

М. Л. Абдулвадудов¹, К.Л. Вахидова², С.М. Ногамирзаев²

¹ Научно-производственный центр «Электронные вычислительно-информационные системы», Россия, г. Зеленоград

² Федеральное государственное бюджетное образовательное учреждение высшего образования "Грозненский государственный нефтяной технический университет имени академика М.Д. Миллионщикова"

364051, Чеченская Республика, г. Грозный, пр-т им. Х.А. Исаева, 100

ПРИМЕНЕНИЕ МЕТОДОВ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ НЕЙРОСЕТЕВОГО АНАЛИЗА ГЕТЕРОГЕННЫХ СИСТЕМ НА КРИСТАЛЛЕ В ЗАДАЧАХ РОБОТОТЕХНИКИ

M. L. Abdulvadudov¹, K. L. Vakhidova², S. M. Nogamirzaev²

¹ Scientific and Production Center "Electronic Computing and Information Systems", Russia, Zelenograd

² Federal State Budgetary Educational Institution of Higher Education "Grozny State Petroleum Technological University named after Academician M.D. Millionshchikov"

364051, Chechen Republic, Grozny, Ave. named after Kh.A. Isaev, 100

APPLICATION OF METHODS FOR INCREASING THE EFFICIENCY OF NEURAL NETWORK ANALYSIS OF HETEROGENEOUS SYSTEMS-ON-CHIP IN ROBOTICS TASKS

В статье представлено исследование методов повышения эффективности выполнения нейросетевых алгоритмов на гетерогенных системах на кристалле (СнК) RoboDeus, ориентированные на применение в робототехнических системах. Приводится анализ архитектуры СнК RoboDeus, включая DSP-ядро Elcore50, и демонстрируется оптимизация работы сверточных нейронных сетей на примере детектора объектов YOLO V5s. Предложенные подходы к распараллеливанию вычислений, эффективному использованию памяти и аккумуляторных регистров позволяют достичь значительного прироста производительности при выполнении задач компьютерного зрения в реальном времени. Результаты экспериментов показывают, что использование многоядерной обработки на базе RoboDeus обеспечивает частоту кадров до 33.3 FPS для задачи детектирования объектов на изображениях разрешения 640×640. Обсуждаются перспективы интеграции оптимизированных нейросетевых алгоритмов в автономные робототехнические комплексы для навигации, распознавания окружения и взаимодействия с объектами

Ключевые слова: гетерогенные вычислительные системы, RoboDeus, YOLO V5s, сверточные нейронные сети, DSP-ядро, архитектура набора инструкций (АНК), система на кристалле (СнК).

The aim of this work is to study methods for improving the efficiency of neural network algorithms on heterogeneous RoboDeus systems-on-a-chip (SoCs), aimed at application in robotic systems. An analysis of the RoboDeus SoC architecture, including the Elcore50 DSP core, is presented, and optimization of convolutional neural networks is demonstrated using the YOLO V5s object detector as an example. The proposed approaches to parallelizing computations, efficient use of memory, and accumulator registers enable significant performance gains when executing computer vision tasks in real time. Experimental results show that the use of multi-core processing based on RoboDeus provides a frame rate of up to 33.3 FPS for the task of detecting objects in 640×640 images. Prospects for integrating optimized neural network algorithms into autonomous robotic systems for navigation, environment recognition, and interaction with objects are discussed.

Keywords: heterogeneous computing systems, RoboDeus, YOLO V5s, convolutional neural networks, DSP core, instruction set architecture (ISA), system on a chip (SoC).

Современная робототехника предъявляет высокие требования к вычислительным платформам: необходимость обработки больших объёмов сенсорных данных (изображений, лидаров, аудио) в реальном времени, энергоэффективность, компактность и надёжность. Ключевую роль в таких системах играют алгоритмы искусственного интеллекта, в частности, свёрточные нейронные сети (CNN), используемые для распознавания образов, детектирования объектов, сегментации сцены. Однако их выполнение на традиционных CPU часто оказывается недостаточно производительным, что стимулирует разработку специализированных гетерогенных архитектур, сочетающих различные типы процессорных ядер.

Искусственный интеллект возник как способ имитации естественного интеллекта, базирующегося на нейронных сетях человеческого мозга и ставшего основой человеческой культуры, науки и искусства. Нейрокибернетика, возникшая во второй половине сороковых годов прошлого века, дала начало исследованиям по технической имитации решения интеллектуальных задач человеком. В шестидесятые годы эта волна исследований уступила место имитации элементов человеческого мышления на основе формальной логики. В начале восьмидесятых (после работ Дж. Хопфилда), в связи с появлением алгоритмов, которые позволили обучать «скрытые» (не связанные со входами и выходами нейросети) слои нейронов, возникла вторая волна интереса к нейронным сетям. Эти алгоритмы значительно расширили возможности нейронных сетей, породив нейроинформатику – науку об обработке информации и управлении с помощью нейронных сетей.

