

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION



УДК 004.93'1; 004.932

<https://doi.org/10.37661/1816-0301-2024-21-2-73-85>

Оригинальная статья
Original Article

Сравнительный анализ производительности одноплатных компьютеров для разработки микроархитектурного вычислительного комплекса обнаружения возгораний

Д. А. Павленко

*Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
E-mail: dmitri.pavlenko@gmail.com*

Аннотация

Цели. Целью работы является выбор базовой вычислительной микроплатформы бортового микроархитектурного вычислительного комплекса для обнаружения аномальных ситуаций на территории Республики Беларусь из космоса на основе методов искусственного интеллекта.

Методы. Для выбора вычислительного комплекса используется метод сравнительного анализа. К выбранному оборудованию применяется серия тестов производительности и проводится сопоставительный анализ (бенчмаркинг). Сравнительный и сопоставительный анализы осуществляются в соответствии с требованиями технического задания на текущий проект.

Результаты. Проведены сравнительный анализ и тестирование производительности одноплатных компьютеров Raspberry Pi 4 Model B и Cool Pi 4 Model B, а также ИИ-ускорителя Google Coral USB Accelerator с Google Edge TPU. Сравнительный анализ показал, что Raspberry Pi 4 Model B и Cool Pi 4 Model B полностью соответствуют требованиям технического задания на разработку бортового микроархитектурного вычислительного комплекса обнаружения аномальных ситуаций. При этом Cool Pi 4 Model B хорошо справляется с нейросетевыми вычислениями, но в четыре раза медленнее, чем Google Coral USB Accelerator. Нейросетевые вычисления на Raspberry Pi 4 Model B в 22 раза медленнее, чем аналогичные вычисления на Google Coral USB Accelerator. Cool Pi 4 Model B опережает Raspberry Pi 4 Model B примерно в два-три раза при решении задач копирования и сжатия данных и почти в шесть раз при нейросетевых вычислениях.

Заключение. Несмотря на то что Raspberry Pi 4 Model B подходит под требования технического задания в качестве вычислительной основы, при разработке бортового микроархитектурного вычислительного комплекса обнаружения аномальных ситуаций стоит использовать более мощные альтернативы со встроенным ускорителем нейронных сетей (например, Radxa Rock 5 Model A) либо с дополнительным внешним ИИ-ускорителем (например, сочетание Cool Pi 4 Model B и Google Coral USB Accelerator). Использование Raspberry Pi 4 Model B с дополнительным ИИ-ускорителем также приемлемо и увеличит

скорость вычислений в десятки раз. ИИ-ускорители обеспечивают самые быстрые нейросетевые вычисления, но есть нюансы, связанные с новизной технологий, которые будут исследоваться при дальнейшей разработке.

Ключевые слова: одноплатные компьютеры, нейронные процессоры, ускорители нейронных сетей, ИИ-ускорители, тестирование производительности, сопоставительный анализ

Благодарности. Работа выполнена при финансовой поддержке научно-технической программы Союзного государства «Комплекс-СГ» в рамках НИР № 9СГ2.1-225 от 24.02.2023. Выражается благодарность сотрудникам ОИПИ НАН Беларуси Эдуарду Витальевичу Снежко и Дмитрию Васильевичу Морозову за ценные замечания, сделанные в ходе исследовательской работы.

Для цитирования. Павленко, Д. А. Сравнительный анализ производительности одноплатных компьютеров для разработки микроархитектурного вычислительного комплекса обнаружения возгораний / Д. А. Павленко // Информатика. – 2024. – Т. 21, № 2. – С. 73–85. <https://doi.org/10.37661/1816-0301-2024-21-2-73-85>

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 05.04.2024
Подписана в печать | Accepted 26.04.2024
Опубликована | Published 28.06.2024

Comparative analysis of single-board computers for the development of a microarchitectural computing system for fire detection

Dzmitry A. Paulenka

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Sarganova, 6, Minsk, 220012, Belarus
E-mail: dmitri.pavlenko@gmail.com*

Abstract

Objectives. The purpose of the work is to select the basic computing microplatform of the onboard microarchitectural computing complex for the detection of anomalous situations in the territory of the Republic of Belarus from space on the basis of artificial intelligence methods.

Methods. The method of comparative analysis is used to select a computing platform. A series of performance tests and comparative analysis (benchmarking) are performed on the selected equipment. The methods of comparative and benchmarking analysis are performed in accordance with the terms of reference to the current project.

Results. A comparative analysis and performance testing of Raspberry Pi 4 Model B and Cool Pi 4 Model B single-board computers, as well as AI-accelerator Google Coral USB Accelerator with Google Edge TPU have been performed. The comparative analysis showed that Raspberry Pi 4 Model B and Cool Pi 4 Model B fully meet the terms of reference to the current project. At the same time Cool Pi 4 Model B handles neural network calculations well, but four times slower than similar calculations on Google Coral USB Accelerator. Neural network computations on the Raspberry Pi 4 Model B are 22 times slower than similar computations on the Google Coral USB Accelerator. Cool Pi 4 Model B outperforms Raspberry Pi 4 Model B by the factor of two to three for data copying and compression and almost six times faster for neural network computations.

Conclusion. Despite the fact that Raspberry Pi 4 Model B meets the terms of reference to the project as a computational basis, when developing an on-board microarchitectural computing system for detecting anomalous situations, it is worth using more powerful alternatives with built-in AI-accelerators (e.g., Radxa Rock 5 Model A) or with an additional external AI-accelerator (e.g., a combination of Cool Pi 4 Model B and Google Coral USB

Accelerator). Using a Raspberry Pi 4 Model B with an additional AI-accelerator is also acceptable and will speed up computations by several dozen times. AI-accelerators provide the fastest neural network computations, but there are features related to the novelty of the technology that will be explored in further development.

