



Умная система видеоаналитики на основе встраиваемого ИИ

Дмитрий Швецов

В статье рассматривается архитектура интеллектуальной системы видеонаблюдения на базе встраиваемого искусственного интеллекта, ориентированной на обнаружение, подсчёт и отслеживание людей в реальном времени. Предложена распределённая конвейерная архитектура обработки видеопотоков с использованием нейронной сети MobileNet-SSD и специализированных аппаратных ускорителей VPU. Описаны аппаратно-программные компоненты системы, методы предварительной и постобработки данных, а также алгоритмы отслеживания объектов на основе фильтров Калмана. Приведён практический пример реализации в промышленной системе контроля опасных зон и обсуждены преимущества пограничных вычислений, включая низкую задержку, энергоэффективность и повышение конфиденциальности данных. Результаты показывают возможность обработки нескольких видеопотоков в реальном времени при низком энергопотреблении, что подтверждает перспективность встраиваемых платформ для интеллектуального видеонаблюдения и систем безопасности.

Введение

В настоящее время глубокое обучение систем с ИИ продемонстрировало значительные преимущества в различных областях исследований, таких как финансы, медицина, автоматическая классификация объектов и многих других. В частности, компьютерное зрение стало первой областью, в которой глубокое обучение получило широкое применение. Новые методы обработки на основе глубокого обучения заменяют традиционные алгоритмы компьютерного зрения, основанные на физических представлениях, моделях и функциях объектов распознавания. Более совершенные системы способны обрабатывать огромные объёмы данных в крупных вычислительных центрах. Текущие задачи связаны с тренировкой систем машинного обучения на достаточном количестве информации, что требует разметки этих данных. С другой стороны, в данной статье внимание концентрируется на узлах умных камер, распределённых по зоне наблюдения, а значит, требующих автономности.

Совершенствование технологий как на программном, так и на аппаратном

уровне позволило проектировать интеллектуальные системы, способные не только управлять видеопотоками с замкнутой цепи камер, но и анализировать и извлекать информацию из видеопотоков в реальном времени.

Эти встраиваемые системы можно использовать как в общественных, так и в частных пространствах, позволяя контролировать количество людей в зоне, движение толпы, обнаруживать аномальное поведение и т.п. В большинстве такие системы являются централизованными, выполняющими алгоритмы компьютерного зрения в одном ЦОДе. Эта центральная система обработки получает и обрабатывает информацию со всей сети камер. Старые системы видеонаблюдения (мониторинга) не обрабатывали информацию и требовали анализа оператора, который был обязан внимательно следить за каждой камерой, и чья эффективность анализа могла снижаться из-за усталости и монотонности.

Современные централизованные системы обработки видео хранят и обрабатывают видеoinформацию, полученную из сети камер. Только некоторые предупреждения или видеофраг-

менты отображаются на центральной консоли/оператору, не требуя высокого уровня внимания к управлению системой наблюдения.

С другой стороны, появление Интернета вещей и концепции вычислений на граничных узлах привело к появлению множества решений, интегрированных в распределённых системах видеонаблюдения. Таким образом, интеллект системы распределён между несколькими узлами, каждый из которых может включать камеру и систему обработки, выполняющую простые задачи перед отправкой информации оператору, облегчая его работу.

Для более сложных задач современные алгоритмы обнаружения систем компьютерного зрения включают нейронные сети с глубоким машинным обучением (DNN). Для выполнения DNN обычно требуются высокопроизводительные аппаратные системы с большой вычислительной мощностью с графическими процессорами. Помимо высокой стоимости, эти аппаратные модули имеют высокое энергопотребление, и их может быть сложно встроить в распределённые узлы умных камер.

В этом контексте в статье рассматривается интеллектуальная система видеонаблюдения для обнаружения, подсчёта и отслеживания людей в реальном времени на встраиваемой аппаратной системе с использованием новых аппаратных модулей – процессоров обработки изображений. Эта система может быть легко установлена и настроена для работы в качестве узлов с умными камерами в распределённой системе видеонаблюдения.

