

Разработка аппаратного обеспечения систем искусственного интеллекта для интеграции нейросетевых моделей в промышленные робототехнические системы

А. Д. Клиновицкий, А. В. Тимофеев, А. Ю. Травин, А. М. Табунщик,
З. Д. Вчерашний, А. А. Акжигитов

Санкт-Петербургский политехнический университет Петра Великого

adk160501@gmail.com

Аннотация. Техническое зрение, в частности использование нейронных сетей для детектирования объектов и их классификации на изображениях, стало ключевым элементом современных мобильных робототехнических систем и конвейерных производств. Эти технологии применяются для позиционирования, анализа данных и построения адаптивных моделей управления. В статье представлено решение на базе микрокомпьютерного модуля (SOM), объединяющее функции обработки видеопотоков, сбора данных с датчиков через низкоуровневые интерфейсы (UART, I2C, CAN, RS485) и выполнения нейросетевых моделей в реальном времени. Разработанный программируемый логический контроллер (ПЛК) исключает промежуточные узлы предобработки, снижая задержки и повышая надежность за счет интеграции промышленных стандартов (виброустойчивость, пассивное охлаждение). Реализация поддерживает фреймворки TensorFlow Lite, PyTorch Mobile и оптимизирована под GPU Mali-G610 MP4. Тестирование на модели YOLOv11 подтвердило возможность обработки видео в реальном времени при сохранении функциональности промышленных интерфейсов. Результаты демонстрируют потенциал предложенного подхода для создания автономных робототехнических систем, адаптированных к условиям динамических производственных сред.

Ключевые слова: техническое зрение; нейронные сети; микрокомпьютер; детектирование объектов; система-на-модуле (SOM); программируемый логический контроллер; промышленные интерфейсы связи; ПЛК; аппаратное ускорение

I. ВВЕДЕНИЕ

Современные промышленные системы, интегрирующие алгоритмы технического зрения и нейронные сети, сталкиваются с архитектурной фрагментацией, обусловленной разделением функциональных модулей. Как правило, такие системы включают:

1. специализированный блок обработки изображений, ответственный за задачи детектирования и классификации объектов;
2. контроллер для сбора данных с датчиков через интерфейсы UART, I2C, CAN;
3. промежуточные узлы предобработки сигналов и преобразования интерфейсов.

Такая архитектура влечет за собой ряд недостатков: увеличение задержек передачи данных, рост стоимости системы из-за дополнительных компонентов, а также снижение надежности вследствие усложнения топологии.

В данной статье предложено аппаратно-интегрированное решение на базе модуля System-on-Module (SOM), объединяющее функции обработки видеопотоков, исполнения нейросетевых моделей и прямого управления исполнительными механизмами. Ключевая инновация заключается в устранении промежуточных узлов предобработки и интерфейсных преобразователей, что обеспечивает:

1. снижение задержек за счет прямой передачи данных от датчиков к нейросетевому процессору;
2. повышение надежности системы, достигаемое за счет минимизации количества компонентов, что снижает потенциальные точки отказа, так как вероятность системного сбоя пропорционально уменьшается при сокращении числа элементов в цепи.
3. соответствие стандартам для промышленных контроллеров.

II. ОБЗОР ЛИТЕРАТУРЫ

Интеграция нейросетевых моделей в промышленные системы активно исследуется в контексте задач автоматизации, где идут разработки нейросетевых регуляторов на существующие ПЛК [1–3]. Работы [4–5] демонстрируют эффективность YOLO и ResNet для детектирования объектов на производственных линиях. Исследования [6–7] подчеркивают важность аппаратного ускорения (например, с использованием TensorRT).

Современные SOM-решения, такие как NVIDIA Jetson [8], предлагают интеграцию вычислений и интерфейсов, но не всегда соответствуют требованиям к надежности, виброустойчивости, наличию промышленных интерфейсов и температурному режиму. Данное исследование заполняет пробел, предлагая модуль, сочетающий нейроускорение, промышленные стандарты и пассивное охлаждение.

III. РАЗРАБОТКА ПЛК

А. Разработка аппаратной платформы

Первостепенной задачей проектирования стало формирование технического задания, учитывающего требования к производительности, надежности и совместимости с промышленными стандартами. Основные критерии включали:

1. Поддержку аппаратного ускорения нейросетевых моделей, например, через встроенный нейронный процессор (NPU) для задач ИИ-инференса, включая обработку изображений, классификацию объектов и выполнение алгоритмов машинного обучения на базе распространенных фреймворков.
2. Интеграцию интерфейсов (UART, I2C, CAN, RS485).
3. Соответствие стандартам виброустойчивости (до 5g), температурного диапазона (-20°C до +60°C) и защиты от внешних воздействий.

