

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Лимоновой Елены Евгеньевны «Биполярная морфологическая аппроксимация нейрона для уменьшения вычислительной сложности глубоких сверточных нейронных сетей», представленную на соискание ученой степени кандидата технических наук по специальности 1.2.2 – «Математическое моделирование, численные методы и комплексы программ»

Актуальность

В своей диссертационной работе Лимонова Елена Евгеньевна исследует вопросы повышения эффективности вычислений в нейросетевых моделях. Сегодня нейронные сети повсеместно используются при решении самых разных задач анализа изображений, видео, аудио и текстовых данных. В частности, при анализе визуальных данных наиболее распространены глубокие нейросетевые модели, содержащие большое количество слоев. Такие сети (например, EfficientNet, YOLO, ResNet) обеспечивают высокую точность решения поставленных задач, требуя при этом значительных вычислительных ресурсов, так как содержат десятки миллионов параметров и выполняют миллиарды операций. По этой причине настройка и применение подобных сетей сопряжено с использованием дорогостоящих рабочих станций или удаленных высокопроизводительных серверов. В последнем случае возникают дополнительные вопросы организации канала передачи данных. В зависимости от характера данных могут возникать юридические ограничения, необходимость решения вопросов информационной безопасности.

По указанным причинам привлекательным решением является использование предварительно настроенных нейронных сетей непосредственно на конечных устройствах. Однако в качестве таких устройств все чаще выступают не высокопроизводительные рабочие станции, а автономные и мобильные устройства, обладающие ограниченными вычислительными ресурсами. Таким образом, актуальной является задача достижения современного высокого уровня качества решения прикладных задач, обеспечиваемого достаточно сложными моделями, при заданных требованиях к скорости обработки и ограничениях на энергопотребление.

Лимонова Е.Е. предлагает решать эту проблему с помощью новой модели биполярного морфологического нейрона. Автор уделяет особое внимание вопросам построения аналогов распространенных сверточных нейросетей с использованием

предложенной модели, вопросам обучения таких нейросетей, а также исследованию точности их работы и вычислительной эффективности на различных платформах. При этом показанная возможность построения аналогов, сопоставимых, а иногда и превосходящих по качеству базовые решения при меньшей латентности (на специализированных устройствах), еще раз подчеркивает актуальность проведенных исследований.

Обзор диссертационной работы

Структура и содержание работы. Диссертационная работа включает в себя введение, 3 главы, заключение и 2 приложения. Текст диссертации изложен на 138 страницах.

Во введении Лимонова Е.Е. представляет тему работы, обосновывает ее актуальность, формулирует цели и задачи диссертационного исследования.

В Главе 1 автор вводит необходимые понятия для постановки задачи повышения вычислительной эффективности нейросетевой модели. Для этого она рассматривает различные модели нейронов: классический математический нейрон, морфологический нейрон, морфологический нейрон с дендритами и спайковые нейроны, структуру нейронных сетей, основные типы слоев и принципы организации их в нейросетевую модель. Далее в первой главе рассматриваются существующие методы повышения вычислительной эффективности нейронных сетей, их достоинства и недостатки. В конце главы формулируются цель и задачи исследования.

В Главе 2 автор рассматривает вычисления в классическом математическом нейроне и предлагает модель биполярного морфологического нейрона, выполняющего преобразование входных данных с использованием операций сложения, взятия максимума, логарифмирования и потенцирования. Такая модель позволяет избавиться от вычислительно сложной операции умножения. Логарифмирование и потенцирование выполняется по основанию 2, что позволяет снизить их вклад в общий объем вычислений. Далее Лимонова Е.Е. рассматривает вычислительную сложность слоев сверточных биполярных морфологических нейронов и показывает, что добиться снижения латентности можно при реализации модели на программируемых логических интегральных схемах или специализированных устройствах. При этом из-за того, что аппроксимация выполняется различным образом для входных сигналов и весов разного

знака, наблюдается небольшой рост числа логических вентилей, необходимых для реализации слоя. Поэтому автор предлагает использовать аппроксимации Митчелла и Шраудольфа для реализации нелинейных функций логарифмирования и потенцирования. Такие кусочно-линейные аппроксимации указанных функций в несколько раз уменьшают сложность и латентность соответствующих модулей.

