

ДОПОЛНИТЕЛЬНОЕ ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА
24.2.437.05, СОЗДАННОГО НА БАЗЕ ФЕДЕРАЛЬНОГО
ГОСУДАРСТВЕННОГО АВТОНОМНОГО ОБРАЗОВАТЕЛЬНОГО
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ «ЮЖНО-УРАЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)» МИНИСТЕРСТВА НАУКИ
И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ,
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА НАУК

аттестационное дело № _____
решение диссертационного совета от 15.11.2024 года, № 57

В соответствии с письмом Департамента аттестации научных и научно-педагогических работников Министерства науки и высшего образования Российской Федерации от 24.09.2024 г. № 10/97/3 на основании рекомендации Президиума ВАК при Минобрнауки России от 20.09.2024 г. № 23/1-зд и заключения экспертного совета ВАК при Минобрнауки России по управлению, компьютерным и информационным наукам от 10.09.2024 г. диссертация Гончаренко Александра Игоревича на тему «Высокопроизводительные нейронные сети глубокого обучения для устройств с низкими вычислительными ресурсами» на соискание ученой степени кандидата технических наук по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ направлена на дополнительное заключение в диссертационный совет 24.2.437.05, созданный на базе федерального государственного автономного образовательного учреждения высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)» Министерства науки и высшего образования Российской Федерации, 454080, г. Челябинск, пр. Ленина, д. 76, приказ от 11 апреля 2012 года № 105/нк.

На заседании 14.05.2024 г. диссертационный совет 24.1.028.01, созданный на базе федерального государственного бюджетного учреждения науки

Института автоматки и электрометрии Сибирского отделения Российской академии наук Министерства науки и высшего образования Российской Федерации (630090, г. Новосибирск, пр. Академика Коптюга, д. 1, приказ от 28 марта 2020 года № 255/нк), принял решение присудить Гончаренко Александру Игоревичу ученую степень кандидата технических наук по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ.

Диссертация «Высокопроизводительные нейронные сети глубокого обучения для устройств с низкими вычислительными ресурсами» по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ принята к защите 08.02.2024 г. (протокол заседания № 1) диссертационным советом 24.1.028.01, созданным на базе федерального государственного бюджетного учреждения науки Института автоматки и электрометрии Сибирского отделения Российской академии наук Министерства науки и высшего образования Российской Федерации (630090, г. Новосибирск, пр. Академика Коптюга, д. 1, приказ от 28 марта 2020 года № 255/нк).

Соискатель Гончаренко Александр Игоревич, 30.04.1992 года рождения, в 2015 г. окончил федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», в 2020 г. освоил программу подготовки научно-педагогических кадров в аспирантуре федерального государственного автономного образовательного учреждения высшего образования «Новосибирский национальный исследовательский государственный университет».

Диссертация выполнена на кафедре систем информатики факультета информационных технологий федерального государственного автономного образовательного учреждения высшего образования «Новосибирский национальный исследовательский государственный университет» Министерства науки и высшего образования Российской Федерации.

Научный руководитель – Нежевенко Евгений Семёнович, доктор технических наук, ведущий научный сотрудник лаборатории информационной оптики федерального государственного бюджетного учреждения науки Института автоматки и электрометрии Сибирского отделения Российской академии наук.

Официальные оппоненты:

– Оселедец Иван Валерьевич, доктор физико-математических наук, профессор РАН, директор АНО «Институт искусственного интеллекта»;

– Куликов Виктор Александрович, кандидат технических наук, Senior ML Scientist 2, Picsart AI Research

дали положительные отзывы о диссертации.

Ведущая организация – федеральное государственное бюджетное образовательное учреждение высшего образования «Новосибирский государственный технический университет», г. Новосибирск, в своем положительном заключении, подписанном Спектором Александром Аншелевичем, доктором технических наук, профессором, профессором кафедры теоретических основ радиотехники и утвержденном Отто Артуром Исааковичем, кандидатом технических наук, проректором по научной работе и инновациям, указала, что, исходя из актуальности, новизны, научной и практической значимости представленной работы, можно сделать заключение, что диссертация «Высокопроизводительные нейронные сети глубокого обучения для устройств с низкими вычислительными ресурсами», представленная на соискание ученой степени кандидата технических наук по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ, выполнена на высоком научном уровне. Она отвечает требованиям пп. 9 – 11, 13, 14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 г. № 842, предъявляемых к кандидатским диссертациям а ее автор Гончаренко Александр Игоревич заслуживает присуждения ученой степени кандидата технических наук по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ.

Результаты диссертационной работы опубликованы в 4 работах, из которых 4 в изданиях, индексируемых международными базами данных:

1. Гончаренко, А.И. Исследование применимости низкоразрядных представлений чисел с плавающей запятой для эффективных вычислений в нейронных сетях / А.И. Гончаренко, А.Ю. Кондратьев // Автометрия. — 2020. —Т. 56, № 1. — С. 93—99. (Scopus, входит в перечень отечественных изданий, которые входят в международные реферативные базы данных и системы цитирования по состоянию на 30 марта 2020 г.)