Несмотря на то, что нейронные сети как научно-технический объект существуют более полувека, основные способы их практической реализации ограничиваются областью программного моделирования в различных средах визуального (и не визуального) проектирования. Программные реализации преобладают над аппаратными. Кроме того, аппаратные реализации вследствие своей относительной дороговизны до сих пор не имеют повсеместного распространения. Однако их удельный вес в общем количестве мировых технологических разработок неуклонно растёт. Впереди планеты всей в данном направлении, как всегда, шагает Япония, где разработки на основе нейросетевых технологий успешно внедряются во множестве образцов бытовой техники, таких как фотоаппараты, микроволновые печи, видеокамеры и т.д. Не отстают и другие развитые страны [22].

Нейросетевые аппаратные решения внедряют в свои продукты такие известные фирмы, как Siemens, Intel, Phillips Research (Нидерланды), 3M Laboratories и многие другие. При этом разработано на удивление много разнообразных аппаратных нейросетевых архитектур. Отечественные фирмы, такие как МЦСТ, НТЦ Модуль и НПЦ ЭЛВИС [1] предлагают свои решения.

Цель работы является повышение эффективности программ нейросетевого анализа для гетерогенных вычислительных систем на основе отечественного процессора 1892BM248 RoboDeus путем внедрения ускоренных алгоритмов нейронных сетей. Для достижения цели в работе необходимо решить следующие задачи: исследовать особенности архитектуры гетерогенных СнК; предложить способы повышения эффективности нейросетевого анализа на гетерогенных СнК; разработка методики реализации нейросетевого алгоритма на гетерогенной СнК; результаты применения методики повышения эффективности нейронной сети YOLO V5s для распознавания и детектирования объектов.

Для решения поставленных задач будут рассмотрены архитектура нейронной сети YOLO V5s и схема вычислителя СнК Robodeus, а также будет составлена схема принципа работы свёрточного нейронного слоя в DSP-ядре СнК RoboDeus.

Гетерогенная топология ЦП

Обычно неоднородность в контексте вычислений упоминается к разным архитектурам набора инструкций (далее АНК), где основной процессор имеет одну, а другие процессоры имеют другую архитектуру (возможно более одной), а не просто другую микроархитектуру (обработка чисел с плавающей точкой — это особый случай, такой процессор не называют гетерогенным).

В прошлом гетерогенные вычисления означали, что разные АНК должны обрабатываться по-разному, в то время как в современном примере системы с гетерогенной системной архитектурой устраняют разницу (для пользователя) при использовании нескольких типов процессоров (обычно CPU и GPU). На одной и той же интегральной схеме, чтобы обеспечить лучшее из 2-х главных задач: общую обработку графического процессора (помимо хорошо известных возможностей графического рендеринга 3D-графики, он также может выполнять математически интенсивные вычисления с очень большими наборами данных), в то время как другие процессоры могут работать с операционной системой и выполнять традиционные последовательные задачи. Недавние результаты показывают [2], что многоядерный процессор с гетерогенной АНК, который использует разнообразие, предлагаемое несколькими АНК, может превзойти лучшую однородную архитектуру с той же АНК на целых 21% с экономией энергии на 23% и сокращением продукта задержки энергии (EDP) на 32%. Пример гетерогенного СнК является RoboDeus, который включает в себя CPU, GPU и DSP.

Архитектура DSP - ядер СнК RoboDeus

В состав DSP-ядра Elcore50 входят следующие основные блоки: процессорное ядро DSP; устройство управления памятью VMMU; L1 кэш инструкций L1 I\$ объемом 4 КВ; L1 кэш данных L1 D\$ объемом 16 КВ; память программ PRAM объемом 64 КВ; память данных XYRAM объемом 512 КВ, которая может быть частично или полностью реконфигурирована в L2 кэш L2\$; контроллер прямого доступа к внешней памяти DMA; коммутатор шин данных MUX; устройство вывода трассы Trace; контроллеры внешних интерфейсов.

DSP-ядро Elcore50 работает в виртуальном адресном пространстве. Трансляция виртуальных адресов в физические выполняется с помощью устройства управления памятью VMMU. В составе DSP-ядра имеется внутренняя статическая память - память программ PRAM и объединенная память данных и программ XYRAM. Обращения к памяти программ выполняются по 128-разрядной шине P, а к памяти данных – по двум 512-разрядным шинам X и Y (в соответствии с рисунком 1).