Keywords: single-board computers, neural processors, neural network accelerators, AI-accelerators, performance testing, comparative analysis

Acknowledgements. The work was carried out with the financial support of the scientific and technical Union State Program "Complex-SG". Special thanks to Eduard Vitalievich Snezhko and Dmitry Vasilievich Morozov, employees of the UIIP NAS of Belarus, for valuable comments made during the research work.

For citation. Paulenka D. A. *Comparative analysis of single-board computers for the development of a microarchitectural computing system for fire detection*. *Informatika [Informatics]*, 2024, vol. 21, no. 2, pp. 73–85 (In Russ.). <https://doi.org/10.37661/1816-0301-2024-21-2-73-85>

Conflict of interest. The author declares of no conflict of interest.

Введение. Современные методы обработки изображений должны использовать нейросетевые методы искусственного интеллекта (ИИ), чтобы быть инновационными и конкурентоспособными. Быстрое увеличение рынка устройств, которые используют специализированные микропроцессоры для эффективного обучения и (или) работы нейронной сети, требует новых способов практической реализации нейросетевых методов на этих устройствах.

Объектом данного исследования являются доступные на рынке микрокомпьютерные решения и одноплатные компьютеры с поддержкой параллельных нейросетевых вычислений. Цель работы – выбор базовой вычислительной микроплатформы бортового микроархитектурного вычислительного комплекса для обнаружения аномальных ситуаций на территории Республики Беларусь из космоса на основе методов ИИ (далее МВК «БортВК»).

В предыдущей научно-технической программе Союзного государства «Технология-СГ» в рамках НИОКР № 3.2.4.1/111/34 от 07.09.2016 был проведен сравнительный анализ одноплатных компьютеров и их аналогов для классификации подстилающих поверхностей Земли [1, 2] и выполнена такая классификация с помощью специализированной сверточной нейронной сети на одноплатном компьютере Raspberry Pi Zero Wireless [3–5]. В результате был разработан микромодуль оперативного распознавания, отбора и сжатия видеоинформации на борту малых космических аппаратов (КА), который является автономным аппаратно-программным комплексом для автоматического обнаружения и классификации изображений подстилающей поверхности Земли¹ (рис. 1).



Рис. 1. Опытный образец микромодуля оперативного распознавания, отбора и сжатия видеоинформации на борту малых КА (полетная часть)

Fig. 1. Prototype of a micromodule for operational recognition, selection and compression of video information on board small spacecraft (flight part)

¹Разработать микромодуль оперативного распознавания, отбора и сжатия видеоинформации на борту малых космических аппаратов : отчет о НИР (заключ.) / ОИПИ НАН Беларуси ; рук. В. А. Ковалев ; исполн.: Д. А. Павленко [и др.]. – Минск, 2020. – 79 с. – № ГР 20164285. – Инв. № 90757.

При дальнейшей разработке автономного аппаратно-программного комплекса данного типа осуществляется переход от задачи классификации подстилающих поверхностей к более сложной задаче обнаружения аномальных ситуаций на спутниковых снимках. После исследований были определены для обнаружения следующие аномальные ситуации:

- последствия пожаров и буреломов в природных экосистемах;
- последствия затопления территорий.

Также предполагается оценка возможности разрабатываемых алгоритмов для обнаружения засухи.

Задача обнаружения аномальных ситуаций использует значительно больше вычислительных мощностей, поэтому для ее решения недостаточно вычислительной мощности одноплатного компьютера Raspberry Pi Zero Wireless.

Общий обзор микрокомпьютерных решений. Областью применения МВК «БортВК» являются поиск и обнаружение аномальных ситуаций и нетипичных визуально детектируемых изменений на спутниковых изображениях на основе методов машинного обучения, ИИ и многоядерного параллелизма. МВК «БортВК» будет выполнен на базе формфактора 1U CubeSat для малых КА. Общая схема работы МВК «БортВК» изображена на рис. 2.

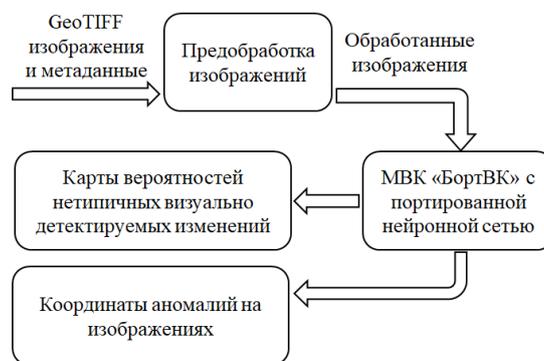


Рис. 2. Общая схема работы МВК «БортВК» без его наземного обеспечения

Fig. 2. General operation scheme of the on-board microcomputing complex MCC "BoardCC" without its ground support

В настоящей статье не рассматривается схема работы наземного программного обеспечения (ПО) для обучения специализированных нейронных сетей и для испытаний функциональных характеристик МВК «БортВК».

Для поддержки параллельных нейросетевых вычислений существуют различные акселераторы нейросетей (AI accelerators), которые отражают эксперименты компаний и стартапов в области аппаратного обеспечения для ИИ. Общее название таких нейросетевых ускорителей по аналогии с центральным процессором (CPU, central processing unit) и графическим процессором (GPU, graphical processing unit) – NPU (neural processing unit), или нейронный процессор.

NPU значительно ускоряет расчеты с использованием нейронных сетей, но есть различные нюансы, связанные с новизной технологий. Важно отметить, что нейронный процессор, в отличие от центрального процессора, нельзя использовать для вычислений общего назначения.

На сегодняшний день существуют следующие разновидности NPU: Tensor Processing Unit (TPU), Neural Network Processor (NNP), Intelligence Processing Unit (IPU), Dataflow Processing Unit (DPU), Vision Processing Unit (VPU), Analog Deep Neural Network (Analog DNN), Associative Processing Unit (APU), AWS Trainium и AWS Inferentia от компании Amazon, Neuromorphic Chip, Quantum Processing Unit (QPU), Photonic Integrated Circuit (PIC) и др.^{2,3}.

