

ВИКИПЕДИЯ

Нейронный процессор

Материал из Википедии — свободной энциклопедии

Нейро́нный проце́ссор (англ. *Neural Processing Unit, NPU* или **ИИ-ускоритель** англ. *AI accelerator*) — это специализированный класс микропроцессоров и сопроцессоров (часто являющихся специализированной интегральной схемой), используемый для аппаратного ускорения работы алгоритмов искусственных нейронных сетей, компьютерного зрения, распознавания по голосу, машинного обучения и других методов искусственного интеллекта^[1].

Содержание

Описание

История

Области применения

Примеры

Существующие продукты

GPU-продукты

ИИ-ускорители в виде внутренних сопроцессоров (аппаратных ИИ-блоков)

Научные исследования и разрабатываемые продукты

Примечания

Ссылки

Описание

Нейронные процессоры относятся к вычислительной технике и используются для аппаратного ускорения эмуляции работы нейронных сетей и цифровой обработки сигналов в режиме реального времени. Как правило, нейропроцессор содержит регистры, блоки памяти магазинного типа, коммутатор и вычислительное устройство, содержащее матрицу умножения, дешифраторы, триггеры и

мультиплексоры^[2].

На современном этапе (по состоянию на 2017 год) к классу нейронных процессоров могут относиться разные по устройству и специализации типы чипов, например:

- Нейроморфные процессоры — построенные по кластерной асинхронной архитектуре, разработанной в Корнеллском университете (принципиально отличающейся от фон Неймановской и Гарвардской компьютерных архитектур, используемых последние 70 лет в IT-отрасли). В отличие от традиционных вычислительных архитектур, логика нейроморфных процессоров изначально узкоспециализирована для создания и разработки разных видов искусственных нейронных сетей. В устройстве используются обычные транзисторы, из которых строятся вычислительные ядра (каждое ядро, как правило, содержит планировщик заданий, собственную память типа SRAM и маршрутизатор для связи с другими ядрами), каждое из ядер эмулирует работу нескольких сотен нейронов и, таким образом, одна интегральная схема, содержащая несколько тысяч таких ядер, алгоритмически может воссоздать массив из нескольких сотен тысяч нейронов и на порядок больше синапсов. Как правило, такие процессоры применяются для алгоритмов глубокого машинного обучения^[3].
- Тензорные процессоры — устройства, как правило, являющиеся сопроцессорами, управляемыми центральным процессором. Тензорные процессоры оперируют тензорами - объектами при помощи которых удобно выполнять преобразование элементов одного векторного пространства в другое, и которые могут быть представлены как многомерные массивы чисел^[4], обработка которых осуществляется с помощью таких программных библиотек, как, например TensorFlow и Caffe 2. В угоду производительности они, обычно, выполняют операции над числами малой разрядности (8 или 16 бит) и специализированы для быстрого выполнения таких операций, как матричное умножение и свёртка, используемых для эмуляции свёрточных нейронных сетей, которые используются для задач машинного обучения^[5].
- Процессоры машинного зрения — во многом похожи на тензорные процессоры, но они узкоспециализированы для ускорения работы алгоритмов машинного зрения, в которых используются методы свёрточных нейронных сетей (CNN) и масштабно-инвариантная трансформация признаков (SIFT). В них делается большой акцент на распараллеливание потока данных между множеством исполнительных ядер, включая использование модели блокнотной памяти — как в многоядерных цифровых сигнальных процессорах, и они так же, как тензорные процессоры, используются для вычислений с низкой точностью, принятой при обработке изображений^[6].

История

Области применения

- Беспилотный автомобиль — например, в этом направлении развивает свои платы Drive PX-series компания Nvidia^{[7][8]}.
- Беспилотный летательный аппарат — например, навигационная система основанная на чипах Movidius Myriad 2 успешно управляет автономными беспилотными летательными аппаратами^[9].
- Диагностика в здравоохранении.