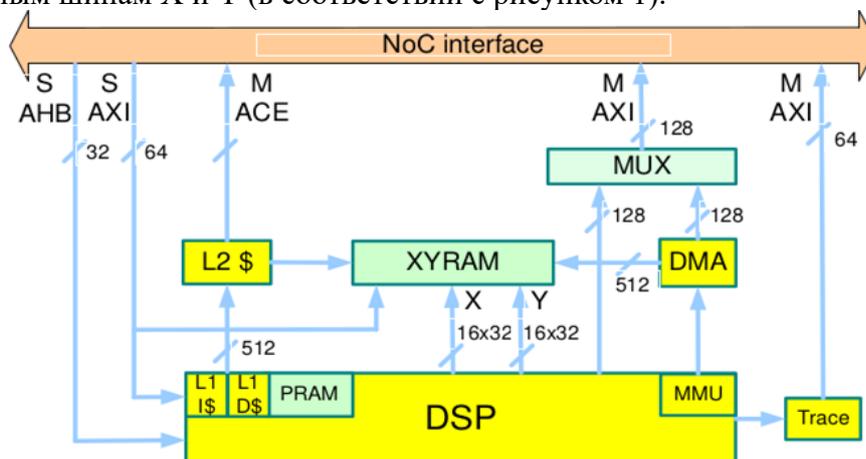


Рисунок 1 – структурная схема DSP – ядра Elcore50

Большинство инструкций выполняются над векторными регистрами в пределах вычислительной секции. Вычислительная секция имеет ширину обработки 64 бита. Данные обрабатываются как упакованный SIMD. В Elcore50 предусмотрено 8 одинаковых вычислительных SIMD секций (в соответствии с рисунком 2). Векторный вычислитель (сопроцессор) EVX оперирует 512-разрядными векторами данных.

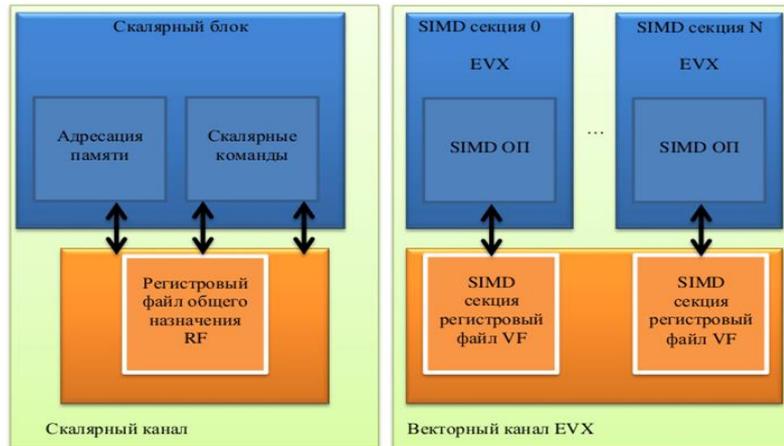


Рисунок 2 – Блоки и регистры одноступенчатого детектора YOLO v5

Архитектура модели YOLO V5s

Поскольку YOLO v5 является одноступенчатым детектором объектов, он состоит из трех важных частей, как и любой другой одноступенчатый детектор объектов: Model Backbone, Model Neck, Model Head (в соответствии с рисунком 3). Модель Backbone в основном используется для извлечения важных функций из заданного входного изображения. В YOLO V5s CSP — частичные сети с перекрестными этапами используются в качестве основы для извлечения богатых информативных функций из входного изображения. Нейросетевой алгоритм CSPNet продемонстрировал значительное улучшение времени обработки в более глубоких сетях. Модель Neck в основном используется для создания пирамид признаков. Пирамиды признаков помогают моделям хорошо обобщаться при масштабировании объектов. Это помогает идентифицировать один и тот же объект с разными размерами и масштабами. Пирамиды функций очень полезны и помогают моделям хорошо работать со скрытыми слоями.

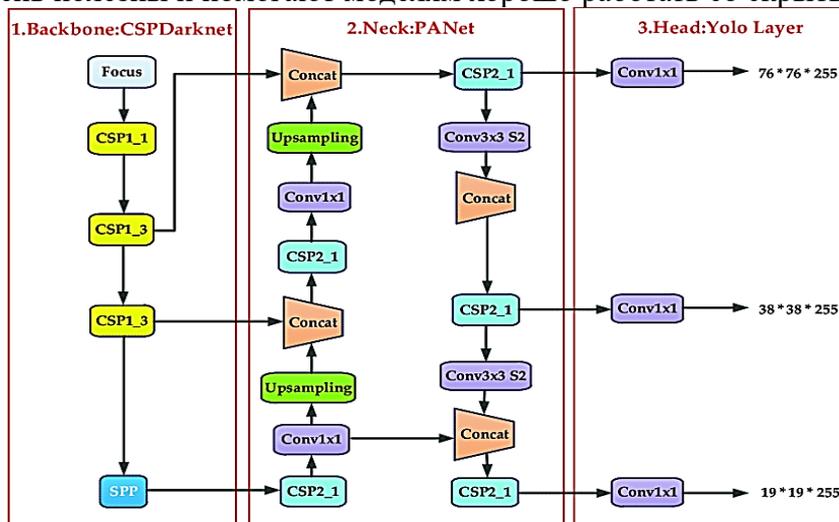


Рисунок 3 – Архитектура нейронной сети YOLO V5s

Модель Head в основном используется для выполнения заключительной части обнаружения. Он применяет рамки привязки к функциям и генерирует окончательные выходные векторы с вероятностями классов, оценками объектности и ограничивающими рамками. В модели YOLO V5s Model Head такая же, как и в предыдущих версиях YOLO V3 и V4.