В качестве примера подобной системы для задачи отслеживания и обнаружения людей рассмотрим архитектуру MobileNet-SSD, оценим производительность системы, сравнив её с другими алгоритмами, оценим потребление энергии граничного узла для конвейерной архитектуры, способной обрабатывать до 12 видеопотоков одновременно.

Обзор технологий встраиваемого ИИ

Для оценки выполнения задач ИИ на встраиваемых устройствах необходимо проанализировать возможности производительности устройств с ограниченными ресурсами в реальном времени.

На рынке существует множество устройств, позволяющих внедрять искусственный интеллект «на борту».

Наиболее известным сопроцессором ускорения является графический процессор (GPU). Большинство GPU предназначены для использования в составе настольных компьютеров, но в последние годы появились различные встраиваемые комплекты разработки, включающие GPU, такие как Jetson AGX Xavier или Jetson Nano, в основном ориентированные на автономные системы.

Другой тип технологии аппаратного ускорения – семейство Coral от Google. Это набор устройств, которые можно подключить через порт USB и создать систему ускорения для архитектур, ориентированных на тензорные вычисления.

Новейшей технологией для разработки систем видеонаблюдения является Vision Processing Unit (VPU), разработанная Intel и ориентированная на параллельную обработку нейронных сетей. VPU ориентированы на высокоскоростную обработку вывода с очень низким энергопотреблением, что позволяет внедрять такие устройства во встраиваемые системы.

Встраиваемый ИИ: аппаратные платформы и тенденции

Термин «встраиваемый искусственный интеллект» охватывает широкий спектр аппаратных и программных решений, предназначенных для выполнения задач машинного обучения непосредственно на устройстве, без необходимости постоянной связи с облаком или центральным сервером.

Ключевые преимущества встраиваемого ИИ:

- 1) низкая задержка: решения принимаются локально, что критично для систем реального времени (наблюдение, автономные транспортные средства);
- 2) конфиденциальность данных: видеопотоки и другие чувствительные данные не покидают устройство;
- 3) надёжность: работа продолжается при отсутствии или ухудшении сетевого соединения.
- 4) энергоэффективность: специализированные процессоры оптимизированы для выполнения операций нейронных сетей с минимальным энергопотреблением.

Основные классы аппаратных ускорителей для встраиваемого ИИ

1. *Vision Processing Units (VPU)*: специализированные процессоры, такие как Intel Movidius Myriad X, спроектированные для высокоэффективного выполнения операций вывода нейронных сетей. Их архитектура включает множество маломощных вычислительных ядер (SHAVE), оптимизированных для потоковой обработки данных компьютерного зрения. Потребляемая мощность часто составляет менее 5 Вт, что делает их идеальными для автономных устройств.
2. *Tensor Processing Units (TPU)*: разработанные Google, эти ASIC (специализированные интегральные схемы) оптимизированы исключительно для матричных умножений, лежащих в основе нейронных сетей. Edge TPU, как в устройствах Coral, предлагают высокую производительность при очень низком энергопотреблении для задач вывода.
3. *Графические процессоры (GPU) для встраиваемых систем*: платформы, такие как NVIDIA Jetson (Nano, AGX Xavier), объединяют мощные GPU архитектуры CUDA с процессорами ARM в форм-факторе, подходящем для встраиваемых и робототехнических

применений. Они обеспечивают большую гибкость для разработки и выполнения сложных моделей, но, как правило, имеют более высокое энергопотребление, чем VPU/TPU.

4. *Нейроморфные процессоры*: инновационные чипы, такие как Intel Loihi, имитируют структуру и работу биологического мозга (спайковые нейронные сети). Они обещают на порядки более высокую энергоэффективность для определённых классов задач, особенно связанных с обработкой сенсорных потоков данных в реальном времени, и являются перспективным направлением для встраиваемого ИИ будущего.

В статье рассмотрим платформу UpSquared2 с VPU Myriad X, что является компромиссом между высокой производительностью в задачах компьютерного зрения, крайне низким энергопотреблением и относительно низкой стоимостью, что соответствует требованиям массового развёртывания в системах видеонаблюдения.