2. On practical approach to uniform quantization of non-redundant neural networks / A. Goncharenko, A. Denisov, S. Alyamkin, E. Terentev // Lecture Notes in Computer Science. — 2019. — Т. 11728. — С. 349—360. (Scopus, материалы конференции)

3. Trainable thresholds to uniform quantization of non-redundant neural networks / A. Goncharenko, A. Denisov, S. Alyamkin, E. Terentev // Lecture Notes in Computer Science. — 2019. — Т. 11507. — С. 302—312. (Scopus, материалы конференции)

4. Low-power computer vision: Status, challenges, and opportunities / S. Alyamkin, ..., A. Goncharenko, G. Xuyang [и др.] // IEEE Journal on Emerging and Selected Topics in Circuits and Systems. — 2019. — Т. 9, № 2. — С. 411—421. (Scopus, 44 соавтора публикации)

На диссертацию и автореферат поступили отзывы:

1) Мулляджанова Рустама Илхамовича, доктора физико-математических наук, заведующего лабораторией прикладных цифровых технологий ММЦ ММФ, федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Бондаренко Ивана Юрьевича, научного сотрудника лаборатории прикладных цифровых технологий ММЦ ММФ, федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет». Отзыв положительный, содержит вопросы:

1. Какие преимущества нового алгоритма квантования, предложенного автором, перед алгоритмами адаптивного квантования на основе гессииана или иных способов анализа чувствительности в глубоких нейронных сетях? 2. Почему не рассматривались наиболее популярные для многих классов задач нейросетевые модели трансформерного типа с механизмом внимания? Возможно, в механизме внимания есть какие-либо особенности, ограничивающие применение предложенного подхода нахождения разрядности для специализированных типов данных?

2) Окунева Алексея Григорьевича, кандидата химических наук, и.о. директора Института интеллектуальной робототехники, федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет». Отзыв положительный, содержит следующие замечания:

1. Низкое качество изображения под номером 2 в автореферате. 2. При описании раздела 2.7 в автореферате следовало бы повторить конкретные

выводы, которые были сделаны в рамках предыдущих глав. Это облегчило бы понимание работы при чтении. 3. Некоторое количество опечаток.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований: предложен и реализован новый алгоритм квантования для моделей произвольного типа на основе тонкой настройки масштабирующих коэффициентов для порогов квантования, при этом время, затраченное на тонкую настройку сети, после применения алгоритма значительно ниже (от 5 до 10 раз, в зависимости от архитектуры нейронной сети), чем в большинстве современных работ в данной области, при незначительном падении точности (менее 1%) относительно оригинальной модели; предложен и реализован алгоритм перемасштабирования весовых коэффициентов для процедуры скалярного квантования для ограниченной функции активации ReLU6, наиболее распространенной в мобильных архитектурах нейронных сетей; предложена и реализована процедура нахождения разрядности для специализированных типов данных. Предложенный механизм не требует дополнительной тонкой настройки сети, что позволяет упростить внедрение нейронных сетей в сложные программно-аппаратные комплексы. Данная процедура применялась к разнообразным архитектурам сверточных нейронных сетей, что может свидетельствовать о ее универсальности; доказана применимость разработанных методов и подходов к созданию нейросетевых моделей, ориентированных на работу в мобильных устройствах в условиях ограниченной оперативной памяти и низкого электропотребления; введено понятие выброса как фактора, снижающего точность нейронных сетей при квантовании.

В то же время необходимо отметить, что диссертация не содержит новых математических методов моделирования изучаемой задачи. В работе отсутствуют исследования, посвященные разработке качественных или аналитических методов исследования математических моделей, описание численных методов, реализованных в предложенных алгоритмах. Не содержится комплексных исследований научных и технических проблем с применением современных технологий математического моделирования и вычислительного эксперимента, реализованных эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента.

Теоретическая значимость исследования обусловлена тем, что предложен новый подход к устранению влияния выбросов посредством обновления порогов квантования, что ведет к уменьшению шага дискретизации при квантовании и, как следствие, снижению количества ошибок, возникающих в нейронной сети. Предложенный подход был теоретически проанализирован при помощи метода обратного распространения ошибки.

В то же время теоретическая значимость исследования недостаточно обоснована, отсутствуют аналитические или экспериментальные исследования, связанные с вопросами устойчивости, сходимости предложенных методов и алгоритмов. Отсутствует обоснование выбора архитектуры нейронных сетей, нейросетевых моделей, реализованных в предложенных алгоритмах.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

разработанный подход был реализован в виде программного комплекса для обучения сверточных нейронных сетей на ЭВМ. Получаемые таким образом нейросетевые модели могут использоваться для произвольных задач на малоомощных вычислителях. В частности, описанный в данной диссертации подход позволил ускорить алгоритм детектирования лица на мобильном устройстве пользователя с ARM процессором. Также разработанный алгоритм стал частью программной платформы EENNT, позволяющей оптимизировать вычислительную сложность нейронных сетей произвольной архитектуры. Процедура подбора разрядности порядка и мантиссы была использована как один из основных модулей для исследования оптимизации архитектуры аппаратных ускорителей на основе систолического массива.