- Машинный перевод.
- Обработка естественного языка.
- Поисковая система — NPU повышают энергоэффективность центров обработки данных, и дают возможность использовать все более сложные запросы.
- Промышленный робот — NPU позволяют расширить спектр задач, которые возможно автоматизировать, путём добавления приспособляемости к меняющимся ситуациям.
- Распознавание по голосу — например, в мобильных телефонах использование технологии Qualcomm Zeroth^[10]
- Сельскохозяйственный робот — например, борьба с сорняками без применения химических средств^[11].



Nvidia Drive PX-series.

Примеры

Существующие продукты

- Процессоры машинного зрения:
 - Intel Movidius Myriad 2, который является многоядерным ИИ-ускорителем, основанным на VLIW-архитектуре, с дополненными узлами, предназначенными для обработки видео^[6].
 - Mobileye EyeQ — это специализированный процессор, ускоряющий обработку алгоритмов машинного зрения для использования в беспилотном автомобиле^[12].
- Тензорные процессоры:
 - Google TPU (англ. *Tensor Processing Unit*) — представлен как ускоритель для системы Google TensorFlow, которая широко применяется для свёрточных нейронных сетей. Сфокусирован на большом объёме арифметики 8-битной точности^[5].
 - Huawei Ascend 310 / Ascend 910 — первые два чипа оптимизированные под решения задач искусственного интеллекта из линейки Ascend компании Huawei^[13].
 - Intel Nervana NNP (англ. *Neural Network Processor*) — это первый коммерчески доступный тензорный процессор, предназначенный для постройки сетей глубокого обучения^[14], компания Facebook была партнёром в процессе его проектирования^{[15][16]}.
 - Qualcomm Cloud AI 100 — ускоритель искусственного интеллекта, предназначенный для использования в составе облачных платформ, поддерживающий программные библиотеки PyTorch, Glow, TensorFlow, Keras и ONNX^[17].
- Нейроморфные процессоры:



16-ядерный чип Adapteva Eriphany (E16G301) на одноплатном компьютере для параллельных вычислений.

- IBM TrueNorth — нейроморфный процессор, построенный по принципу взаимодействия нейронов, а не традиционной арифметики. Частота импульсов представляет интенсивность сигнала. По состоянию на 2016 год среди исследователей ИИ нет консенсуса, является ли это правильным путём для продвижения^[18], но некоторые результаты являются многообещающими, с продемонстрированной большой экономией энергии для задач машинного зрения^[19].
- Adapteva Epiphany — предназначен как сопроцессор, включает модель блокнотной памяти сети на кристалле, подходит к модели программирования потоком информации, которая должна подходить для многих задач машинного обучения.
- ComBox x64 Movidius PCIe Blade board (https://combox.io/upload/combox_x64_movidius_blade_board.pdf) — плата расширения PCI Express с максимальной плотностью VPU Intel Movidius (MyriadX) для инференса сверхточных нейронных сетей в ЦОД
- Cambricon MLU100 — карта расширения PCI Express с ИИ-процессором мощностью 64 TFLOPS с половинной точностью или 128 TOPS для вычислений INT8^[20].
- Cerebras Wafer Scale Engine (WSE, CS-1) — экспериментальный суперпроцессор компании Cerebras, содержит 1,2 трлн транзисторов, организованных в 400 000 ИИ-оптимизированных вычислительных ядер и 18 Гбайт локальной распределённой памяти SRAM, и всё это связано ячеистой сетью с общей производительностью 100 петабит в секунду. Чип Cerebras — это фактически суперкомпьютер на чипе, где вычислительные ядра SLAC (Sparse Linear Algebra Cores) — полностью программируемые и могут быть оптимизированы для работы с любыми нейронными сетями^[21].
- KnuPath — процессор компании KnuEdge, предназначен для работы в системах распознавания речи и прочих отраслях машинного обучения, он использует соединительную технологию LambdaFabric и позволяет объединять в единую систему до 512 тысяч процессоров^[22].