В качестве базового вычислительного модуля выбран SOM Firefly iCore-3588Q, сочетающий четырехъядерные процессоры ARM Cortex-A76 (до 2.6 ГГц) и Cortex-A55 (1.8 ГГц) с графическим ускорителем Mali-G610 MP4. Данная платформа обеспечивает 6 ТОПС для INT8-инференции, что подтверждено ее применением в автономных бортовых системах, БПЛА и медицинских комплексах с ИИ [9].

В. Реализация промышленных интерфейсов

Среди промышленных интерфейсов реализованы:

- 1) CAN:
 - 2 изолированных канала;
 - Напряжение пробоя изоляции: 2 кВ;
 - Режимы: внешнее/внутреннее питание (переключение через тумблер).
- 2) RS485:
 - 2 канала с поддержкой Modbus и скоростью 8 Мбит/с;
 - встроенный шунт 120 Ом;
 - изоляция до 1.5 кВ.
- 3) Ethernet:
 - 4 порта со скоростью 1000 Мбит/с и связью с контроллером по GMAC;
 - протоколы: Modbus TCP, весь стек IP (включая TCP, UDP).

С. Проектирование интерфейсов и печатной платы

Для обработки видеопотоков интегрирован маршрутизатор со связью с контроллером по GMAC и 4 портами Ethernet (до 1000 Мбит/с), обеспечивающий передачу данных в режиме реального времени. В структуру интерфейсов включены:

- 8 программируемых GPIO-портов с возможностью конфигурации под UART/I2C;
- 2 силовых ключа (24 В, 5 А) для управления внешней нагрузкой;

- шестипиновый разъем для подключения дополнительных модулей.

Дифференциальные пары линии Ethernet и USB2.0 имеют характер сильного взаимодействия, рассчитаны при идеальных условиях. Печатная плата спроектирована в четырехслойной топологии на базе стеклотекстолита FR4 ($\epsilon = 4.5$, $\tan\delta = 0.02$) с медным опорным слоем. Параметры дифференциальных пар (Ethernet, USB 2.0) рассчитаны в KiCad:

- ширина дорожек: 0.15 мм, расстояние между парами: 0.1 мм;
- толщина металлизации: 18 мкм, диэлектрик: 115 мкм.

Результирующий импеданс составил 100.9 Ом (при вариациях $\epsilon=4-5$ и толщине диэлектрика до 200 мкм - до 120 Ом), что соответствует требованиям высокоскоростных протоколов. В дальнейшем результаты показали скорость у разработанного встроенного маршрутизатора до 1000 Мбит/с.

Д. Разработка электрических и функциональных характеристик

Для обеспечения универсальности и совместимости с промышленными стандартами входное напряжение устройства было разделено на два диапазона: VIN12 (9–15 В) и VIN48 (36–60 В). Выбор диапазонов обусловлен распространенностью данных напряжений в промышленной автоматизации (например, 12 В для датчиков, 48 В для силовых приводов). Для стабилизации и защиты от перегрузок внедрены:

DC/DC-преобразователи с КПД $\geq 92\%$, обеспечивающие регулировку напряжения.

TVS-диоды и самовосстанавливающиеся предохранители для подавления импульсных помех и коротких замыканий.

Чтение ШИМ до 80 кГц обеспечено аппаратными захватами сигналов и DMA-передачами для минимизации задержек;

Аналоговые входы (4x 16-бит АЦП) оснащены многоуровневыми RC-фильтрами и программной калибровкой, что снизило уровень шума до 2 LSB.

Е. Проектирование программной экосистемы

Программный стек разрабатывался с учетом гибкости и производительности. Критичные к времени задачи (например, обработка данных сенсоров в реальном времени) реализованы через обертки на Cython, связывающие Python-код с Linux-драйверами (через ioctl, sysfs или libgpiod).

Интеграция нейроускорения потребовала настройки Vulkan API под Mali-G610 MP4 через специализированные слои (layers) для распределения вычислительных задач между шейдерными блоками и тензорными ядрами, а также адаптации моделей YOLOv11 под TensorFlow Lite с применением квантования INT8, что сократило потребление памяти.

Ф. Обеспечение эксплуатационных параметров

Температурная стабильность достигнута благодаря пассивному радиатору из анодированного алюминия.

Расчет системы охлаждения проводился с помощью инструмента ПО SolidWorks flow simulation, при тепловыделении вычислительным модулем 15 Вт тепловой энергии. Теплопроводящим элементов является полимерная термопрокладка толщиной 0,5 мм с теплопроводностью 5 Вт/м*К. Графическое изображение результатов расчета представлено на рис. 1. По результатам расчета максимальная температура исследуемой системы составляет 66,68 °С. Данная температура превышает максимально разрешенную производителем рабочую температуру в 60 °С, следовательно при программной настройке системы необходимо ограничить энергопотребление вычислительного модуля.