²Аппаратное ускорение глубоких нейросетей: GPU, FPGA, ASIC, TPU, VPU, IPU, DPU, NPU, RPU, NNP и другие буквы [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/455353>. – Дата доступа: 04.04.2024.

³Processing Units – CPU, GPU, APU, TPU, VPU, FPGA, QPU [Electronic resource]. – Mode of access: https://primo.ai/index.php?title=Processing_Units_-_CPU,_GPU,_APU,_TPU,_VPU,_FPGA,_QPU. – Date of access: 04.04.2024.

Некоторые нейронные процессоры (Google TPU, Intel Movidius, Nvidia NVDLA, Amazon AWS) являются составной частью компьютеров в виде интегральных схем либо поставляются отдельно как ИИ-ускорители: платы расширения, USB-ускорители, дополнительное оборудование для масштабирования. Примеры USB ИИ-ускорителей: Google Coral USB Accelerator с вычислительной мощностью 4 TOPS (trillions operations per second), RK1808 NPU с вычислительной мощностью 3 TOPS, Orange Pi AI Stick Lite с вычислительной мощностью 2,8 TOPS, Intel Neural Stick 2 с вычислительной мощностью 1 TOPS и др. Современные встроенные NPU одноплатных компьютеров имеют вычислительную мощность 6 TOPS.

Сравнительный анализ, тестирование производительности и выбор компьютерных решений для МВК «БортВК» (см. рис. 2) осуществляются в соответствии с требованиями технического задания (ТЗ) (табл. 1).

Таблица 1
 Требования к составу и параметрам бортовой части МВК «БортВК»

Table 1
 Requirements to the composition and parameters of the on-board part of the MCC "BoardCC"

Пункт ТЗ ToR item	Наименование требования Requirement name	Значение Value
4.1.4	Базовое ПО	TensorFlow, PyTorch, Keras, Python, OpenCV, TensorFlow Lite, TensorRT, NumPy
4.1.5	Архитектуры нейронных сетей, которые используют особенности современных нейронных сетей	MobileNet, EfficientNet, DeepLab, SSD
4.4.1	Масса изделия, г, не более	400
4.4.2	Мощность энергопотребления, Вт, не более	20
4.4.4	Размеры экспериментального образца в конструктиве наноспутника со сторонами, см, не более	10×10×10
4.4.5	МВК «БортВК» должен функционировать под управлением операционной системы (ОС) семейства Linux x86-64 и иметь характеристики не хуже	ОС Linux x86-64, CPU ARM v8 64-bit, RAM 4 Гб, память 64 Гб, разъем USB 2.0, сетевой адаптер wi-fi
4.5.1	На вход экспериментального образца МВК «БортВК» подаются изображения подстилающей поверхности Земли в формате GeoTIFF и вспомогательные файлы с метаданными. На выходе после обработки изображений формируются карты вероятности нетипичных визуально детектируемых изменений и соответствующие относительные координаты этих изменений на изображениях	<i>Входные данные:</i> снимки GeoTIFF и вспомогательные файлы с метаданными <i>Выходные данные:</i> карты вероятности нетипичных визуально детектируемых изменений и соответствующие относительные координаты этих изменений на изображениях

Примечание. В таблице приведены только требования к бортовой части МВК «БортВК» без пояснений, которые используются для выбора микроархитектурного решения. Остальные требования ТЗ не приводятся.

Note. The table shows only the requirements for the onboard part of the MCC "BoardCC" without explanations, which are used to select a microarchitectural solution. The remaining requirements of the technical specifications are not given.

При выборе вычислительного комплекса учитывалась его доступность на коммерческом рынке. В силу существующих ограничений на поставку ряда категорий вычислительной техники в Республике Беларусь их приобретение оказалось затруднительным. В итоге для тестов производительности были отобраны и куплены два одноплатных компьютера и один нейронный процессор (рис. 3):

- одноплатный компьютер Raspberry Pi 4 Model B (далее Pi4)⁴;
- одноплатный компьютер Cool Pi 4 Model B (далее CoolPi)⁵;
- ИИ-ускоритель Google Coral USB Accelerator с Google Edge TPU (далее Coral)⁶.

⁴Raspberry Pi 4 [Electronic resource]. – Mode of access: <https://www.raspberrypi.com/products/raspberrypi-4-model-b>. – Date of access: 05.04.2024.

⁵Cool Pi 4 Model B – A much faster alternative to Raspberry Pi 4 SBC [Electronic resource]. – Mode of access: <https://www.cnx-software.com/2022/12/04/cool-pi-4-model-b-powerful-raspberrypi-4-alternative>. – Date of access: 05.04.2024.

⁶Coral USB Accelerator [Electronic resource]. – Mode of access: <https://coral.ai/products/accelerator>. – Date of access: 05.04.2024.

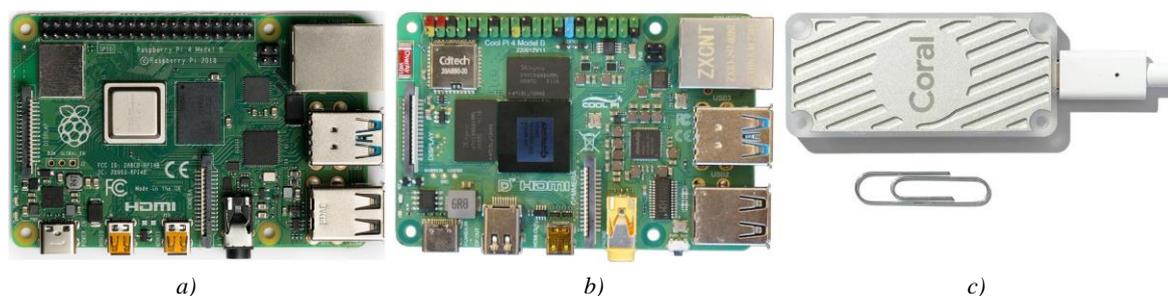


Рис. 3. Raspberry Pi 4 Model B (a), Cool Pi 4 Model B (b) и Google Coral USB Accelerator (c)
Fig. 3. Raspberry Pi 4 Model B (a), Cool Pi 4 Model B (b) and Google Coral USB Accelerator (c)

Среди одноплатных компьютеров, которые доступны экономически и поддерживают параллельные нейросетевые вычисления, можно также выделить одноплатные компьютеры Orange Pi 5 Plus⁷ и Radxa Rock 5 Model A⁸.