GPU-продукты

- Nvidia Tesla — серия специализированных GPGPU-продуктов компании Nvidia^[23]:
 - Nvidia Volta — графические процессоры (GPU) архитектуры Volta (2017 год) компании Nvidia (такие как Volta GV100), содержат до 640 специальных ядер для тензорных вычислений^[1].
 - Nvidia Turing — графические процессоры архитектуры Turing (2018 год) компании Nvidia (такие как Nvidia TU104), содержат до 576 специальных ядер для тензорных вычислений^[24].
 - Nvidia DGX-1 — специализированный сервер, состоящий из 2 центральных процессоров и 8 GPU Nvidia Volta GV100 (5120 тензорных ядер), связанных через быструю шину NVLink^[25]. Специализированная архитектура памяти у этой системы является особенно подходящей для построения сетей глубокого обучения^{[26][27]}.



Nvidia Tesla C870.

- AMD Radeon Instinct — специализированная GPGPU-плата компании AMD, предлагаемая как ускоритель для задач глубокого обучения^{[28][29]}.

ИИ-ускорители в виде внутренних сопроцессоров (аппаратных ИИ-блоков)

- Cambricon-1A — NPU-блок в ARM-чипах Huawei Kirin 970, разработанный компанией Cambricon Technologies^[30].
- CEVA NeuPro — семейство лицензируемых ИИ-процессоров для глубокого обучения компании CEVA, Inc.^[31].
- Neural Engine — ИИ-ускоритель внутри ARM-чипов Apple A11 Bionic и A12 Bionic SoC^[32].
- PowerVR 2NX NNA (Neural Network Accelerator) — семейство лицензируемых IP-модулей для машинного обучения компании Imagination Technologies^[33].

Научные исследования и разрабатываемые продукты

- Индийский технологический институт в Мадрасе разрабатывает ускоритель на импульсных нейронах для новых систем архитектуры RISC-V, направленных на обработку больших данных на серверных системах^[34].
- Eyeriss — разработка, направлена на свёрточные нейронные сети с применением блокнотной памяти и сетевой архитектуры в пределах кристалла.
- Fujitsu DLU — многоблочный и многоядерный сопроцессор компании Fujitsu использующий вычисления с низкой точностью и предназначенный для глубокого машинного обучения^[35].
- Intel Loihi — нейроморфный процессор компании Intel, который сочетает процессы обучения, тренировки и принятия решений в одном чипе, позволяя системе быть автономной и «сообразительной» без подключения к облаку. Например, при обучении с помощью базы данных MNIST (Mixed National Institute of Standards and Technology) процессор Loihi оказывается в 1 млн раз лучше, чем другие типичные спайковые нейронные сети^[36].
- Kalray — показала MPPA^[37] и сообщила о повышении эффективности свёрточных нейронных сетей в сравнении с GPU.
- SpiNNaker — массово-параллельная компьютерная архитектура, которая сочетает ядра традиционной ARM-архитектуры с усовершенствованной сетевой структурой, специализированной для моделирования крупной нейронной сети.
- Zeroth NPU — разработка компании Qualcomm, направленная непосредственно на привнесение возможностей распознавания речи и изображений в мобильные устройства^[38].
- IVA TPU — тензорный процессор, над созданием которого работает^{[39][40]} российская компания IVA Technologies. В октябре 2020 года были опубликованы результаты^{[41][42]} тестирования архитектуры ускорителя расчета нейронных сетей IVA TPU, проведенного



6-ядерный SoC Apple A11 Bionic с Neural Engine

международным консорциумом MLPerf (учрежден в 2018 году [Baidu](#), [Google](#), [Harvard University](#), [Stanford University](#), [University of California, Berkeley](#)).