Выбор функций активации является наиболее важным в любой глубокой нейронной сети. Недавно было введено множество функций активации, таких как: Leaky ReLU, swish и т.д. В YOLO V5s функция активации Leaky ReLU используется в средних/скрытых слоях, а сигмовидная функция активации используется в последнем слое обнаружения.

Алгоритм сверточной нейронной сети

Свёрточные нейронные сети (СНС) составляют вычислительное ядро большинства современных систем машинного зрения, являясь ключевым инструментом для решения задач распознавания образов, детектирования объектов и семантической сегментации в реальном времени. В робототехнике, где требуется высокая скорость реакции на изменяющееся окружение, энергоэффективность и надёжность алгоритмов, оптимизация выполнения операций свёртки приобретает особую значимость. Однако традиционные подходы к реализации СНС на универсальных процессорах зачастую не позволяют достичь требуемой производительности при ограниченных аппаратных ресурсах, что стимулирует активный поиск специализированных методов вычислений, адаптированных под архитектуру гетерогенных систем на кристалле.

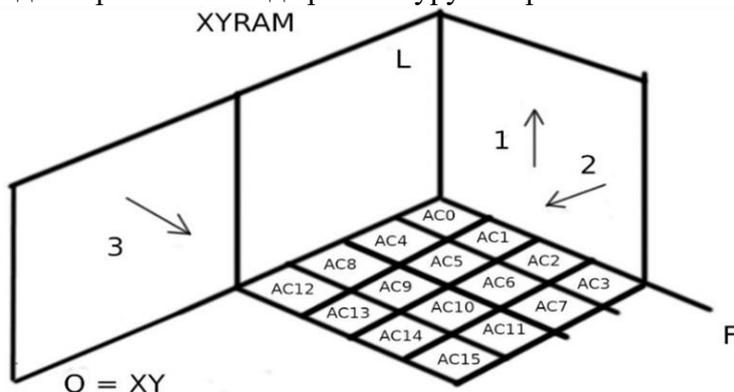


Рисунок 4 – 3-х мерное представление вычислений свертки и расположения аккумуляторов

В соответствии с рисунком 4 изображено 3-х мерное представление сверточного слоя. Пространство LF — данные весов нейронного слоя; пространство QL — входные данные, где $Q = X*Y$; пространство QF — выходные данные. В DSP – ядре SnK RoboDeus присутствуют 16 аккумуляторов, которые накапливают с каждой итерацией скалярно-векторное произведение матриц. Преимущество этой схемы в том, что при расположении аккумуляторов в пространстве QF в виде 4x4, получается оптимальный параллельный код вычислений и загрузок данных. Т.е. весь код загрузок работает параллельно с инструкциями вычислений с помощью VLIW пакетов. Данные в аккумуляторах локализуются по оси QF. Скалярно-векторные операции с накоплением в аккумуляторах проходят первоначально в сторону оси L, затем локализуются данные по оси LF в памяти XYRAM. Со следующей итерацией в пространстве QF в сторону Q можно работать с новыми выходными данными (2-е направление из рисунка) эффективно, т.к. данные весов уже хранятся в памяти XYRAM и времени на загрузку в

память XYRAM не потребуется. И наконец локализируются данные по оси QL в L3 кэше, и работа со следующими данными в пространстве QF в сторону F выполняется эффективно, т.к. входные данные в пространстве QL уже хранятся в L3 кэше и времени на загрузку в L3 кэш не потребуется.

Таким образом для повышения эффективности работы нейросетевых алгоритмов на СнК RoboDeus, а именно на DSP-ядрах ELscore применены следующие подходы: максимальное использование аккумуляторных регистров, оптимальное расположение аккумуляторов (в данном случае по схеме 4x4) для объединения кода загрузок с вычислительными инструкциями в VLIW пакеты, размещение данных фильтров в память XYRAM для повторного использования в итерациях вычислений по направлению Q, размещение входных данных в L3 кэш для повторного использования в итерациях вычислений по направлению F, они позволили значительно повысить производительность алгоритмов на аппаратной платформе. Для улучшения эффективности можно предложить также в начале вычислений инициализировать аккумуляторы нейронами смещения (biases), а в конце вычислений выполнить функцию активации для аккумуляторов.

