоков нейросетевого процессора
- 4) Потактовый план работы блоков нейросетевого процессора

Выводы.

Разработанная архитектура может быть использована в учебном процессе для изучения технологии аппаратного ускорения работы нейронных сетей, а также для дальнейшей реализации прототипа нейросетевого процессора и тестирования его эффективности на реальных задачах.