Отметим, что в диссертации отсутствуют документы, подтверждающие внедрение или использование разработанного программного обеспечения в компаниях, указанных на странице 10 текста диссертации. Акты о внедрении результатов диссертационной работы и свидетельство о регистрации программы для ЭВМ были продемонстрированы соискателем на заседании диссертационного совета.

Оценка достоверности результатов исследования выявила: достоверность полученных результатов подтверждена проведением большого числа тестов на реальных данных. Для измерения точности использовались общепринятые метрики. Полученные экспериментальные результаты хорошо согласуются с

теоретическими выкладками. Основная идея предложенного подхода состоит в использовании концепции дистилляции, в соответствии с которой квантовая модель представляет собой сеть-ученик, а оригинальная модель – сеть-учитель.

В то же время в работе не использованы ранее известные подходы к решению поставленной задачи на основе обучения с подкреплением и эволюционным программированием, которые дают лучшее качество обучения во многих случаях. Не обоснован выбор автора в пользу подхода обучения с учителем и используемых нейросетевых моделей. Согласно п. 10 Положения о присуждении ученых степеней предложенные автором диссертации решения должны быть аргументированы и оценены по сравнению с другими известными решениями. К сожалению, никаких сравнений с аналогичными известными решениями в диссертации не предоставлено.

Личный вклад соискателя состоит в разработке и реализации подходов к оптимизации производительности нейронных сетей. Как указано в диссертации и автореферате разработанные автором подходы изложены в главах 2 и 4 диссертации. Также автором проведен теоретический анализ разработанной процедуры дообучаемых порогов на основе алгоритма обратного распространения ошибки.

Однако, все указанные соискателем публикации выполнены в соавторстве. В диссертации, автореферате, заключении организации не указана степень авторского вклада, не указано, какие из выносимых на защиту результатов диссертации опубликованы, что не соответствует п. 11 Положения о присуждении ученых степеней.

В диссертационной работе А.И. Гончаренко не содержатся все три составляющие специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ: математическое моделирование, численные методы, комплексы программ. Отсутствуют новые научные результаты в области численных методов; исследования, посвященные разработке качественных, аналитических и экспериментальных методов исследования математических моделей для оценки достоверности; обоснования отличий полученных результатов от аналогов, что не позволило автору в достаточной мере доказать и обосновать основные положения и результаты, выносимые на защиту в соответствии с паспортом специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ. Диссертация не

соответствует пп. 9 – 11 Положения о присуждении ученых степеней по научной специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ.

В ходе дополнительного рассмотрения диссертации А.И. Гончаренко были высказаны следующие критические замечания:

1. Математическая модель какой нейронной сети исследуется?
2. Можем ли мы считать, что предложенный Вами метод квантования порогов является численным?
3. В качестве научной новизны результата Вы предлагаете метод квантования порогов. В тексте диссертации не сказано, как Вы это используете в стохастическом градиентном спуске.

Соискатель Гончаренко А.И. ответил на задаваемые ему в ходе заседания вопросы и привел собственную аргументацию:

1. Данные модели ближе всего к системе зрения человека. В частности, мы говорим про сверточные нейронные сети. Если же мы говорим про модели для обработки последовательностей, то там это, скорее, близко именно к мозгу человека, к глубокой лимбической системе.

2. Да, потому что он отделен от метода численной оптимизации. Суть во взаимной синергии заключается в том, что предложен поиск порогов как обучаемых величин. Для того чтобы их искать как обучаемые величины мы модифицируем градиентный спуск или его аналоги.

3. Согласен, что в диссертации это явно не отражено. У меня была ориентация на практический аспект использования результатов моей диссертационной работы.

На основании результатов тайного голосования по вопросу соответствия диссертации соискателя Гончаренко Александра Игоревича установленным критериям, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ, диссертационный совет пришел к заключению о несоответствии диссертации соискателя Гончаренко Александра Игоревича на тему «Высокопроизводительные нейронные сети глубокого обучения для устройств с низкими вычислительными ресурсами» установленным критериям, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук по

специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ.

При проведении тайного голосования диссертационный совет в количестве 15 человек, из них 6 докторов наук по специальности и отрасли наук рассматриваемой диссертации, участвовавших в заседании, из 20 человек, входящих в состав совета, по вопросу «Диссертация соискателя Гончаренко Александра Игоревича соответствует установленным критериям, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук по специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ» проголосовали: «Да» – 2, «Нет» – 13.

Председатель

диссертационного совета



Шестаков Александр Леонидович

Ученый секретарь

диссертационного совета

Манакова Наталья Александровна

Дата оформления заключения 15 ноября 2024 г.