Примечания

1. Популярность машинного обучения влияет на эволюцию архитектуры процессоров (<https://servernews.ru/957816>). Servernews. (31 августа 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117065650/https://servernews.ru/957816>) 17 ноября 2017 года.
2. Нейропроцессор, устройство для вычисления функций насыщения, вычислительное устройство и сумматор (<http://www.findpatent.ru/patent/213/2131145.html>). FindPatent.RU. Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171201030914/http://www.findpatent.ru/patent/213/2131145.html>) 1 декабря 2017 года.
3. IBM поставила LLNL нейропроцессоры TrueNorth за \$1 млн (http://www.computerra.ru/143633/ibm_truenorth_in_llnl/). Компьютерра. (31 марта 2016). Дата обращения: 17 ноября 2017. Архивировано (https://web.archive.org/web/20171119163854/http://www.computerra.ru/143633/ibm_truenorth_in_llnl/) 19 ноября 2017 года.
4. Intel разрабатывает тензорные процессоры для ИИ (<https://www.itweek.ru/ai/article/detail.php?ID=190242>). PC Week/RE. (22 ноября 2016). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171201032547/https://www.itweek.ru/ai/article/detail.php?ID=190242>) 1 декабря 2017 года.
5. Подробности о тензорном сопроцессоре Google TPU (<https://servernews.ru/957543>). Servernews. (25 августа 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117065046/https://servernews.ru/957543>) 17 ноября 2017 года.
6. Intel анонсировала процессор машинного зрения Movidius Myriad X (<https://3dnews.ru/957694>). 3DNews. (29 августа 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117122425/https://3dnews.ru/957694>) 17 ноября 2017 года.
7. Nvidia Drive PX: Scalable AI Supercomputer For Autonomous Driving (<http://www.nvidia.com/object/drive-px.html>). Nvidia. Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20160716182319/http://www.nvidia.com/object/drive-px.html>) 16 июля 2016 года. (англ.)
8. NVIDIA представила Drive PX Pegasus — платформу для автопилота нового поколения (<https://3dnews.ru/959767>). 3DNews (10 октября 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117065100/https://3dnews.ru/959767>) 17 ноября 2017 года. (рус.)
9. Movidius powers worlds most intelligent drone (<https://www.siliconrepublic.com/machines/movidius-dji-drone>). Дата обращения: 15 ноября 2017. Архивировано (<https://web.archive.org/web/20160809224917/https://www.siliconrepublic.com/machines/movidius-dji-drone>) 9 августа 2016 года. (англ.)
10. Qualcomm Research brings server-class machine learning to everyday devices (<https://www.qualcomm.com/news/onq/2015/10/01/qualcomm-research-brings-server-class-machine-learning-everyday-devices-making>). Дата обращения: 15 ноября 2017. Архивировано (<https://web.archive.org/web/20160808154205/https://www.qualcomm.com/news/onq/2015/10/01/qualcomm-research-brings-server-class-machine-learning-everyday-devices-making>) 8 августа 2016 года. (англ.)
11. Design of a machine vision system for weed control (<https://web.archive.org/web/20100623062608/http://www.abe.ufl.edu/wlee/Publications/ICAME96.pdf>). Дата обращения: 15 ноября 2017. Архивировано из оригинала (<http://www.abe.ufl.edu/wlee/publications/icame96.pdf>) 23 июня 2010 года. (англ.)
12. The Evolution of EyeQ (<https://www.mobileye.com/our-technology/evolution-eyeq-chip/>). Дата обращения: 18 ноября 2017. Архивировано (<https://web.archive.org/web/20171207090412/http://www.mobileye.com/our-technology/evolution-eyeq-chip/>) 7 декабря 2017 года.

