

## ОТЗЫВ

официального оппонента Исупова Константина Сергеевича  
на диссертационную работу Валуевой Марии Васильевны  
«Разработка методов и алгоритмов построения цифровых устройств интеллектуального  
анализа визуальных данных»,  
представленную на соискание ученой степени кандидата технических наук по  
специальности 2.3.5 – Математическое и программное обеспечение вычислительных  
систем, комплексов и компьютерных сетей

**Актуальность темы.** Сверточные нейронные сети (СНС) являются перспективным инструментом для решения задач распознавания изображений и получили в последние годы широкое применение во встроенных системах обработки визуальных данных. Вместе с тем постоянно растущая практическая потребность в увеличении скорости работы устройств, а также постоянное усложнение архитектуры и увеличение размеров нейронных сетей делает актуальной проблему проектирования высокопроизводительных аппаратных ускорителей нейросетевых устройств для обеспечения быстрой работы и экономически выгодной эксплуатации таких систем. В связи с этим тема диссертационного исследования Валуевой Марии Васильевны, посвященная проектированию быстродействующих алгоритмов и аппаратных реализаций глубоких нейронных сетей для систем обработки визуальных данных в условиях ограниченных ресурсов, является актуальной. В своей работе Валуева М.В. для улучшения характеристик нейронной сети при ее аппаратной реализации использует систему остаточных классов (СОК) с модулями специального вида.

**Научная новизна.** Основными результатами, составляющими научную новизну диссертации, являются следующие:

1. Разработан метод проектирования сверточных слоев СНС и его программная реализация на языке описания аппаратуры VHDL, который основан на методе Винограда и арифметике СОК с модулями специального вида.

2. Разработан алгоритм проектирования аппаратной реализации СНС, в котором учитывается выбор модулей СОК, способ представления весовых коэффициентов в памяти устройства и предложенные в диссертации методы проектирования отдельных компонентов СНС.

**Практическая значимость.** В диссертационной работе представлены оригинальные программные реализации разработанных подходов и алгоритмов, что подтверждается наличием 6 свидетельств о государственной регистрации программ для ЭВМ, а также актами о внедрении. Полученные результаты можно рекомендовать к

использованию учащимися ВУЗов по направлениям «Прикладная математика и информатика», «Информатика и вычислительная техника».

**Достоверность и обоснованность научных результатов.** Все основные теоретические положения и выводы, сформулированные в диссертационной работе Валуевой М.В., являются обоснованными. При получении основных теоретических положений автор использовал методы высшей алгебры, теории чисел, вычислений в системе остаточных классов, а также методы математического, программного и аппаратного моделирования. Математические выкладки выполнены с должной степенью строгости. Рекомендации по использованию результатов диссертации сформулированы на основании совокупности полученных аналитических оценок и результатов моделирования, поэтому их обоснованность не вызывает сомнения.

Достоверность результатов, полученных в диссертации, подтверждается их внедрением на предприятии, апробацией на научных мероприятиях международного и всероссийского уровня, а также публикациями автора, среди которых имеются статьи в журналах из перечня ВАК РФ, статьи в изданиях, включенных в аналитические базы Web of Science и Scopus, материалы конференций и свидетельства о государственной регистрации программы для ЭВМ.

**Содержание и структура диссертации.** Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы (114 наименований), списка рисунков, списка таблиц и двух приложений. Общий объем диссертации составляет 154 страницы машинописного текста.

В первой главе представлено описание СНС и проведен анализ методов улучшения технических характеристик СНС. Предложено использование СОК с модулями специального вида  $2^\alpha$  и  $2^\alpha - 1$  для улучшения характеристик аппаратной реализации СНС. Сформулирован алгоритм процесса проектирования аппаратной реализации СНС с вычислениями в СОК.

Во второй главе рассмотрена задача увеличения производительности цифровых фильтров, которые применяются для реализации операции свертки в СНС. Предложены архитектуры устройств, реализующих операцию свертки на основе метода Винограда и арифметики СОК с модулями специального вида. Проведен теоретический анализ параметров разработанных устройств на основе «unit-gate»-модели и аппаратного моделирования на ПЛИС. Полученные результаты показывают, что разработанные цифровые фильтры в большинстве случаев имеют лучшую производительность по сравнению с известными аналогами.