Orange Pi 5 Plus идентичен CoolPi. У них одинаковые модели чипсетов и GPU, быстрая память типа eMMC, достаточное количество разъемов для масштабирования. В отличие от CoolPi Orange Pi 5 Plus имеет встроенный NPU мощностью 6 TOPS, что на 2 TOPS (или на 33 %) больше, чем у Coral. Операционная система OrangePi OS похожа на оперативную систему Debian и Ubuntu. Продукты компании Orange хорошо задокументированы и поддерживаются сообществом разработчиков. Предполагается, что Orange Pi 5 Plus должен работать быстрее, чем сочетание CoolPi и Coral.

Очевидным недостатком Orange Pi 5 Plus является размер 7,5×10,0 см, в результате чего могут возникнуть трудности при установке его в корпус малого КА размерами 10×10×10 см (см. табл. 1, п. 4.4.4 ТЗ). Адаптер питания к этому одноплатному компьютеру имеет мощность 5 В · 4 А = 20 Вт, что граничит с техническим требованием к энергопотреблению (см. табл. 1, п. 4.4.2 ТЗ).

Одноплатный компьютер Radxa Rock 5 Model A имеет встроенный NPU 6 TOPS, меньшие размеры 8,5×5,6×1,7 см, идентичен Orange Pi 5 Plus (одинаковый чипсет, но меньше разъемов) и экономически доступен. Он менее популярен и менее задокументирован, чем Raspberry Pi и Orange Pi. Имеет модификацию Radxa Rock 5 Model B на том же чипсете RK3588, но с большим количеством разъемов и размерами 10,0×7,5×2,0 см (формфактор Pico-ITX).

В настоящее время нет необходимости приобретать и тестировать Radxa Rock 5 Model A и Orange Pi 5 Plus. Имеющееся в наличии сочетание CoolPi и Coral будет работать на 30 % медленнее при расчетах на NPU и приблизительно одинаково при расчетах на CPU.

Pi4 и CoolPi удовлетворяют всем требованиям ТЗ. Pi4 и его более новый аналог Raspberry Pi 5 (далее Pi5) являются самыми популярными одноплатными компьютерами, но отстают в вычислительной мощности от более современных моделей одноплатных компьютеров. Чипсеты Broadcom BCM2711 у Pi4 и Broadcom BCM2712 у Pi5 не самые мощные на сегодняшний день, но их требования к энергопотреблению значительно ниже, чем у чипсета Rockchip RK3588s компьютера CoolPi.

CoolPi хуже задокументирован и протестирован, чем Pi4 и Pi5. Один пример из многих, с которыми пришлось столкнуться: включение и работа CoolPi под ОС Ubuntu 22.04 выявили частые зависания системы. Информации о причине зависаний на специализированных сайтах найдено не было. На устранение проблемы с зависаниями было потрачено много времени. Выяснилось, что зависания системы происходят из-за графического интерфейса пользователя (GUI) GNOME. Выполнение ТЗ и функционирование МВК «БортВК» не требуют GUI на одноплатном компьютере. Однако разработка и тестирование будут проходить значительно легче

⁷Orange Pi 5 Plus (32GB) [Electronic resource]. – Mode of access: <http://www.orangepi.org/html/hardWare/computerAndMicrocontrollers/details/Orange-Pi-5-plus-32GB.html>. – Date of access: 05.04.2024.

⁸Radxa ROCK 5A [Electronic resource]. – Mode of access: <http://radxa.com/products/rock5/5a>. – Date of access: 05.04.2024.

при использовании какого-либо интерфейса пользователя. Удаление сервисов GNOME и замена их на аналогичные менеджеры входа в систему LXDM и среду рабочего стола LXDE устранили зависания системы для CoolPi.

Coral не является самостоятельным компьютером, а представляет собой ускоритель нейросетевых вычислений. Он подключается к компьютеру через разъем USB 3.0 Type-C и используется как дополнительный ИИ-ускоритель при вычислениях через библиотеку машинного обучения TensorFlow Lite. Из недостатков Coral стоит отметить, что данный TPU поддерживает только библиотеку TensorFlow Lite и не поддерживает другие библиотеки машинного обучения.

Технические характеристики тестируемых устройств. Сравнение технических характеристик выполнено для Pi4, CoolPi, Coral и настольного персонального компьютера (ПК) (далее Desktop). Тесты для ПК Desktop в НИР не требуются и выполнены исключительно для наглядности и сравнения. Основные технические характеристики тестируемых устройств приведены в табл. 2.