13. Huawei создала первые в мире ИИ-процессоры, пойдя по пути разработчиков «Эльбрус» (https://www.cnews.ru/news/top/2018-10-23_huawei_sozdala_pervye_v_mire_iiprotssorypojdyu), *CNews* (23 октября 2018). Архивировано (https://web.archive.org/web/20181023144345/http://www.cnews.ru/news/top/2018-10-23_huawei_sozdala_pervye_v_mire_iiprotssorypojdyu) 23 октября 2018 года. Дата обращения: 24 октября 2018.
14. До конца года Intel выпустит «первую в отрасли микросхему для обработки нейронных сетей» — Intel Nervana Neural Network Processor (<http://www.ixbt.com/news/2017/10/18/intel-intel-nervana-neural-network-processor.html>). *iXBT.com* (18 октября 2017). Дата обращения: 21 ноября 2017. Архивировано (<https://web.archive.org/web/20171115020800/http://www.ixbt.com/news/2017/10/18/intel-intel-nervana-neural-network-processor.html>) 15 ноября 2017 года.
15. Intel unveils purpose-built Neural Network Processor for deep learning (<https://techreport.com/news/32704/intel-unveils-purpose-built-neural-network-processor-for-deep-learning>), *Tech Report* (17 октября 2017). Архивировано (<https://web.archive.org/web/20171124103059/http://techreport.com/news/32704/intel-unveils-purpose-built-neural-network-processor-for-deep-learning>) 24 ноября 2017 года. Дата обращения: 17 ноября 2017.
16. Intel Nervana Neural Network Processors (NNP) Redefine AI Silicon (https://www.intelnervana.com/intel-nervana-neural-network-processors-ntp-redefine-ai-silicon/?_ga=2.62312428.1380200850.1508486032-2008757629.1504021982) (17 октября 2017). Архивировано (https://web.archive.org/web/20171020135824/https://www.intelnervana.com/intel-nervana-neural-network-processors-ntp-redefine-ai-silicon/?_ga=2.62312428.1380200850.1508486032-2008757629.1504021982) 20 октября 2017 года. Дата обращения: 17 ноября 2017.
17. Qualcomm представила ускоритель искусственного интеллекта Cloud AI 100 (<https://servernews.ru/985614>), *Servernews.ru* (10 апреля 2019). Архивировано (<https://web.archive.org/web/20190410085332/https://servernews.ru/985614>) 10 апреля 2019 года. Дата обращения: 16 апреля 2019.
18. Ян ЛеКун про IBM TrueNorth (<https://www.facebook.com/yann.lecun/posts/10152184295832143>). Дата обращения: 15 ноября 2017. Архивировано (<https://web.archive.org/web/20150705024137/https://www.facebook.com/yann.lecun/posts/10152184295832143>) 5 июля 2015 года. (англ.)
19. IBM cracks open new era of neuromorphic computing (<http://www.extremetech.com/extreme/187612-ibm-cracks-open-a-new-era-of-computing-with-brain-like-chip-4096-cores-1-million-neurons-5-4-billion-transistors>). — «TrueNorth is incredibly efficient: The chip consumes just 72 milliwatts at max load, which equates to around 400 billion synaptic operations per second per watt — or about 176,000 times more efficient than a modern CPU running the same brain-like workload, or 769 times more efficient than other state-of-the-art neuromorphic approaches». Дата обращения: 15 ноября 2017. Архивировано (<https://web.archive.org/web/20160709120314/http://www.extremetech.com/extreme/187612-ibm-cracks-open-a-new-era-of-computing-with-brain-like-chip-4096-cores-1-million-neurons-5-4-billion-transistors>) 9 июля 2016 года. (англ.)
20. Китайская компания Cambricon разрабатывает чипы ИИ для дата-центров. (<https://web.archive.org/web/20180616030139/http://www.hardwareluxx.ru/index.php/news/hardware/prozessoren/44821-cambricon-ai-chip-.pdf>) Дата обращения: 15 июня 2018. Архивировано из оригинала (<http://www.hardwareluxx.ru/index.php/news/hardware/prozessoren/44821-cambricon-ai-chip-.pdf>) 16 июня 2018 года.
21. Serebras — процессор для ИИ невероятных размеров и возможностей (<https://3dnews.ru/992698>). *3DNews*. (20 августа 2019). Дата обращения: 21 августа 2019. Архивировано (<https://web.archive.org/web/20190820193943/https://3dnews.ru/992698>) 20 августа 2019 года.
22. КнуPath — нейроморфный процессор военного назначения (<https://3dnews.ru/934373>). *3DNews*. (9 июня 2016). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117064831/https://3dnews.ru/934373>) 17 ноября 2017 года.