Результаты вычислений нейронной сети YOLO V5s

Custom Dataset, 4 класса:

Ошибка: диапазон значений (0 - 1): 0.01

1 ядро, типы данных — half(fp16), размеры картинок = 640x640x3: FPS = 3.07

16 ядер, типы данных — half(fp16), размеры картинок = 640x640x3: FPS = 33.3

Coco Dataset, 80 классов

Ошибка: диапазон значений (0 - 1): 0.0016

1 ядро, типы данных — half(fp16), размеры картинок = 640x640x3: FPS = 2.16

16 ядер, типы данных — half(fp16), размеры картинок = 640x640x3: FPS = 21.7

Таблица – Производительность нейросетевых алгоритмов на СнК RoboDeus

Задача	Тип нейросети	1892BM248 «RoboDeus», fps		Jetson AGX Xavier, fps	Jetson Nano, fps
		single stream	пакетная обработка		
Классификация	ResNet 18	31,1	432,5	743 (fp16)/338(fp32)	90,4(fp16)/53
	ResNet 34	21,3	276,6	419(fp16)/191,5(fp32)	50,4(fp16)/28,5
	ResNet 50	17,3	204,5	325(fp16)/115	33,5(fp16)/18,4
	MobileNetv1	43,26	583,2	513 (fp32)	72 (fp32)
Обнаружение	SSD-ResNet34	1,15	11,68	149(fp16)/51,47(fp32)	12(fp16)
	SSD-MobileNetv2	1,28	12,03	232,3/118,2	

Заключение

Стоит отметить, что использование гетерогенных СнК, таких как RoboDeus, в сочетании с оптимизированными нейросетевыми алгоритмами позволяет достичь высокой производительности в задачах компьютерного зрения, актуальных для робототехники. Ключевыми преимуществами являются энергоэффективность, возможность параллельной обработки нескольких потоков данных и адаптируемость архитектуры под различные типы нейронных сетей. Перспективным исследованием является использование режима single-stream обработки, когда несколько DSP-ядер работают над

одним кадром, что может дополнительно повысить эффективность для задач с высоким разрешением. Также перспективным направлением является разработка фреймворка для обучения нейронных сетей непосредственно на целевом аппаратном обеспечении, что позволит сократить цикл разработки робототехнических систем с интеллектуальными функциями.

В работе были представлены архитектура нейронной сети YOLO V5s, архитектура гетерогенного СнК RoboDeus, пример оптимизации сверточной нейронной сети на DSP-ядре СнК RoboDeus. Нейронная сеть YOLO V5s была оптимизирована в режиме multi stream, т.е. каждое DSP - ядро обрабатывало по одному фрейму. Такие внутренние нейронные слои в YOLO V5s, как транспонирование матрицы, умножение и суммирование тензора на тензор стали эффективными с помощью векторных инструкций и приемами программирования кэша XYRAM. В формуле sigmoid экспонента была представлена в виде $e^x = e^{k \ln 2 + r} = e^{k \ln 2} e^r = 2^k e^r$, где $k = \text{round}\left(\frac{x}{\ln 2}\right) = \left\lfloor \frac{x}{\ln 2} + 0.5 \right\rfloor$, $|r| \leq 0.5 \ln 2$ и $r = x - k \ln 2$. Число k было найдено с помощью ассемблерных команд DSP-ядра СнК RoboDeus, а число r было найдено с помощью аппроксимации Лагранжа.

В дальнейшем планируется оптимизация нейронных сетей в режиме single stream, т.е. каждое DSP - ядро будет обрабатывать один и тот же frame по кусочкам. С представленным опытом эффективности работы нейронных слоев можно применить в таких нейронных сетях, которые применяются в NLP, т.е. распознавание речи и обработка текстовых документов, а также в разработке фреймворка обучения нейронных сетей.