Таблица 2
 Основные технические характеристики тестируемых устройств

Table 2
 Main technical specifications of the tested devices

Техническая характеристика <i>Technical specification</i>	Pi4	CoolPi	Coral	Desktop
Дата выпуска, месяц/год	06/2019	12/2022	01/2020	06/2013
Цена на сайте AliExpress от 24.03.2024, бел. руб.	263	544	309	Примерно 1300
Компания производитель, страна	Raspberry Pi Foundation, Великобритания	Shenzhen Yanyi Technology Co. Ltd, КНР	Google LLC, США	Intel Corporation, США
Физические размеры, см	8,6×5,7×1,7	8,8×5,7×1,7	6,5×3,0×0,8	41,6×41,0×18,0
Масса, г	46,00	54,00	19,53 (с кабелем 36,37)	Более 6000
ОС	Debian GNU / Linux 11 (bullseye)	Ubuntu Linux 22.04.3 LTS (Jammy Jellyfish)	Нет	Microsoft Windows 10
NPU, TOPS	Нет	Нет	4	Нет
Чипсет	Broadcom BCM2711	Rockchip RK3588s	Google Edge TPU and PMIC	Intel Lynx Point Z87, Intel Haswell
Модель CPU	Cortex-A72	Cortex-A76 и Cortex-A55	Нет	Intel i7-4770
Частота CPU, ГГц	1,8	2,4 и 1,8	Нет	3,6
Кол-во ядер CPU, шт.	4	4 и 4	Нет	4
Модель GPU	Broadcom VideoCore VI	Arm Mali-G610	Нет	Nvidia GeForce GTX 760
Размер RAM, ГБ	3,71	3,63	Нет	15,8
Тип памяти, размер	SanDisk SC64G UHS DDR50 SDXC card, 59,5 ГБ	AT2S9C HS400 eMMC 5.1 card, 58,2 ГБ	Нет	Toshiba THNSNJ512GCST SATA-3, 475 ГБ
Разъемы USB, шт.×тип	2×USB3.0, 2×USB2.0, USB-C OTG	2×USB3.0, 2×USB2.0, USB-C OTG	USB-C	4×USB3.0, 4×USB2.0
Сетевой адаптер wi-fi	Есть	Есть	Нет	Есть

Технические характеристики для табл. 2 взяты из спецификаций продукции на сайтах производителей и других источников. Чипсет Rockchip RK3588s обладает четырьмя ядрами Arm Cortex-A76 и еще четырьмя ядрами Arm Cortex-A55. Всего у него восемь ядер двух типов.

Реальные размеры оперативной памяти (RAM) и памяти хранения немного меньше заявленных, однако это обычная практика изготовителей.

Потребляемая электроэнергия. В соответствии с п. 4.4.2 ТЗ (см. табл. 1) мощность энергопотребления вычислительного устройства не должна превышать 20 Вт. По этой причине были проведены измерения электропитания для одноплатных компьютеров Pi4 и CoolPi.

Измерения проводились на сертифицированном источнике питания МНИПИ Б5-84. Измерялись электрическое напряжение в вольтах (В) и сила электрического тока в амперах (А). Мощность энергопотребления в ваттах (Вт) получена перемножением напряжения и силы тока: $Вт = В \cdot А$. Результаты измерений приведены в табл. 3.

Таблица 3
Потребляемая одноплатами компьютерами электроэнергия

Table 3
Electricity consumption by single board computers

Измерение, Вт <i>Measurement, Wt</i>	Pi4	CoolPi
Пиковое электропотребление при нагрузке	5,12 В · 1,20 А = 6,14	5,00 В · 2,06 А = 10,30
Электропотребление при частичной нагрузке	5,00 В · 1,15 А = 5,75	5,00 В · 1,83 А = 9,15
Пиковое электропотребление при включении	Измерение не проводилось	5,00 В · 1,68 А = 8,40
Энергопотребление в выключенном состоянии	5,00 В · 0,32 А = 1,60	Измерение не проводилось

Максимальная и частичная нагрузки на одноплатные компьютеры производились с помощью ПО stress через следующую команду с различными опциями:

```
# sudo apt install stress # install stress software
stress --cpu 8 --io 4 --vm 4 --vm-bytes 256M --hdd 4 --hdd-bytes 1024M --timeout 10s
```

CoolPi способен работать при напряжениях больше 5 В, но измерения силы тока при напряжении 12 В и выше не проводились, потому что напряжение бортовой (внешней) системы питания наноспутника CubeSat будет не более 5 В.

В соответствии с официальной документацией⁹ максимальная мощность энергопотребления Coral составляет 2 Вт. Одновременное измерение энергопотребления одноплатного компьютера и Coral не проводилось. Поэтому при использовании Coral необходимо прибавить еще 2 Вт к пиковому электропотреблению одноплатного компьютера при нагрузке.

Таким образом, мощность энергопотребления Pi4 и CoolPi с дополнительным ИИ-ускорителем Coral не превышает 13 Вт, что соответствует требованию ТЗ (не более 20 Вт).

Тестирование производительности. В ТЗ нет требований к скорости обработки данных. Однако чем быстрее будут обрабатываться данные, тем лучше.

Тесты производительности (бенчмарки) ПК Desktop проведены для наглядности и сравнения, чтобы показать, что современные одноплатные компьютеры не уступают, а иногда даже превосходят настольные компьютеры с датой выпуска более пяти лет. Для ОС Windows ПК Desktop не удалось найти бесплатное кроссплатформенное ПО, поэтому для тестов производительности на ОС Windows были использованы контейнеры Docker: phoronix/pts¹⁰ и ubuntu¹¹.

В табл. 4–6 приведены результаты тестов производительности для трех имеющихся в наличии микроплатформ (Pi4, CoolPi, Coral) и ПК Desktop.

Таблица 4
Тестирование производительности тензорного сопроцессора Coral

Table 4
Performance testing of the Coral tensor coprocessor

Тест, мс <i>Test, ms</i>	Pi4	CoolPi	Desktop
TPU+TFLite	5,2	4,8	4,7
CPU+TF	117,8	19,7	215,5
CPU+TFLite	140,8	23,9	6420,3

Примечание. TPU может работать только с библиотекой TFLite, поэтому четвертое сочетание TPU+TF не применяется.

Note. TPU can only work with the TFLite library, so the fourth combination TPU+TF does not apply.

⁹Coral USB Accelerator datasheet [Electronic resource]. – Mode of access: <https://coral.ai/static/files/Coral-USB-Accelerator-datasheet.pdf>. – Date of access: 05.04.2024.

¹⁰phoronix/pts [Electronic resource]. – Mode of access: <https://hub.docker.com/r/phoronix/pts>. – Date of access: 05.04.2024.

¹¹ubuntu [Electronic resource]. – Mode of access: https://hub.docker.com/_/ubuntu. – Date of access: 05.04.2024.