В Главе 3 приводятся численные эксперименты с сетями на основе биполярных морфологических нейронов. Автор показывает, что и прямая аппроксимация классических нейронов, и обучение сети целиком методом градиентного спуска не позволяют добиться качественного распознавания. Поэтому Лимонова Е.Е. предлагает выполнять аппроксимацию послойно и сквозным образом обучать полученную сеть. Результаты работы этого метода автор демонстрирует на сверточных нейронных сетях с различным количеством слоев. В диссертации рассмотрены классические модели с архитектурой LeNet для базовых задач распознавания, более сложные и допускающие построение глубоких моделей остаточные архитектуры ResNet и предназначенная для задач сегментации архитектура U-Net. Результаты обучения показали, что для более простых задач аппроксимацию можно выполнить без заметного снижения точности, в то время как качество полностью преобразованных моделей с архитектурами ResNet и U-Net существенно снизилось. По этой причине рекомендовано использовать частично преобразованные сети, обеспечивающие сопоставимые, а иногда и превосходящие по характеристики качества при меньшей латентности. В конце главы кратко описан реализованный автором программный комплекс, с помощью которого были выполнены численные эксперименты.

В заключении Лимонова Е.Е. резюмирует основные результаты работы.

В списке использованной литературы содержится 135 наименований.

Приложения содержат информацию о двух зарегистрированных программах для ЭВМ и внедрении результатов диссертационной работы.

Апробация работы. Основные результаты работы докладывались на профильных российских и международных конференциях:

- 39-я междисциплинарная школа-конференция «Информационные технологии и системы» (Сочи, Россия, 2015),
- The 9th International Conference on Machine Vision (Ницца, Франция, 2016),

- The 12th International Conference on Machine Vision (Амстердам, Нидерланды, 2019),
- The 13th International Conference on Machine Vision (Рим, Италия, 2020),
- The 25th International Conference on Pattern Recognition (Милан, Италия, 2020).

Публикации. Результаты по теме диссертации изложены в 10 печатных работах. Из них одна работа опубликована в журнале, рекомендованном ВАК, 8 работ – в изданиях, индексируемых международными базами цитирования Scopus и/или Web of Science, 1 – в сборнике трудов конференции. Получено два свидетельства о регистрации программ для ЭВМ.

Значимость полученных результатов

Теоретическая значимость работы состоит в разработке нейрона с морфологической структурой, позволяющего создавать нейросети с морфологическими слоями. Приведенные результаты дают основания полагать, что нейронные сети, включающие такие слои, могут обладать точностью, сопоставимой с известными решениями, однако требуют адаптации существующих методов обучения. Считаю, что изучение предложенной модели представляет интерес и потенциально способно повысить качество и эффективность решения прикладных задач.

Практическая значимость работы определяется разработанным алгоритмическим и программным обеспечением, позволяющим выполнять инициализацию и настройку нейросетей на основе предложенной модели нейрона. Практическая значимость работы подтверждается использованием результатов работы в приложениях ООО «Смарт Энджинс Сервис» и АО «Тинькофф Банк», а также в процессе принятия решений при проектировании новых устройств АО «МЦСТ».

Степень обоснованности и достоверности результатов работы

Все положения и выводы диссертации детально описаны и полностью обоснованы. Достоверность результатов подтверждается точностью приведенных математических доказательств, а также результатами численных экспериментов, выполненных с использованием разных входных данных и различных нейросетевых моделей. Результаты работы Лимоновой Е.Е. являются непротиворечивыми, согласуются с результатами,

публикуемыми в научных изданиях. Работа апробирована на профильных российских и международных конференциях.