Список литературы

1. АО НПЦ «ЭЛВИС». Электронный ресурс: <https://elvees.ru> (дата обращения: 24.10.2025).
2. Галушкин А. И. Теория нейронных сетей: учеб. пособие для вузов / под общ. ред. А. И. Галушкина. М.: ИПРЖР, 2000. Кн. 1. 416 с.
3. Осовский С. Нейронные сети для обработки информации. // Пер. с польского И.Д. Рудинского. - М.: Финансы и статистика, 2002. 344 с.
4. Ростовцев В. С. Искусственные нейронные сети [Электронный ресурс]: учеб. для вузов. 2-е изд., стер. СПб: Лань, 2021. 216 с. Режим доступа: <https://e.lanbook.com/book/160142> (дата обращения: 20.08.2022).
5. Цуриков А. Н. Моделирование и обучение искусственных нейронных сетей [Электронный ресурс]: учеб. пособие. Ростов-н/Д.: РГУПС, 112 с. Режим доступа: <https://e.lanbook.com/book/140610> (дата обращения: 20.08.2022).
6. A Heterogeneous Multi-Core System-on-Chip for Energy Efficient Brain Inspired Computing / Antonio Pullini, Francesco Conti, David Rossi, Michael Gautschi / ResearchGate, 2017. – 6 p.
7. Towards AI. Официальный сайт. Электронный ресурс: <https://towardsai.net> (дата обращения: 24.10.2022).
8. Kunzman, D.M. (2011). Programming Heterogeneous Systems. International Symposium on Parallel and Distributed Processing Workshops. doi:10.1109/IPDPS.2011.377
9. Федоров, М. В. (член-корр. РАН) Новые методы машинного обучения: использование фундаментального знания для решения практических задач: [тезисный онлайн-доклад] [Электронный ресурс] // Программа II Всероссийской школы-семинара по искусственному интеллекту и большим данным в технических, промышленных, природных и социальных системах. – Национальный центр физики и математики (НЦФМ). Дата проведения: 25-29 ноября 2024 г. (дата обращения: 04.04.2025).
10. Чудина, О.В. Теория и практика термической обработки металлов: учебно-методическое пособие к мультимедийному изданию / О.В. Чудина, Г.В. Гладова, А.В. Остроух. М.: МАДИ, 2013. 64 с.
11. Сергеев, А. П. Введение в нейросетевое моделирование: учеб. пособие / А. П. Сергеев, Д. А. Тарасов ; под общ. ред. А. П. Сергеева. Екатеринбург : Изд-во Урал. ун-та, 2017. 128 с. ISBN 978-5-7996-2124-7
12. Хливненко Л. В. Практика нейросетевого моделирования : учебное пособие для вузов / Л. В. Хливненко, Ф. А. Пятакович. 2е изд., стер. Санкт-Петербург : Лань, 2021. 200 с.: ил. Текст : непосредственный. ISBN 978.5.8114.8264.1
13. Вичугов В.Н Модифицированный градиентный алгоритм обучения радиально-базисных нейронных сетей. Известия Томского политехнического университета. Управление, вычислительная техника и информатика. 2009. Т. 315. № 5 с150-152
14. Сидоров Д.И., Новикова О.И. Применение методов машинного обучения для предиктивной диагностики оборудования. Проблемы искусственного интеллекта, 2023, №1(28), С.23-34.

15. Козлов С.В., Фролова М.П. Оптимизация энергозатрат промышленных установок с использованием нейронных сетей. Проблемы искусственного интеллекта, 2021, №4(27), С.67-78.
16. Белов И.А., Орлова Т.К. Адаптивные системы автоматического регулирования с использованием гибридных алгоритмов. Проблемы искусственного интеллекта, 2022, №2(29), С.18-29.
17. Никитина Л.В., Павлов В.Г. Моделирование и анализ сложных технических систем в условиях неопределенности. Проблемы искусственного интеллекта, 2023, №4(31), С.78-90.
18. Федоров К.А., Семенова А.А. Цифровые двойники как основа повышения эффективности производственных линий. Проблемы искусственного интеллекта, 2021, №1(24), С.55-66.
19. Григорьев П.С., Медведева Е.В. Интеллектуальный анализ данных в задачах мониторинга технологических параметров. Проблемы искусственного интеллекта, 2022, №4(31), С.34-45.
20. Титов В.М., Ветрова Н.С. Разработка алгоритмов управления на основе обучения с подкреплением для промышленных роботов. Проблемы искусственного интеллекта, 2023, №2(29), С.61-72.
21. Орлова С.К., Крылов Д.В. Методы повышения устойчивости систем автоматического управления в условиях возмущений. Проблемы искусственного интеллекта, 2021, №3(26), С.41-52.
22. Нурулин Ю. Р., Скворцова И. В., Нурулин М. Ю., Наумов А. С. Разработка человеко-машинных интерфейсов для систем спектрального анализа радиосигналов в реальном времени // Современные наукоемкие технологии. – 2025. – № 3. – С. 40-46. – DOI 10.17513/snt.40321.