Таблица 5
 Тестирование производительности с помощью скрипта sbc-bench.sh

Table 5
 Performance testing with the use of sbc-bench.sh script

Тест Test	Pi4	CoolPi	Desktop
memcpy, Мбайт/с	2469	7829	7407
memset, Мбайт/с	3077	24 766	14 722
7-zip, MIPS	5720	15 100	16 870
AES-256, 16кБ, мегахеш	36	1091	584
Троттлинг	Нет	Да	Да

Таблица 6
 Тестирование производительности с помощью ПО Phoronix Test Suite, тест stress-ng

Table 6
 Performance testing with the Phoronix Test Suite software, stress-ng test

Тест, Bogo Ops/s Test, Bogo Ops/s	Pi4	CoolPi	Desktop
CPU Stress	392	852	6806
Matrix Math	1342	14 838	18 742
Memory Copying	395	719	1260

Тесты производительности в табл. 4 проведены для обученной модели нейронной сети MobileNet v3, которая находится в файлах репозитория Coral¹²: tf2_mobilenet_v3_edgetpu_1.0_224_ptq_edgetpu.tflite для TPU и tf2_mobilenet_v3_edgetpu_1.0_224_ptq.tflite для CPU. ПО для проведения тестов производительности Coral TPU разработано специально для данного испытания и находится в файлах tf_lite_benchmarks.py и tf_lite.py репозитория автора исследования¹³.

В первом ряду табл. 4 TPU+TFLite приведены результаты тестов производительности для TPU Coral и библиотеки TensorFlow Lite (TFLite), во втором ряду CPU+TF – тесты производительности для CPU и библиотеки TensorFlow (TF), в третьем ряду CPU+TFLite – тесты производительности для CPU и библиотеки TensorFlow Lite (TFLite).

Для тестов производительности в табл. 5 использовано ПО Томаса Кайзера sbc-bench.sh¹⁴, которое разработано специально для тестов различных моделей одноплатных компьютеров.

В табл. 5 первый ряд memcpy – это скорость копирования данных из одного блока памяти в другой с помощью функции memcpy языка программирования C/C++, второй ряд memset – скорость инициализации блока памяти с помощью функции memset языка программирования C/C++, третий ряд 7-zip – скорость сжатия данных с помощью ПО 7-zip, четвертый ряд AES-256 – скорость шифрования данных с помощью алгоритма AES-256, пятый ряд Троттлинг – наличие или отсутствие троттлинга при большой нагрузке на CPU в течение пяти минут. Троттлинг отсутствует только для Pi4, что связано с плохой системой охлаждения для CoolPi и Desktop.

Для тестов производительности в табл. 6 использовано ПО Phoronix Test Suite¹⁵. Оно включает в себя сотни различных тестов производительности, из которых был выбран тест stress-ng. С помощью стресс-теста stress-ng проведены тесты производительности CPU (ряд CPU Stress), матричной математики (ряд Matrix Math) и копирования блоков памяти (ряд Memory Copying).

¹²google-coral/test_data [Electronic resource]. – Mode of access: https://github.com/google-coral/test_data. – Date of access: 27.03.2024.

¹³Paulenka, D. A. Coral TPU project [Electronic resource] / D. A. Paulenka. – Mode of access: https://github.com/foobar167/junkyard/tree/master/coral_tpu. – Date of access: 05.04.2024.

¹⁴ThomasKaiser/sbc-bench [Electronic resource]. – Mode of access: <https://github.com/ThomasKaiser/sbc-bench>. – Date of access: 05.04.2024.

¹⁵Phoronix Test Suite [Electronic resource]. – Mode of access: <https://www.phoronix-test-suite.com>. – Date of access: 05.04.2024.

В ПО Phoronix Test Suite производительность измеряется величиной Vogo Ops/s (bogus operations per second), которая является способом измерения скорости исполнения инструкций на компьютере в ядре Линукс.

Журнальные файлы с результатами тестирования находятся в файлах репозитория автора исследования¹⁶:

- data/coral-tpu-benchmark-results.txt для табл. 4;
- data/Thomas-Kaiser-sbc-bench-results.txt для табл. 5;
- data/phoronix-stress-ng-results.txt для табл. 6.

Для наглядности данные из табл. 4–6 показаны в виде диаграмм на рис. 4–6. На рис. 4 чем меньше столбик диаграммы, тем лучше. На рис. 5 и 6 чем выше столбик диаграммы, тем лучше.

Для наглядности на рис. 4 шкала времени представлена в логарифмическом масштабе. Видно, что скорости расчетов на Coral TPU через библиотеку TFLite (TPU + TFLite) отличаются незначительно для различных устройств и значительно опережают расчеты на CPU. Расчеты на CPU для одной и той же модели нейросети MobileNet v3 через библиотеку TensorFlow (CPU + TF) производятся немного быстрее, чем расчеты через библиотеку TensorFlow Lite (CPU + TFLite), кроме ПК Desktop.

Возможно, для ПК Desktop значительное отличие почти в 30 раз между скоростью расчетов CPU + TF (215,5 мс) и CPU + TFLite (6420,3 мс) связано с какими-то внутренними оптимизациями библиотек TF и TFLite.