Обнаружение людей

Обнаружение объектов определяется как способность распознавать и локализовать определённый тип объекта на изображении. В этом контексте литература, связанная с обнаружением людей в приложениях видеонаблюдения, очень обширна. Среди различных предложенных методов можно выделить две большие группы: основанные на классических методах, которые используют математические техники и моделирование особенностей людей, и современные подходы, основанные на глубоком обучении.

Принадлежащий к классическим методам распознавания образов на изображении алгоритм, определяющий особенности изображения с использованием гистограммы ориентированных градиентов, а затем применяющий классификатор на основе метода опорных векторов, является одним из самых популярных. В алгоритме предполагается использование признаков для обнаружения головы и плеч людей в сочетании с классификатором на основе анализа главных компонент.

В последние годы увеличилось использование камер RGBD, которые предоставляют как RGB-изображения, так и информацию о глубине с использованием различных технологий, таких как стереоскопическое зрение или время пролёта (время, необходимое свету

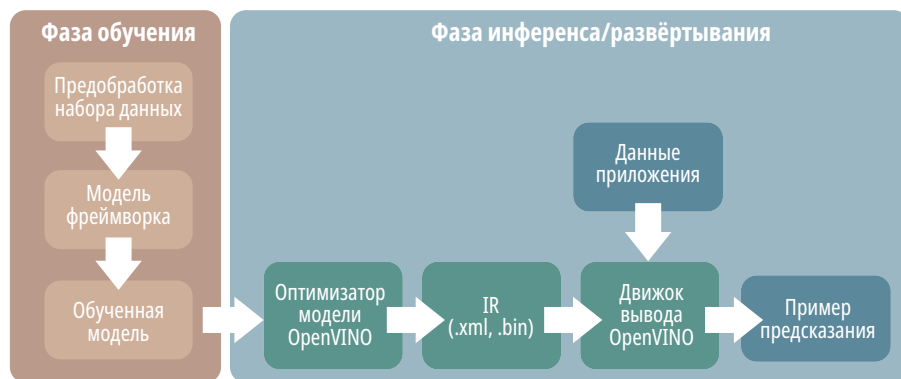


Рис. 1. Архитектура OpenVINO™

для прохождения до объекта и обратно к камере). Улучшение мощности систем обработки привело к смене парадигмы в компьютерном зрении. В частности, классификация изображений и распознавание объектов перешли на использование нейронных сетей. Одной из первых таких сетей была так называемая AlexNet в 2012 году. Кроме того, этот тип сети можно использовать для обнаружения конкретных элементов, таких как пешеходы.

Этот тип сети использует наборы свёрточных слоёв, извлекающих данные, которые используются плотной нейронной сетью для классификации изображения. Для обучения этих сетей используются размеченные наборы данных, что позволяет получить классификаторы с высокой способностью к обобщению и значениями точности, далеко выходящими за рамки тех, что дают классические методы. Для правильного обучения эти наборы данных должны быть большого объёма, а в случае их недостаточности существуют методы, позволяющие улучшить обучение с меньшим количеством изображений, решая проблему переобучения. Несмотря на хорошие результаты в точности, необходимо учитывать высокую вычислительную стоимость при работе с этим типом сети, что делает её дорогой и сложной для работы в реальном времени.

Для улучшения работы в реальном времени и ускорения вычислений было предложено несколько математических моделей. Они обычно называются алгоритмами «один выстрел» из-за их способности локализовать область с наиболее релевантной информацией на изображении. Среди этих алгоритмов наиболее известны Faster R-CNN, SSD и YOLO. Эти алгоритмы могут обрабатывать изображения в реальном времени на устройствах с ограниченными ресурсами. Рассмотрим алгоритм для обнаружения людей SSD. По сравнению с классическими

методами, он медленнее в обучении, но имеет лучшие результаты как по времени, так и по точности в выводе.