References

1. JSC Scientific and Production Center "ELVEES." Electronic resource: <https://elvees.ru> (date accessed: 24.10.2025).
2. Galushkin, A. I. Neural Network Theory: a textbook for universities / edited by A. I. Galushkin. Moscow: IPZhR, 2000. Book 1. 416 p.
3. Osovsky, S. Neural Networks for Information Processing. // Translated from Polish by I.D. Rudinsky. Moscow: Finance and Statistics, 2002. 344 p.
4. Rostovtsev, V. S. Artificial Neural Networks [Electronic resource]: a textbook for universities. 2nd ed., reprinted. St. Petersburg: Lan, 2021. 216 p. – Access mode: <https://e.lanbook.com/book/160142> (date accessed: 20.08.2022).
5. Tsurikov A. N. Modeling and training of artificial neural networks [Electronic resource]: textbook. manual. Rostov-on-Don: RGUPS, 112 p. Access mode: <https://e.lanbook.com/book/140610> (date accessed: 20.08.2022).
6. A Heterogeneous Multi-Core System-on-Chip for Energy Efficient Brain Inspired Computing / Antonio Pullini, Francesco Conti, David Rossi, Michael Gautschi / ResearchGate, 2017. – 6 p.
7. Towards AI. Official website. Electronic resource: <https://towardsai.net> (date accessed: 24.10.2022).
8. Kunzman, D.M. (2011). Programming Heterogeneous Systems. International Symposium on Parallel and Distributed Processing Workshops. doi:10.1109/IPDPS.2011.377
9. Fedorov, M. V. (Corresponding Member of the Russian Academy of Sciences) New Methods of Machine Learning: Using Fundamental Knowledge to Solve Practical Problems: [online abstract report] [Electronic resource] / M. V. Fedorov // Program of the II All-Russian School-Seminar on Artificial Intelligence and Big Data in Technical, Industrial, Natural, and Social Systems. – National Center for Physics and Mathematics (NCPM). – Date: November 25-29, 2024. – (date of access: 04.04.2025).
10. Chudina, O.V. Theory and Practice of Heat Treatment of Metals: a teaching aid for a multimedia publication / O. V. Chudina, G. V. Gladova, A. V. Ostroukh. Moscow: MADI, 2013. 64 p.
11. Sergeev, A. P. Introduction to Neural Network Modeling: a tutorial / A. P. Sergeev, D. A. Tarasov; edited by A. P. Sergeev. Yekaterinburg: Ural University Publishing House, 2017. 128 p. ISBN 978-5-7996-2124-7
12. Khlivnenko, L. V. Practice of Neural Network Modeling: a tutorial for universities / L. V. Khlivnenko, F. A. Pyatakovich. 2nd ed., reprinted. Saint Petersburg: Lan, 2021. 200 p.: ill. Text: direct. ISBN 978.5.8114.8264.1
13. Vichugov V.N. Modified Gradient Algorithm for Training Radial Basis Neural Networks. Bulletin of Tomsk Polytechnic University. Management, Computer Science, and Information Technology. 2009. Vol. 315. No. 5, pp. 150-152
14. Sidorov D.I., Novikova O.I. Application of Machine Learning Methods for Predictive Equipment Diagnostics. Problems of Artificial Intelligence, 2023, No. 1(28), pp. 23-34.
15. Kozlov S.V., Frolova M.P. Optimization of Energy Consumption of Industrial Installations Using Neural Networks. Problems of Artificial Intelligence, 2021, No. 4 (27), pp. 67-78.
16. Belov I. A., Orlova T. K. Adaptive automatic control systems using hybrid algorithms. Problems of Artificial Intelligence, 2022, No. 2 (29), pp. 18-29.
17. Nikitina L. V., Pavlov V. G. Modeling and analysis of complex technical systems under uncertainty. Problems of Artificial Intelligence, 2023, No. 4 (31), pp. 78-90.
18. Fedorov K. A., Semenova A. A. Digital twins as a basis for increasing the efficiency of production lines. Problems of Artificial Intelligence, 2021, No. 1 (24), pp. 55-66.
19. Grigoriev P. S., Medvedeva E. V. Intelligent data analysis in process parameter monitoring tasks. Problems of Artificial Intelligence, 2022, No. 4 (31), pp. 34-45.
20. Titov V.M., Vetrova N.S. Development of control algorithms based on reinforcement learning for industrial robots. Problems of Artificial Intelligence, 2023, No. 2 (29), pp. 61-72.

21. Orlova S.K., Krylov D.V. Methods for increasing the stability of automatic control systems under disturbances. *Problems of Artificial Intelligence*, 2021, No. 3 (26), pp. 41-52.
22. Nurulin Yu. R., Skvortsova I. V., Nurulin M. Yu., Naumov A. S. Development of human-machine interfaces for real-time radio signal spectral analysis systems // *Modern High Technologies*. – 2025. – No. 3. – P. 40-46. – DOI 10.17513/snt.40321.

RESUME

M. L. Abdulvadudov, K. L. Vakhidova, S. M. Nogamirzaev

Application of methods for increasing the efficiency of neural network analysis of heterogeneous systems-on-chip in robotics tasks

Modern robotic systems place high demands on the performance and energy efficiency of computing platforms, especially when processing computer vision data in real time. Traditional processor architectures often fail to handle complex neural network algorithms, such as convolutional neural networks, limiting the navigation and interaction capabilities of autonomous robots. Therefore, a pressing issue is the development and optimization of methods for performing neural network computations on specialized energy-efficient heterogeneous SoCs combining various types of processor cores. Objective, Methodology, and Novelty

The objective of this study is to improve the efficiency of neural network algorithms on the domestically produced 1892BM248 "RoboDeus" heterogeneous system-on-chip (SoC) for robotics applications. The scientific novelty lies in the development and application of a set of optimization methods, including efficient parallelization of computations, optimal memory management, and the use of accumulator registers adapted to the architecture of the Elcore50 DSP core of the RoboDeus SoC.

The methodology of this study included an analysis of the architecture of the heterogeneous RoboDeus SoC and the Elcore50 DSP core, a study of convolutional neural network algorithms (using the YOLO V5s object detector as an example), and the development and practical implementation of optimization methods for this hardware. Experiments were conducted using various datasets to evaluate performance.