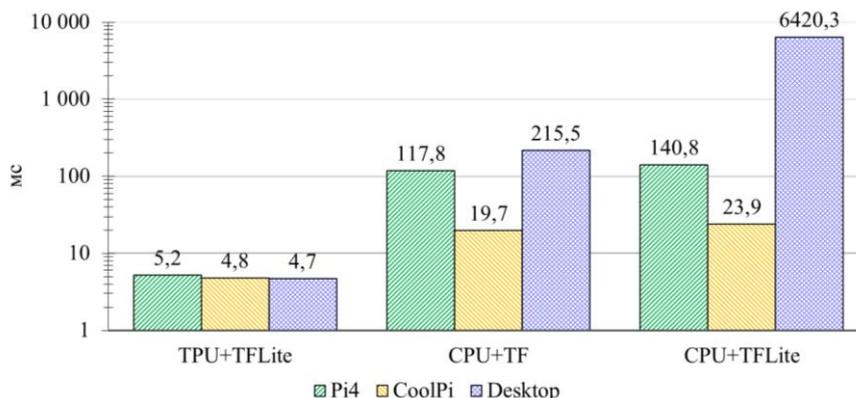


Рис. 4. Тестирование производительности тензорного сопроцессора Coral TPU

Fig. 4. Performance testing of the Coral TPU tensor coprocessor

На рис. 4 видно, что восемь ядер одноплатного компьютера CoolPi хорошо справляются с параллельными нейросетевыми вычислениями, но эти расчеты в четыре раза медленнее, чем аналогичные расчеты на Coral. Вычисления на процессоре Pi4 в 22 раза медленнее, чем аналогичные расчеты на Coral, и в шесть раз медленнее, чем расчеты на CoolPi. Более современные одноплатные компьютеры CoolPi и Pi4 значительно опережают по нейросетевым вычислениям устаревший процессор Intel i7-4770 ПК Desktop.

На рис. 5 и 6 видно, что более быстрый одноплатный компьютер CoolPi опережает Pi4 примерно в два-три раза для задач копирования (memcpy, Memory Copying), сжатия данных (7-zip) и нагрузки на центральный процессор (CPU Stress). CoolPi опережает Pi4 в восемь раз для задачи инициализации (выделения) блока памяти (memset) и в 11 раз для матричных операций (Matrix Math). CoolPi в 30 раз быстрее шифрует данные с помощью алгоритма AES-256, чем Pi4, но этот показатель важен при защищенной передаче данных и не важен для нейросетевых расчетов. Скорость матричных операций восьмиядерного процессора CoolPi сравнима со скоростью восьмиядерного процессора Intel i7-4770 ПК Desktop.

¹⁶URL: https://github.com/foobar167/junkyard/tree/master/coral_tpu

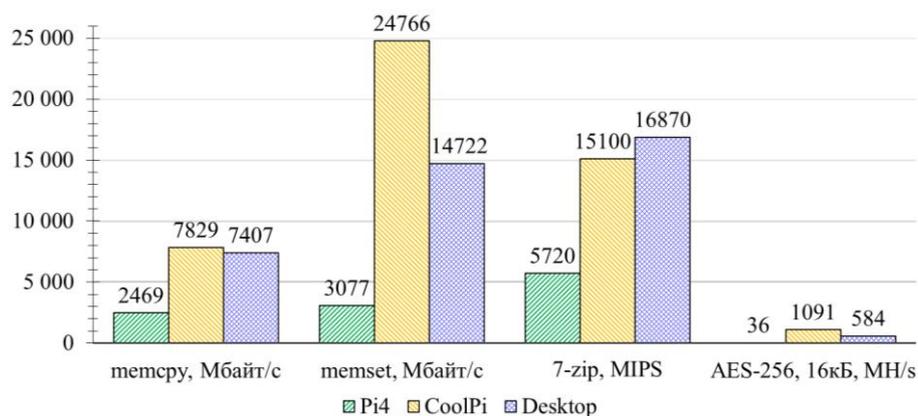


Рис. 5. Тестирование производительности с помощью скрипта sbc-bench.sh Томаса Кайзера
 Fig. 5. Performance testing with the use of Thomas Kaiser's sbc-bench.sh script

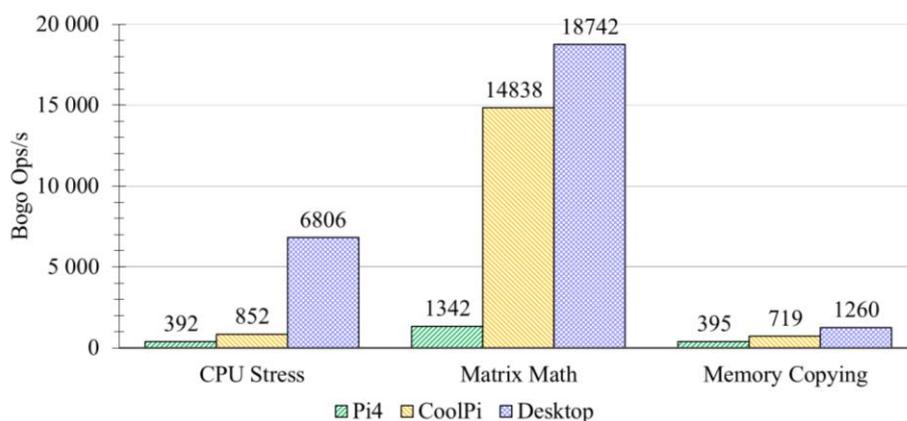


Рис. 6. Тестирование производительности с помощью ПО Phoronix Test Suite, тест stress-ng
 Fig. 6. Performance testing with the Phoronix Test Suite software, stress-ng test

Несмотря на то что Pi4 подходит под требования ТЗ в качестве вычислительной основы, при разработке МВК «БортВК» стоит использовать более мощные альтернативы со встроенным NPU (например, Radxa Rock 5 Model A) либо без встроенного NPU, но с дополнительным ИИ-ускорителем (сочетание CoolPi и Coral TPU). Использование Pi4 с дополнительным ИИ-ускорителем (сочетание Pi4 и Coral TPU) также приемлемо и увеличит скорость вычислений в десятки раз.

Coral дает значительное ускорение нейросетевых расчетов, однако есть особенности, исследование которых будет проводиться на следующих этапах НИР:

- Coral работает только с библиотекой TensorFlow Lite;
- для использования Coral TPU нейронная сеть должна быть предварительно квантована (quantization), когда все веса нейронной сети переводятся из чисел с плавающей запятой (float или double) в целочисленные значения (uint8 или int8);
- если по какой-либо причине не все веса нейронной сети преобразовались из float в uint8, например какой-то слой нейросети не поддерживает квантование, то такая частично квантованная сеть будет выполняться не на Coral TPU, а на CPU;
- тесты производительности Coral в табл. 4 и на рис. 4 выполнены при помощи готовых квантованных нейронных сетей, но, чтобы выполнить задачу адаптации (портирования) нейросети для одноплатного компьютера, необходимо научиться самостоятельно квантовать веса нейронной сети.