Встраиваемая система обработки: аппаратные и программные компоненты

Поскольку одним из требований к предлагаемой системе является её портативность и гибкость для развёртывания, выбор аппаратной встраиваемой платформы является одной из основных задач, решаемых в разработке. Одной из целей была разработка приложения ИИ для встраиваемой системы, которое могло бы обнаруживать, отслеживать и подсчитывать людей. Предложенная архитектура была основана на распараллеливании нескольких процессов для наиболее эффективного использования доступных аппаратных модулей. Анализ был разделён на два процесса, которые общаются через использование независимых буферов.

Для тестирования алгоритмов искусственного интеллекта на VPU наиболее часто используется платформа OpenVINO™, упрощающая оптимизацию и развёртывание свёрточных нейронных сетей (CNN). Платформа OpenVINO™ включает два различных инструмента, показанных на рис. 1: оптимизатор моделей (Model Optimizer) и движок вывода (Inference Engine). Эти инструменты позволяют оптимизировать модель и снизить время выполнения, используя для этого различные аппаратные платформы, например процессоры Intel (CPU), ускорители нейросетей (VPU) или графические процессоры (GPU). Одно из преимуществ этой платформы заключается в том, что её можно установить на любое устройство, отвечающее минимальным требованиям. Это также позволяет устанавливать оба модуля отдельно: оптимизатор моделей – на компьютере, используемом для обучения сети, а движок вывода – во встраиваемой системе.

Оптимизатор моделей представляет собой консольное приложение, которое позволяет адаптировать и оптимизировать нейронные модели для ускорения вывода. Оптимизатор моделей OpenVINO™ поддерживает различные фреймворки, такие как TensorFlow, PyTorch или Caffe. В результате его работы создаются файлы промежуточного представления: файл с расширением .xml, описывающий слои, размеры и соединения архитектуры, и файл с расширением .bin, содержащий веса параметров этой архитектуры. Что касается движка вывода, он может быть установлен на любом устройстве независимо от оптимизатора моделей. Движок вывода загружает IR-файлы и выполняет вывод на выбранном пользователем оборудовании, будь то CPU, VPU или GPU. Кроме того, движок вывода отвечает за балансировку нагрузки при выводе, чтобы не перегружать отдельные устройства. Таким образом, вычислительная нагрузка распределяется либо между ядрами CPU, либо между несколькими VPU.

Обработка изображений с применением ИИ для обнаружения и отслеживания людей на встраиваемых системах

Основной идеей разработки приложения на основе ИИ является создание инструментария, способного обнаруживать, отслеживать и подсчитывать людей во встраиваемой системе. На рис. 2 представлена архитектура решения, основанная на параллелизации нескольких процессов с целью максимально эффективного использования доступных аппаратных модулей. Анализ был разделён на два взаимосвязанных процесса, обменивающихся данными через промежуточные буферы.

В первом процессе – так называемом процессе аналитики данных – выполнялись предварительная обработка изображения и постобработка информации, возвращённой движком вывода ИИ. До предварительной обработки к изображению могли применяться различные алгоритмы: снижение шума, детектор границ или другие методы низкоуровневой обработки пикселей для улучшения качества изображения. Однако, поскольку такая низкоуровневая предобработка зависит от конкретного сценария применения и привнесит дополнительные вычислительные затраты, она была сведена к минимуму: выбор области интереса (ROI), 300 пикселей и нормализация значений пиксе-