The optimizations resulted in a significant increase in performance. For the YOLO V5s neural network, running on 16 RoboDeus SoC cores with half-precision (fp16) data and a 640x640 image resolution, the processing frequency was 33.3 frames per second for the custom dataset and 21.7 frames per second for the COCO dataset. A comparative analysis with other platforms (NVIDIA Jetson AGX Xavier and Jetson Nano) demonstrated the solution's competitiveness in object classification and detection tasks.

Thus, applying the proposed optimization methods on the heterogeneous RoboDeus SoC enables high performance in computer vision tasks relevant to robotics. Key advantages include energy efficiency, parallel processing, and architectural adaptability. Promising areas for further research include single-stream optimization (distributing the processing of a single frame across multiple cores) and adapting the methods to neural network algorithms for natural language processing (NLP). The developed approaches facilitate the creation of high-performance and energy-efficient computing systems for autonomous robotic systems.

РЕЗЮМЕ

М. Л. Абдулвадулов, К.Л. Вахидова, С.М. Ногамирзаев

Применение методов повышения эффективности нейросетевого анализа гетерогенных систем на кристалле в задачах робототехники

Современные робототехнические системы предъявляют высокие требования к производительности и энергоэффективности вычислительных платформ, особенно при обработке данных компьютерного зрения в реальном времени. Традиционные процессорные архитектуры зачастую не справляются с выполнением сложных нейросетевых алгоритмов, таких как свёрточные нейронные сети, что ограничивает возможности

автономных роботов по навигации и взаимодействию с окружением. В связи с этим актуальной задачей является разработка и оптимизация методов выполнения нейросетевых вычислений на специализированных энергоэффективных гетерогенных системах на кристалле, сочетающих различные типы процессорных ядер.

Целью работы является повышение эффективности выполнения нейросетевых алгоритмов на отечественной гетерогенной системе на кристалле (СнК) 1892ВМ248 «RoboDeus» для задач робототехники. Научная новизна заключается в разработке и применении комплекса методов оптимизации, включающих эффективное распараллеливание вычислений, оптимальное управление памятью и использование аккумуляторных регистров, адаптированных под архитектуру DSP-ядра Elcore50 СнК RoboDeus.

Методология работы включала анализ архитектуры гетерогенного СнК RoboDeus и DSP-ядра Elcore50, исследование алгоритмов свёрточных нейронных сетей (на примере детектора объектов YOLO V5s), а также разработку и практическую реализацию методик оптимизации для данного аппаратного обеспечения. Эксперименты проводились с использованием различных наборов данных для оценки производительности.

В результате проведённых оптимизаций достигнуто значительное повышение производительности. Для нейронной сети YOLO V5s при работе на 16 ядрах СнК RoboDeus с данными в формате половинной точности (fp16) и разрешении изображений 640×640 частота обработки составила 33.3 кадра в секунду для кастомного набора данных и 21.7 кадра в секунду для набора данных COCO. Сравнительный анализ с другими платформами (NVIDIA Jetson AGX Xavier и Jetson Nano) показал конкурентоспособность решения в задачах классификации и обнаружения объектов.

Таким образом, применение предложенных методов оптимизации на гетерогенной СнК RoboDeus позволяет достичь высокой производительности в задачах компьютерного зрения, актуальных для робототехники. Ключевыми преимуществами являются энергоэффективность, возможность параллельной обработки и адаптируемость архитектуры. Перспективными направлениями дальнейших исследований являются оптимизация в режиме single-stream (распределение обработки одного кадра между несколькими ядрами), а также адаптация методик для нейросетевых алгоритмов в области обработки естественного языка (NLP). Разработанные подходы способствуют созданию высокопроизводительных и энергоэффективных вычислительных систем для автономных робототехнических комплексов.

Абдулвадуов Магомед Лечиевич, ведущий инженер-программист по встраиваемым системам, АО НПЦ «ЭЛВИС», эл. почта abdulvadudovmagomed@mail.ru тел 8(915)178-67-77 адрес: 124460, г Москва, Зеленоград, ул Конструктора Лукина, 14, стр. 14

Область научных интересов: искусственный интеллект, компьютерное зрение, встраиваемые системы

Вахидова Карина Лечиевна, доцент кафедры «Автоматизация технологических процессов и производств», Грозненский государственный нефтяной технический университет имени академика М.Д. Миллионщикова, 364051, Чеченская Республика, г. Грозный, пр-т им. Х.А. Исаева, 100, тел 8(938)894-22-72 *Область научных интересов:* моделирование технических систем

Ногамирзаев Сулиман Мовлаевич – старший преподаватель кафедры «Технологии машиностроения и транспортных процессов» Грозненский государственный нефтяной технический университет имени академика М.Д. Миллионщикова, эл. почта kafedra2020@mail.ru, адрес: 364051, Чеченская Республика, г. Грозный, пр-т им. Х.А. Исаева, 100, *Область научных интересов:* компьютерное моделирование систем и процессов

Статья поступила в редакцию 23.09.2025