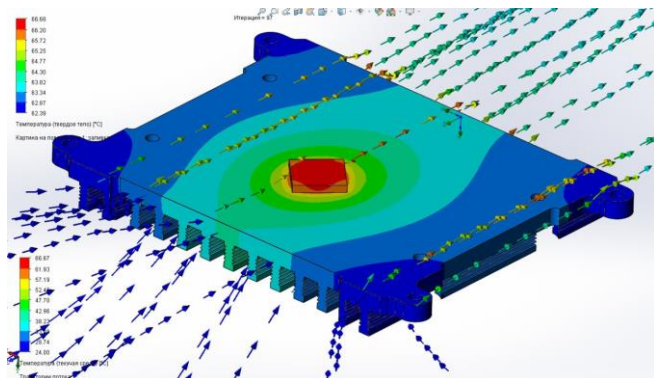


Рис. 1. Расчет теплораспределения при 15 Вт потребления

Виброустойчивость достигнута благодаря системе виброразвязок, поглощающих вибрации.

Расчет виброустойчивости проводился через уравнение гармонического анализа при помощи инструмента Ansys, а именно Harmonic Response и Transient Analysis. Были проведены два расчета - с учетом переходных колебаний в начале возбуждения, физической и геометрической нелинейности, а также без этого учета. Для гашения вибраций в корпусе предусматривались демпферы из TPU GF-10. Решение по выбору материала выносилось по полученным результатам. Далее были проведены тесты с установкой, показанной на рис. 2, создающей постоянные колебания частотой 4 Гц.

Внутренние компоненты залиты термостойким компаундом для защиты от механических деформаций.



Рис. 2. Установка для тестирования виброустойчивости

IV. РЕЗУЛЬТАТЫ

В результате разработаны печатные платы, для интерфейсов (рис. 3) и управляющего модуля ПЛК (рис. 4).

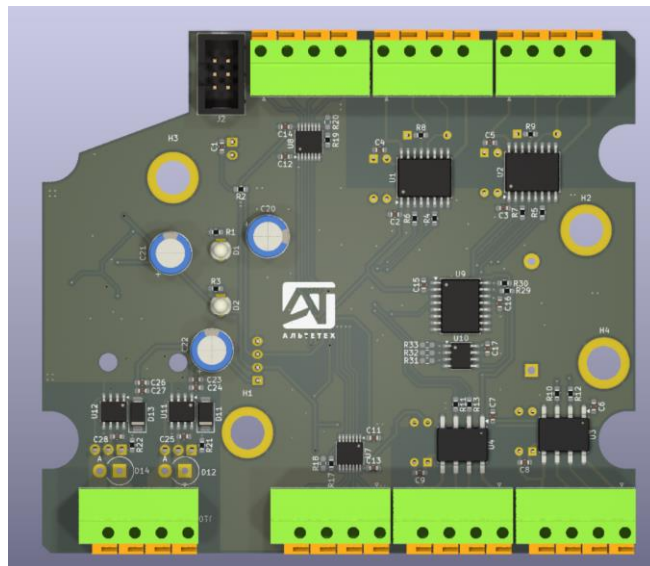


Рис. 3. Плата-расширения интерфейсов

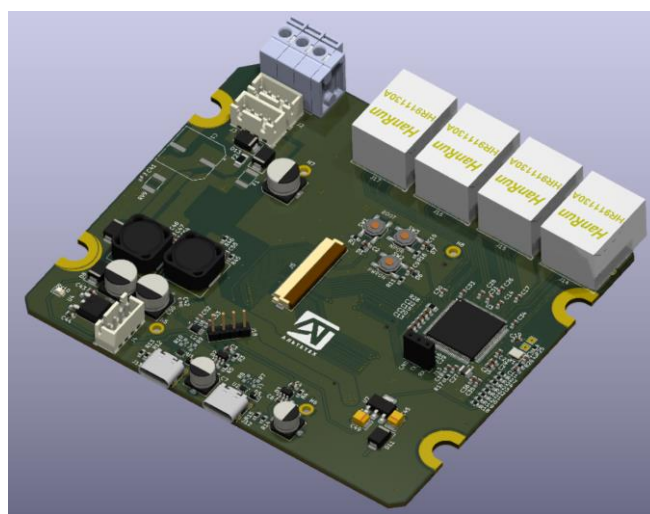


Рис. 4. Плата управляющего модуля

Для интеграции в промышленную инфраструктуру разработан корпус модуля с оптимизированной геометрией, с возможностью выполнения его методом литья. Конструкция предусматривает:

- пассивный радиатор с ребристой поверхностью;
- унифицированный крепеж по стандарту IEC 60715 (DIN-рейка 35 мм), обеспечивающий совместимость с типовыми промышленными шкафами управления (рис. 5).