Как было показано выше, акселераторы нейросетей обеспечивают самые быстрые нейросетевые вычисления, однако есть нюансы, связанные с новизной технологий: отсутствие подробной документации, малое количество примеров, слабая поддержка пользователя, редкое обновление аппаратного и программного обеспечения и др. Со временем ситуация в области разработки NPU будет улучшаться, а использование дополнительных ИИ-ускорителей упрощаться и лучше документироваться.

Заключение. Обзор доступных на рынке микрокомпьютерных решений, поддерживающих параллельные нейросетевые вычисления, показал, что на данный момент оптимальной вычислительной платформой для разработки МК «БортВК» в соответствии с ТЗ является одноплатный компьютер Cool Pi 4 Model B с подключенным к нему тензорным сопроцессором Google Coral USB Accelerator для ускорения параллельных нейросетевых вычислений. Альтернативой ему служит менее мощный, но более задокументированный и протестированный одноплатный компьютер Raspberry Pi 4 Model B или его аналог Raspberry Pi 5 с подключенным к нему тензорным сопроцессором Google Coral USB Accelerator.

Список использованных источников

1. Сравнительный анализ вычислительных платформ для бортового микро модуля предварительного распознавания изображений / В. А. Ковалев [и др.] // Информатика. – 2018. – Т. 15, № 3. – С. 7–21.
2. Comparative analysis of budget computing platforms for a portable micromodule of on-board image classification / V. A. Kovalev [et al.] // BIG DATA and Advanced Analytics : Collection of Materials of the Fourth Intern. Scientific and Practical Conf., Minsk, Belarus, 3–4 May 2018 / editorial board: M. Batura [et al.]. – Minsk, BSUIR, 2018. – P. 31–42.
3. Распознавание подстилающей поверхности Земли с помощью сверточной нейронной сети на одноплатном микрокомпьютере / Д. А. Павленко [и др.] // Информатика. – 2020. – Т. 17, № 3. – С. 36–43. <https://doi.org/10.37661/1816-0301-2020-17-3-36-43>
4. Интеллектуальная технология распознавания подстилающей поверхности Земли / С. В. Кругликов [и др.] // Радиоэлектронные технологии. – 2019. – № 1. – С. 90–94.
5. Recognition of underlying surface using a convolutional neural network on a single-board computer / D. A. Paulenka [et al.] // BIG DATA and Advanced Analytics : сб. материалов VI Междунар. науч.-практ. конф., Минск, Беларусь, 20–21 мая 2020 г. : в 3 ч. Ч. 1 / редкол.: В. А. Богущ [и др.]. – Минск : Бестпринт, 2020. – С. 71–77.

References

1. Kovalev V. A., Paulenka D. A., Snezhko E. V., Liauchuk V. A., Kalinovski A. A. *Comparative analysis of computing platforms for onboard micromodule of provisional image recognition*. Informatika [Informatics], 2018, vol. 15, no. 3, pp. 7–21 (In Russ.).
2. Kovalev V. A., Paulenka D. A., Snezhko E. V., Liauchuk V. A. Comparative analysis of budget computing platforms for a portable micromodule of on-board image classification. *BIG DATA and Advanced Analytics : Collection of Materials of the Fourth International Scientific and Practical Conference, Minsk, Belarus, 3–4 May 2018*. Editorial board: M. Batura [et al.]. Minsk, Belorusskij gosudarstvennyj universitet informatiki i radiojelektroniki, 2018, pp. 31–42.
3. Paulenka D. A., Kovalev V. A., Snezhko E. V., Liauchuk V. A., Pechkovsky E. I. *Recognition of the Earth's underlying surface using a convolutional neural network on a single-board microcomputer*. Informatika [Informatics], 2020, vol. 17, no. 3, pp. 36–43 (In Russ.). <https://doi.org/10.37661/1816-0301-2020-17-3-36-43>
4. Kruglikov S. V., Kovalev V. A., Paulenka D. A., Snezhko E. V., Liauchuk V. A. *Intelligent technology for recognizing the underlying surface of the Earth*. Radioelektronnye tehnologii [Radioelectronic Technology], 2019, no. 1, pp. 90–94 (In Russ.).
5. Paulenka D. A., Kovalev V. A., Snezhko E. V., Liauchuk V. A., Pechkovsky E. I. *Recognition of underlying surface using a convolutional neural network on a single-board computer*. BIG DATA and Advanced Analytics : sbornik materialov VI Mezhdunarodnoj nauchno-prakticheskoy konferencii, Minsk, Belarus', 20–21 maja 2020 goda : v 3 chastjah. Chast' 1 [BIG DATA and Advanced Analytics : Collection of Materials of the VI International Scientific and Practical Conference, Minsk, Belarus, 20–21 May 2020) : in 3 Parts. Part 1]. Editorial board: V. A. Bogush [et al.]. Minsk, Bestprint, 2020, pp. 71–77.

Информация об авторе

Павленко Дмитрий Анатольевич, ведущий инженер-программист, лаборатория анализа биомедицинских изображений, Объединенный институт проблем информатики Национальной академии наук Беларуси.

E-mail: dmitri.pavlenko@gmail.com

<https://www.researchgate.net/profile/Dzmitry-Paulenka>

<https://scholar.google.com/citations?user=2AX0it0AAAAJ>

<https://orcid.org/0009-0007-9911-4356>

Information about the author

Dzmitry A. Paulenka, Lead Software Engineer, Laboratory of Biomedical Images Analysis, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.

E-mail: dmitri.pavlenko@gmail.com

<https://www.researchgate.net/profile/Dzmitry-Paulenka>

<https://scholar.google.com/citations?user=2AX0it0AAAAJ>

<https://orcid.org/0009-0007-9911-4356>