

ОТЗЫВ

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

В настоящее время нейронные сети укрепились в качестве одного из основных механизмов решения задач распознавания образов и обнаружения событий, перейдя из разряда перспективных технологий, применяемых в исследовательских проектах, в разряд технологий продуктовых, используемых как в составе больших корпоративных вычислительных центров, так и в пользовательских мобильных устройствах. При этом как в первом, так и во втором случае, определяющим фактором применимости нейросетей является соотношение производительности, стоимости и потребляемой мощности вычислительного оборудования, необходимого для выполнения нейросетевых вычислений с требуемой пропускной способностью и задержкой.

Типичные нейросети состоят из множества однотипных вычислительных блоков, допускающих параллельное выполнение, при этом зачастую слабо чувствительных к погрешностям вычислений. В связи с этим использование микропроцессоров общего назначения для нейросетевых вычислений неэффективно. В качестве промежуточного этапа в специализации вычислительного оборудования для нейросетевых вычислений стали популярны графические процессоры, но и они обладают избыточной с точки зрения нейросетей общностью функциональности, теряя в эффективности. В последние годы все больше распространены становятся специализированные аппаратные акселераторы, предназначенные для оптимального применения уже обученных и оптимизированных нейросетей к пользовательским данным. Данные акселераторы применяются либо в серверной среде вместо графических процессоров для снижения стоимости потребляемой электроэнергии, либо, типично для мобильных устройств, дополняют графический процессор, экономя за счет своей эффективности энергию батареи. Соответственно, поиск эффективно вычисляемых представлений нейросетей (в т.ч. допуская необходимость разработки специализированной аппаратуры) в настоящее время является весьма актуальной задачей.

Решению именно этой задачи — разработки эффективно вычисляемых элементов нейросетей — посвящена диссертация Лимоновой Е.Е. В отличие от многих современных работ по данной теме, где исследуется возможность использования вычислений с малой разрядностью без потери точности нейросети, в данном диссертационном исследовании рассматриваются возможные альтернативы более фундаментальной вычислительной структуре — нейрону, основанному на умножении. Поскольку умножение требует довольно большого количества логических вентилях сравнительно с такими операциями, как сложение или взятие максимума, рассматривается возможность использования нейронов на основе именно этих двух более легковесных операций, т.н. биполярных морфологических нейронов.

Особенностью исследования, повышающей его практическую ценность, является сфокусированность на использовании морфологических нейронов как механизма аппроксимации классических нейронов с сохранением архитектуры нейросети, т.е. накопленная экспертиза по обучению нейросетей не перестает быть релевантной при использовании предлагаемой модели — она явно предлагает методику постепенной замены

традиционных нейронов на морфологические с одновременным дообучением сети, что приводит к высокой степени сохранения исходного качества распознавания при снижении требований к вычислительной платформе.

Из автореферата диссертации Лимоновой Е.Е. видно, что проведена значительная теоретическая и экспериментальная работа: 1) предложена аппроксимация классической модели нейрона – биполярный морфологический нейрон; 2) доказана теорема, показывающая, что выразительная мощность биполярных морфологических нейронных сетей не уступает выразительной мощности классических моделей; 3) предложен метод обучения для рассмотренной конструкции, который проверен на трех классах нейросетевых архитектур, используемых в решении задач машинного зрения. Более того, автор провел первичное моделирование аппаратной реализации биполярного морфологического нейрона и показал, что наблюдается снижение латентности сверточных слоев на 30-40%. Результаты изложены в 10 работах (1 ВАК РФ, 8 публикаций в международных базах цитирования, 1 работа в сборнике трудов конференции), а также используются в 2 зарегистрированных программах для ЭВМ.

В работе есть следующие недостатки: 1) в кратком описании главы 3 указано, что при сравнении времени работы классической и преобразованных моделей на мобильном процессоре с ARM время снизилось на 20%, однако из краткого описания, представленного в автореферате, не совсем ясно для какой конкретной задачи проводились эти замеры; 2) на стр.12 используется термин латентность модулей, при этом не вполне понятно, как связано уменьшение латентности схемы ВМ-нейрона в сравнении с классическим нейроном, и итоговое возможное быстродействие устройства, реализующего вм-нейроны. Однако указанные недостатки не влияют на суть работы и являются незначительными.

На основании вышеизложенного можно сделать вывод, что работа Лимоновой Е.Е. на тему «Биполярная морфологическая аппроксимация нейрона для уменьшения вычислительной сложности глубоких сверточных нейронных сетей» выполнена на высоком научном уровне и отвечает всем требованиям ВАК, предъявляемым к диссертационным исследованиям на соискание ученой степени кандидата технических наук, полученные результаты особенно ценны в области проектирования микропроцессоров следующих поколений, 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».

Логинов Вадим Евгеньевич

к.т.н.

начальник отдела "Компоненты системы программирования и библиотеки" отделения "Системы программирования" ПАО ИНЭУМ им И.С.Брука

«13» февраля 2023 г.

Сведения об организации:

ПАО «Институт электронных управляющих машин им. И.С. Брука»

Адрес: 119334, Москва, ул. Вавилова, д. 24

Телефон: +7 (499) 135 44 49

E-mail: ineum@ineum.ru