23. Computex: Глава Nvidia не видит угрозы в «тензорном» процессоре Google (<https://www.osp.ru/news/2016/0601/13032932/>). «Открытые системы». (1 июня 2016). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171201035122/https://www.osp.ru/news/2016/0601/13032932/>) 1 декабря 2017 года.
24. Что принесёт на рынок новая архитектура NVIDIA Turing? (<http://3dnews.ru/973967/>) 3DNews. (14 августа 2018). Дата обращения: 17 августа 2018. Архивировано (<https://web.archive.org/web/20190323115004/https://3dnews.ru/973967/>) 23 марта 2019 года.
25. Эра NVIDIA Volta началась с ускорителя Tesla V100 (<https://servernews.ru/952008/>). Servernews. (11 мая 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171201032027/https://servernews.ru/952008/>) 1 декабря 2017 года.
26. GTC Europe 2017: библиотека NVIDIA TensorRT 3 ускоряет работу нейросетей в 18 раз по сравнению с универсальным решением (<https://servernews.ru/959837/>). Servernews. (12 октября 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117065013/https://servernews.ru/959837/>) 17 ноября 2017 года.
27. Новый российский суперкомпьютер предназначен для обучения нейросетей (<https://servernews.ru/957932/>). Servernews. (1 сентября 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117122728/https://servernews.ru/957932/>) 17 ноября 2017 года.
28. AMD Announces Radeon Instinct: GPU Accelerators for Deep Learning, Coming in 2017 (<http://www.anandtech.com/show/10905/amd-announces-radeon-instinct-deep-learning-2017/>), Anandtech (12 декабря 2016). Архивировано (<https://web.archive.org/web/20161212212552/http://www.anandtech.com/show/10905/amd-announces-radeon-instinct-deep-learning-2017/>) 12 декабря 2016 года. Дата обращения: 12 декабря 2016.
29. Radeon Instinct Machine Learning GPUs include Vega, Preview Performance (<https://www.pcper.com/reviews/Graphics-Cards/Radeon-Instinct-Machine-Learning-GPUs-include-Vega-Preview-Performance/>), PC Per (12 декабря 2016). Архивировано (<https://web.archive.org/web/20170811183801/https://www.pcper.com/reviews/Graphics-Cards/Radeon-Instinct-Machine-Learning-GPUs-include-Vega-Preview-Performance/>) 11 августа 2017 года. Дата обращения: 12 декабря 2016.
30. Huawei представляет будущее мобильного искусственного интеллекта на IFA 2017. (<https://bad-android.com/news/7301-huawei-predstavliaet-bydyshee-mobilnogo-iskysstvennogo-intellekta-na-ifa-2017/>) Дата обращения: 15 июня 2018. Архивировано (<https://web.archive.org/web/20180616030225/https://bad-android.com/news/7301-huawei-predstavliaet-bydyshee-mobilnogo-iskysstvennogo-intellekta-na-ifa-2017/>) 16 июня 2018 года.
31. CEVA NeuPro. A Family of AI Processors for Deep Learning at the Edge. (<https://www.ceva-dsp.com/product/ceva-neupro/>) Дата обращения: 15 июня 2018. Архивировано (<https://web.archive.org/web/20180616030155/https://www.ceva-dsp.com/product/ceva-neupro/>) 16 июня 2018 года.
32. The iPhone X's new neural engine exemplifies Apple's approach to AI (<https://www.theverge.com/2017/9/13/16300464/apple-iphone-x-ai-neural-engine>), *The Verge* (13 сентября 2017). Архивировано (<https://web.archive.org/web/20170915010751/https://www.theverge.com/2017/9/13/16300464/apple-iphone-x-ai-neural-engine>) 15 сентября 2017 года. Дата обращения: 17 ноября 2017.
33. Imagination представила новые ИИ-ускорители PowerVR 2NX (<https://3dnews.ru/970916/>), 3DNews (8 июня 2018). Архивировано (<https://web.archive.org/web/20180616001935/https://3dnews.ru/970916/>) 16 июня 2018 года. Дата обращения: 15 июня 2018.
34. India preps RISC-V Processors - Shakti targets servers, IoT, analytics (https://web.archive.org/web/20170703232121/http://www.eetimes.com/document.asp?doc_id=1328790&page_number=2). — «The Shakti project now includes plans for at least six microprocessor designs as well as associated fabrics and an accelerator chip». Дата обращения: 15 ноября 2017. Архивировано из оригинала (http://www.eetimes.com/document.asp?doc_id=1328790&page_number=2) 3 июля 2017 года. (англ.)
35. Fujitsu разрабатывает специализированный процессор для систем ИИ (<https://servernews.ru/955927/>). Servernews. (24 июля 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117122909/https://servernews.ru/955927/>) 17 ноября 2017 года.