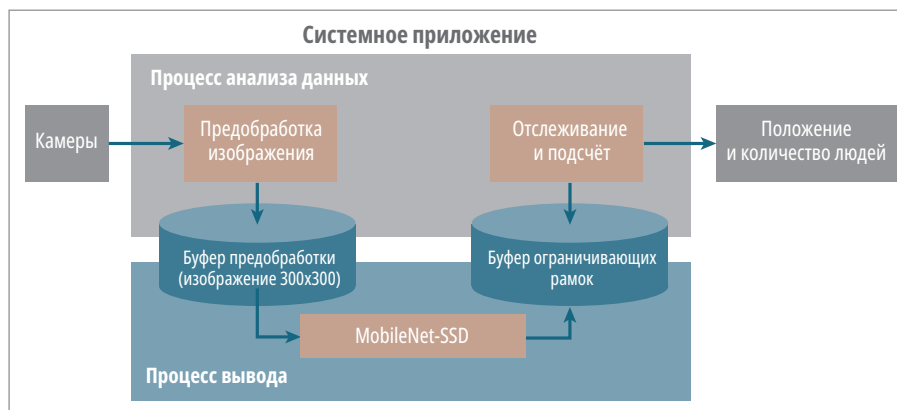


Рис. 2. Поток данных и многопоточная обработка

лей в диапазон $(-1,0, +1,0)$ – всё это делалось для соответствия требованиям входного слоя сети MobileNet-SSD. Постобработка включает реорганизацию данных, полученных от сети MobileNet-SSD, которые содержат ограничивающие рамки, предсказывающие расположение людей. Эта информация далее анализируется банком фильтров Калмана, прогнозирующим движения и возможные перекрытия объектов, что позволяет получить более точный и надёжный результат. Весь первый процесс полностью выполняется на центральном процессоре (CPU) системы.

Второй процесс – это непосредственно выполнение вывода нейронной сети. Он предназначен для работы на границе (edge) с использованием VPU, хотя также допускается его выполнение на CPU. После того как MobileNet-SSD завершает вывод, ограничивающие рамки сохраняются в общий буфер. В соответствии с архитектурным решением, представленным на рис. 2, использование таких компонентов, как VPU, позволяет одновременно выполнять более одного вывода, обрабатывая несколько видеопотоков параллельно. При использовании CPU количество одновременных выводов определяется вычислительной мощностью процессора и числом его ядер.

Обнаружение людей с использованием MobileNet-SSD

Для разработки была выбрана архитектура MobileNet, которая использует метод SSD для обнаружения объектов. Причин использования этой архитектуры две: во-первых, требовалась быстрая архитектура, которую можно было бы реализовать с выбранными алгоритмами, такими как SSD; и, во-вторых, была критична требовательность к ресурсам, потому что устройства, используемые в этом проекте, являются портативными, и их оборудование не такое

мощное, как у высокопроизводительного ПК. По этой причине была выбрана архитектура MobileNet, основной особенностью которой является скорость вычислений и использование типа свёрточных слоёв, позволяющих использовать меньше ресурсов. Эта архитектура снижает интенсивность обработки данных благодаря использованию разделимых поэлементных и точечных свёрточных слоёв.

Отслеживание людей с использованием банка фильтров Калмана

Для отслеживания людей в данной системе обнаружения использовались фильтры Калмана. Основные причины использования этого типа фильтров: устойчивость к перекрытиям и низкая вычислительная стоимость фильтра. В рассматриваемой системе предполагается, что движение людей будет происходить с постоянной и линейной скоростью. Кривые зависимости предсказания состояния и обновления состояния показаны на рис. 3, а числовые данные – в табл. 1. Эти зависимости отслеживания одного человека с помощью фильтра Калмана представлены на фото рис. 3 в виде рамки, выделяющей объект. Таким образом, для каждого обнаружения, выполненного на MobileNet-SSD, фильтр Калмана будет отслеживать каждый объект в виде ограничивающих рамок обнаруженных объектов.