Новизна исследования

Проведенные в рамках диссертационной работы исследования и полученные результаты являются новыми с научной точки зрения. Считаю, что научной новизной обладают следующие полученные автором результаты:

- новая модель биполярного морфологического нейрона, выполняющая преобразование входных данных с использованием операций сложения, взятия максимума, логарифмирования и потенцирования;
- доказательство того, что нейронная сеть из биполярных морфологических нейронов позволяет равномерно приблизить произвольную непрерывную на компакте функцию с заранее заданной точностью;
- новый метод аппроксимации нейросетевых моделей, в котором исходные слои обученной нейросетевой модели послойно преобразуются в их биполярные морфологические аппроксимации, и результирующая модель далее дообучается сквозным образом;
- результаты численных экспериментов, включая исследование метода послойного преобразования и дообучения, а также решение задач классификации (на наборах данных MNIST, CIFAR10, машиночитаемых строк паспортов Российской Федерации) и бинаризации рукописных текстов (на наборе данных DIBCO 2017);
- оценки количества вентиля и латентности, полученные для рассмотренных слоев, состоящих из биполярных морфологических нейронов.

Замечания по содержанию и оформлению диссертации

В диссертационной работе Лимоновой Е.Е. можно выявить следующие недостатки:

1. В работе рассматривается такой известный способ оптимизации нейросетей как квантование, и приводятся результаты его применения для изображений символов из паспортов РФ (п.3.2.2), но сравнение с ним модели на основе биполярного морфологического нейрона отсутствует.

2. В работе не рассматривается вопрос применения разработанного подхода на графических ускорителях (GPU). Было бы интересно увидеть границы эффективного применения предложенного подхода к построению нейросетей для распространенных архитектур GPU фирм NVIDIA и AMD.
3. В главе 2 описание рисунков 2.7 и 2.8 не содержит обоснования выбранного диапазона аргумента, неясно насколько он информативен.
4. Автор приводит соотношение характеристик (числа вентиля и латентности) классических и биполярных морфологических сверточных слоев в таблице 8, однако было бы интересно также привести оценку соотношения характеристик для какой-либо модели нейронной сети целиком, а также для частично преобразованных моделей, которые рассматриваются в главе 3.
5. В работе присутствует ряд описок и опечаток: «широко используется в современные нейросетевых моделях» (стр. 23), «сеть из БМ нейронов с одним входом» (стр. 44), «Благодаря этому можно рассматривать достаточно рассматривать» (стр. 46), «стандарт был разработан Институт инженеров» (с. 52).

Указанные выше замечания носят, в основном, рекомендательный характер и не оказывают существенного влияния на общую положительную оценку работы.

Общая оценка работы

Диссертационную работу Лимоновой Е.Е. «Биполярная морфологическая аппроксимация нейрона для уменьшения вычислительной сложности глубоких сверточных нейронных сетей» следует квалифицировать как завершенное научное исследование, выполненное на актуальную тему. Полученные результаты обоснованы и имеют теоретическое и практическое значение. Положения, выносимые на защиту, обладают научной новизной и отражают результаты работы. Эти результаты опубликованы в рецензируемых научных изданиях, в полной мере прошли апробацию на профильных семинарах и международных конференциях, а также внедрены в деятельность ряда организаций. Автореферат достаточно полно отражает содержание, основные научные положения и результаты диссертационной работы. Работа выполнена по специальности 1.2.2 – «Математическое моделирование, численные методы и комплексы программ» и соответствует паспорту специальности.

Заключение

Диссертация «Биполярная морфологическая аппроксимация нейрона для уменьшения вычислительной сложности глубоких сверточных нейронных сетей» является законченной научно-исследовательской работой, выполненной на хорошем уровне, и соответствует требованиям Положения о присуждении ученых степеней, а ее автор, Лимонова Елена Евгеньевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 1.2.2 – «Математическое моделирование, численные методы и комплексы программ».

Официальный оппонент,
доцент кафедры геоинформатики и
информационной безопасности Федерального
государственного автономного образовательного
учреждения высшего образования
«Самарский национальный исследовательский
университет имени академика С.П. Королева»
(Самарский университет),
кандидат технических наук,

 / Мясников Е.В.
« 1 » Февраль 2023 г.

Сведения об организации:

Федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» (Самарский университет)

Адрес: ул. Московское шоссе, д. 34, г. Самара, 443086

ssau@ssau.ru

+7 (846) 267-43-70

Подпись Мясникова Е.В. удостоверяю.

Ученый секретарь Самарского университета  *Васильев ИИ*



02.02.2023