36. Intel представила нейроморфный процессор Loihi (<https://3dnews.ru/959077>). *3DNews*. (26 сентября 2017). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117122017/https://3dnews.ru/959077>) 17 ноября 2017 года.
37. Kalray MPPA (http://www.hotchips.org/wp-content/uploads/hc_archives/hc27/HC27.24-Monday-Epub/HC27.24.20-Multimedia-Epub/HC27.24.240-DeDinechin-KalrayMPPA-Kalray-v4.pdf) . Дата обращения: 15 ноября 2017. Архивировано (https://web.archive.org/web/20160423124257/http://www.hotchips.org/wp-content/uploads/hc_archives/hc27/HC27.24-Monday-Epub/HC27.24.20-Multimedia-Epub/HC27.24.240-DeDinechin-KalrayMPPA-Kalray-v4.pdf) 23 апреля 2016 года. (англ.)
38. Qualcomm показала нейропроцессор Zeroth (<http://logmag.net/qualcomm-pokazala-nejroprocessor-zeroth/>). *Logmag.net* (16 октября 2013). Дата обращения: 17 ноября 2017. Архивировано (<https://web.archive.org/web/20171117174613/http://logmag.net/qualcomm-pokazala-nejroprocessor-zeroth/>) 17 ноября 2017 года.
39. *embedded world*. IVA TPU – DNN inference accelerator // *NeuroMatrix Architecture for Neural Network Applications | embedded world* (<https://www.embedded-world.de/en/events/vortrag/iva-tpu--dnn-inference-accelerator--neuromatrix-architecture-for-neural-network-applications/767250>) (англ.). *www.embedded-world.de*. Дата обращения: 30 ноября 2020. Архивировано (<https://web.archive.org/web/20210121141831/https://www.embedded-world.de/en/events/vortrag/iva-tpu--dnn-inference-accelerator--neuromatrix-architecture-for-neural-network-applications/767250>) 21 января 2021 года.
40. В России создана оригинальная процессорная архитектура, способная потеснить NVidia (https://www.cnews.ru/news/top/2020-11-06_sozdannaya_v_rossii_novejshaya). *CNews.ru*. Дата обращения: 30 ноября 2020. Архивировано (https://web.archive.org/web/20201125080957/https://www.cnews.ru/news/top/2020-11-06_sozdannaya_v_rossii_novejshaya) 25 ноября 2020 года.
41. *Inference Results* (<https://mlperf.org/inference-results/>) (англ.). *MLPerf*. Дата обращения: 30 ноября 2020. Архивировано (<https://web.archive.org/web/20201128082119/https://mlperf.org/inference-results>) 28 ноября 2020 года.
42. *Sally Ward-Foxton*. Machine learning benchmark expands support for edge, data center workloads (<https://www.embedded.com/machine-learning-benchmark-expands-support-for-edge-data-center-workloads/>) (амер. англ.). *Embedded.com* (3 ноября 2020). Дата обращения: 30 ноября 2020. Архивировано (<https://web.archive.org/web/20201125113744/https://www.embedded.com/machine-learning-benchmark-expands-support-for-edge-data-center-workloads/>) 25 ноября 2020 года.

Ссылки

- Популярность машинного обучения влияет на эволюцию архитектуры процессоров (<https://servernews.ru/957816>). *Servernews*. (31 августа 2017). Дата обращения: 17 ноября 2017.

Источник — https://ru.wikipedia.org/w/index.php?title=Нейронный_процессор&oldid=133203392

Эта страница в последний раз была отредактирована 22 сентября 2023 в 18:43.

Текст доступен по лицензии Creative Commons «С указанием авторства — С сохранением условий» (CC BY-SA); в отдельных случаях могут действовать дополнительные условия.