Для распараллеливания каждого из элементов, составляющих систему, ис-

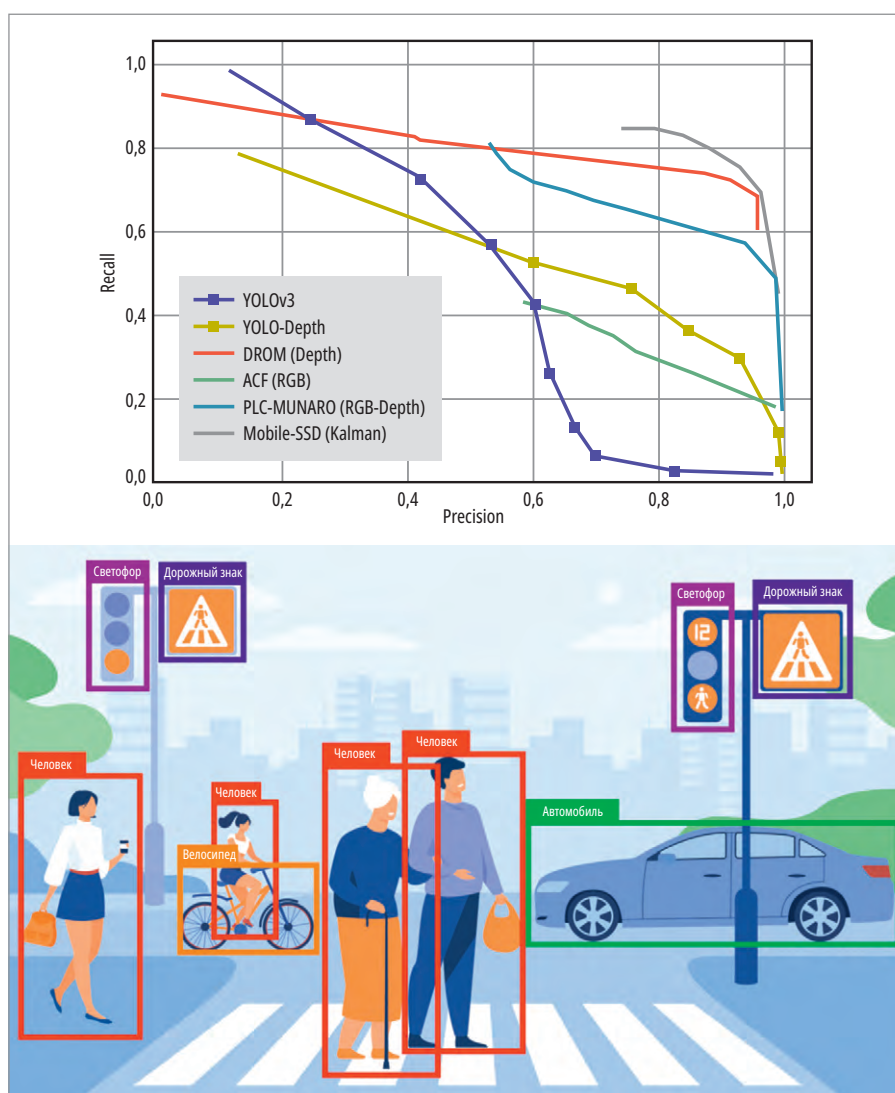


Рис. 3. Полнота и точность распознавания объектов различных нейросетей

Таблица 1. Полнота и точность набора данных EPFL

Методы	EPFL-объект		EPFL-стенд	
	Точность	Полнота	Точность	Полнота
MobileNet-SSD+Kalman	87,82	88,14	81,3	80,6
ACF	83,8	86,4	66,3	40,3
DPOM	98,5	85,4	96,3	70,9
PCL-MUNARO	88,61	82,36	92	56
YOLOv3	89,7	90,3	58,6	59,8
YOLO-depth	89,8	88,6	78,4	47,9

пользуется конвейерная работа. Основными элементами, составляющими архитектуру, являются предварительная обработка, фильтр Калмана и вывод MobileNet-SSD. Распараллеливание вычислительных процессов позволяет обрабатывать более одного видеопотока одновременно. Эту функцию ранее выполняли с помощью очередей CPU, теперь это выполняется с помощью VPU, которая может обрабатывать более одного вывода одновременно, а также более оптимально, чем через CPU. Это позволяет не только обрабатывать больше видеопотоков, но и более эффективно использовать аппаратные средства устройства вычислительной платформы.

В качестве наглядного примера успешной реализации принципов встраиваемого ИИ в специализированном оборудовании можно рассмотреть российскую разработку – бортовой комплекс контроля опасных зон «ВИЗИОН». Данное решение демонстрирует, как современные технологии компьютерного зрения и локального вывода нейронных сетей воплощаются в компактных автономных устройствах для решения конкретной прикладной задачи – повышение безопасности при эксплуатации транспорта и техники в ограниченных пространствах.