Данный конструктив обеспечивает компактный монтаж (габариты 155×110×80 мм) и соответствие требованиям Industry 4.0 для развертывания в распределенных системах автоматизации, а также мобильных робототехнических системах, где требуется использование нейронных систем. Изготовленный прототип, удовлетворяющий заявленным требованиям, показан на рис. 6.

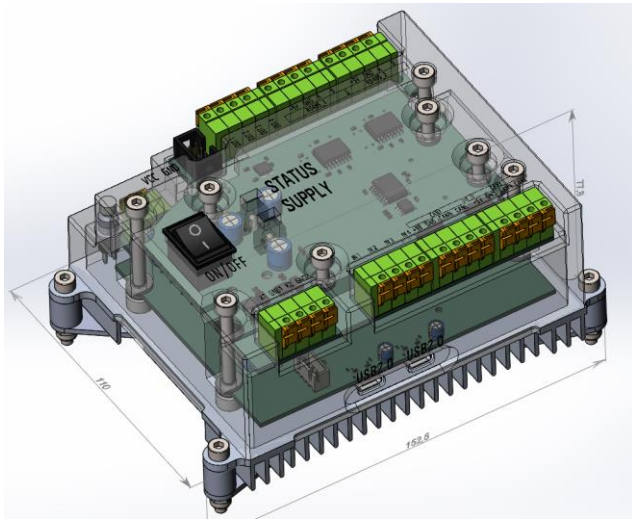


Рис. 5. Разработанная модель



Рис. 6. Прототипы ПЛК

V. ЗАКЛЮЧЕНИЕ

Разработанный SOM-модуль демонстрирует эффективность интеграции нейросетевых моделей в промышленные системы. Устранение промежуточных узлов позволило снизить задержки, повысить надежность и упростить масштабирование. Решение поддерживает современные фреймворки ИИ и соответствует промышленным стандартам.

VI. ДАЛЬНЕЙШИЕ ПЕРСПЕКТИВЫ

Дальнейшие перспективы развития ПЛК включают в себя:

- Внедрение 5G*GSM-модема для удаленного управления и передачи данных.
- Интеграция моделей предиктивной аналитики для обслуживания оборудования.
- Расширение поддерживаемых интерфейсов (например, PROFIBUS, Промышленный Ethernet, RS422).
- Оптимизация энергопотребления для портативных робототехнических систем.

СПИСОК ЛИТЕРАТУРЫ

- [1] Толстель О.В., Ширкин А.Е., Калабин А.Л. Построение системы технического зрения для выравнивания содержимого упаковок дельта-манипулятором на пищевом производстве // Программные продукты и системы. 2023. Т. 36. № 2. С. 334-341.
- [2] Volkov A.K., Mironova L.V., Potapova S.E. The use of pretrained neural networks for solving the problem of reverse searching of X-ray images of prohibited items and substances // May 2024 Civil Aviation High TECHNOLOGIES 27(2):8-24. DOI:10.26467/2079-0619-2024-27-2-8-24
- [3] Мокрецов Н.С., Татарникова Т.М. Оптимизация процесса обучения при ограниченном объеме вычислительных ресурсов. // SCM-25, с.205-208. Available at: <https://scm.etu.ru/assets/files/2024/sbornik-novyj/205-208.pdf> (accessed 23.05.2025)
- [4] Степаненко С.О., Якимов П.Ю. Использование высокопроизводительной платформы глубокого обучения для ускорения обнаружения объектов //Сборник трудов ИТНТ-2019. 2019. С. 624-630.
- [5] Kortli Y. et al. Deep embedded hybrid CNN–LSTM network for lane detection on NVIDIA Jetson Xavier NX // Knowledge-based systems. 2022. Т. 240. С. 107941.
- [6] Ismagilov R. Performance Evaluation of the Rockchip Systems-on-Chip Through YOLOv4 Object Detection Model //2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT). IEEE, 2023. P. 241-243.
- [7] Савиных А.А., Астахов А.М. Прочностной и термодинамический расчет теплообменника // ББК 1 Н 34. С. 5247.
- [8] Тестоведов Н.А. и др. Контроль зазоров в конструкциях технических изделий в процессе вибрационных испытаний // Обработка металлов: технология, оборудование, инструменты. 2021. Т. 23. №. 2. С. 40-53.
- [9] Ibrahim M.Q. et al. Optimizing Convolutional Neural Networks: A Comprehensive Review of Hyperparameter Tuning Through Metaheuristic Algorithms // Archives of Computational Methods in Engineering. 2025. С. 1-38.