В третьей главе рассмотрены методы реализации вычислительно сложных «немодульных» операций в СОК и предложены устройства для их выполнения. Разработан метод преобразования чисел из ПСС в СОК с модулями специального вида и его программная реализация на языке описания аппаратуры. Предложены архитектуры устройств определения знака числа, сравнения чисел в СОК, а также архитектура устройства обратного преобразования в СОК на основе Китайской теоремы об остатках с дробными величинами (КТОд). Разработаны методы проектирования устройств, реализующих функцию активации ReLU на основе КТОд и операцию выбора максимального элемента из окрестности на основе КТОд. Приведенный теоретический анализ и аппаратное моделирование показали, что разработанные архитектуры устройств реализации вычислительно сложных операций в СОК обладают меньшей задержкой.

В четвертой главе представлен алгоритм проектирования компонентов СНС с вычислениями в СОК и архитектура СНС, разработанная по предложенному алгоритму. Использование модулей специального вида позволяет избежать необходимости выполнения трудоемкой операции нахождения остатка от деления по модулю. Результаты эксперимента показали, что разрядность весовых коэффициентов может быть уменьшена без потери точности распознавания СНС. В сверточных слоях сети предложено использование цифровых фильтров на основе предлагаемого модифицированного метода Винограда с вычислениями в СОК. Приведенные результаты аппаратного моделирования показали, что использование метода Винограда позволяет увеличить производительность устройства в 4,3–5,56 раз, а применение предлагаемого подхода организации вычислений в СОК увеличивает производительность устройства на 2,84–12,13% по сравнению с рассмотренными аналогами.

В заключении приведены основные результаты и выводы по диссертационному исследованию. В приложениях представлены акт о внедрении результатов диссертации и свидетельства о регистрации программ для ЭВМ по теме диссертации.

**Замечания по диссертации.** По тексту работы имеются следующие замечания:

1. В качестве одной из причин, приводящих к необходимости разработки специализированных устройств нейросетевой обработки данных, автор называет недостаточную производительность процессоров общего назначения (CPU) и графических ускорителей (GPU). При этом не приводит какие-либо оценки, подтверждающие данное заявление. Начиная с микроархитектуры Volta все GPU компании NVIDIA оснащаются специальными блоками тензорных ядер, позволяющими повысить эффективность выполнения операций машинного обучения. Более того, для повышения производительности вычислений одновременно может быть задействовано множество

GPU. Аналогично, многие современные CPU поддерживают широкие векторные регистры и арифметику половинной точности для повышения производительности машинного обучения. Поэтому, при всей важности задачи разработки эффективных ускорителей нейросетевых устройств на ПЛИС, требует уточнения утверждение, что производительность CPU и GPU недостаточна для решения задач машинного обучения.

2. В разделе 2 разработанные архитектуры устройств фильтрации по методу Винограда сравниваются с известными аналогами при разрядности входных данных 8, 16 и 32 бита. Однако не указано, как данная разрядность соотносится с произведением модулей СОК (M). Если входные данные могут изменяться в полном диапазоне СОК (разрядность входных данных соответствует разрядности M), то в процессе вычисления свёртки необходимо выполнять контроль переполнения динамического диапазона, что существенно повлияет на технические характеристики разработанных устройств. Если же допустимые значения входных данных ограничиваются некоторой величиной для предотвращения переполнения, то следует показать, как вычисляется эта величина, и выбирать набор модулей СОК заведомо большим, чем разрядность входных данных.

3. В тексте диссертации имеется ряд грамматических неточностей и опечаток, не оказывающих влияния на общее восприятие работы.

Отмеченные недостатки не влияют на положительную оценку научных и практических результатов диссертационного исследования.

**Заключение.** Диссертация Валугевоу М.В. «Разработка методов и алгоритмов построения цифровых устройств интеллектуального анализа визуальных данных» выполнена на актуальную тему, является законченной научно-квалификационной работой, обладает научной новизной и имеет практическую ценность. Автореферат отражает основные полученные научные результаты диссертационной работы.

Таким образом, считаю, что диссертация Валугевоу Марии Васильевны соответствует всем требованиям Положения о порядке присуждения ученых степеней, предъявляемым к кандидатским диссертациям, а её автор заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Официальный оппонент

кандидат технических наук,

доцент кафедры электронных вычислительных машин,

ФГБОУ ВО Вятский государственный университет

Собственноручную подпись  
*Михайленко Е.Н.* заверяю  
Начальник управления по работе  
с персоналом  
Михайленко Е.Н.

Исупов Константин Сергеевич

29.09.2023