Ключевые особенности и архитектура решения системы

«ВИЗИОН» представляет собой законченный встраиваемый комплекс, построенный на рассмотренных ранее принципах.

Её архитектура включает в себя:

- блок управления с модулем ИИ, выполняющий функции, аналогичные описанному выше конвейеру данных. На этом блоке, оснащённом специализированным аппаратным ускорителем (например, VPU), работает сконфигурированная нейронная сеть для детекции людей;

- набор камер (от 1 до 6 шт.), формирующих видеопотоки для анализа. Система поддерживает суммарный угол обзора до 360° и может работать в условиях средней освещённости, при необходимости используя инфракрасные камеры;
- локальный коммуникационный блок и модуль светозвуковой сигнализации для оповещения;
- автономный или сетевой источник питания с широким диапазоном входного напряжения питания.

В данном случае реализация принципов встраиваемого ИИ в системе «ВИЗИОН» на примере локального и полностью автономного комплекса, который не требует облачных вычислений или постоянного сетевого подключения. Все операции – от захвата изображения до анализа сцены и принятия решения – выполняются на борту. Это обеспечивает минимальную задержку, критически важную для систем безопасности, и независимость от качества каналов связи. Для выполнения задач компьютерного зрения в реальном времени используются энергоэффективные аппаратные ускорители, такие как VPU или аналоги. Это позволяет системе иметь низкое энергопотребление при высокой вычислительной производительности, делая возможным питание от аккумуляторов или использование в удалённых локациях. Следует отметить, что видеопотоки с камер обрабатываются локально и не передаются в облако или внешние центры обработки данных. Это гарантирует сохранность информации и соответствует строгим требованиям безопасности на промышленных объектах. Как упоминалось ранее, в основе этой системы лежит нейронная сеть, оптимизированная для детекции человека. Сеть анализирует данные в реальном времени, прогнозируя возникновение опасных ситуаций. Как и в рассмотренной архитектуре на базе MobileNet-SSD, сеть использует опти-

мизированные свёрточные слои для баланса между точностью и скоростью работы на ограниченных ресурсах. Важной особенностью является возможность дополнительного «дообучения» сети под индивидуальные требования заказчика для распознавания специфических объектов или сценариев (предметы, животные).

Отличительной особенностью описанного примера является большее количество зон контроля, их размеры и пороги срабатывания сигнализации (световой индикатор на расстоянии 5-3 метра, звуковой – на 3 метра и ближе). Это позволяет адаптировать систему к работе как в крупных цехах, так и в тесных помещениях или проходах. Таким образом, эта система служит отличным примером того, как теоретические и аппаратно-программные разработки в области встраиваемого ИИ находят практическое применение. Она решает актуальную проблему безопасности, сочетая высокую точность обнаружения (до 99,9%), работу в реальном времени, энергоэффективность и соответствие требованиям к защите данных, демонстрируя все ключевые преимущества пограничных вычислений в сфере промышленного видеонаблюдения и автоматизации.

Заключение

В статье был рассмотрен метод создания систем с ИИ на встраиваемой платформе, которая может обнаруживать и отслеживать людей надёжным и точным способом. Новая компьютерная технология, используемая на встраиваемых платформах, – это аппаратные модули VPU, которые позволяют выполнять вывод CNN быстрее и эффективнее, чем CPU, в устройствах начального уровня. Разработанная система позволяет реализовать приложение компьютерного зрения с низким энергопотреблением и высокой вычислительной производительностью.

Другим важным моментом системы является обработка нескольких видеопотоков в реальном времени с использованием интегрированного VPU. Возможность автономного питания системы от портативных батарей или возобновляемых систем обеспечивает большую гибкость в распределённых системах видеоналиктики. ●

Автор – сотрудник фирмы ПРОСОФТ
Телефон: (495) 234-0636
E-mail: info@prosoft.ru