

Аспирант 2 курса кафедры информатизации образования
Московский городской педагогический университет (МГПУ)

(г. Москва, Россия)

АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ОБРАБОТКИ МУЛЬТИМЕДИЙНЫХ ДАННЫХ ДЛЯ РОБОТИЗИРОВАННЫХ ЛИНИЙ ПРОИЗВОДСТВА

Аннотация: В статье представлена архитектура интеллектуальной системы обработки мультимедийных данных для роботизированных линий производства. Описана последовательная организация модулей сенсорного ввода, предобработки, семантической сегментации, аналитики и принятия решений. Показаны требования к вычислительной инфраструктуре, методы оптимизации обработки в реальном времени и особенности интеграции системы в промышленную среду. Предлагаемая архитектура обеспечивает устойчивость к шумам, повышение точности анализа и адаптивность к изменяющимся условиям производства.

Ключевые слова: семантическая сегментация, интеллектуальные системы, мультимедийные данные, роботизированные линии производства, компьютерное зрение, архитектура систем, анализ изображений, обработка данных в реальном времени, промышленная автоматизация.

ВВЕДЕНИЕ

Современные роботизированные линии производства являются ключевым элементом концепции «Индустрия 4.0», предполагающей глубокую интеграцию методов искусственного интеллекта и средств компьютерного зрения в автоматизированные технологические процессы. Обработка мультимедийных данных — изображений, видеопотоков, данных глубины и трёхмерных моделей — становится основным источником информации для принятия решений в реальном времени, обеспечивая контроль качества, мониторинг состояния оборудования, навигацию и взаимодействие роботов с объектами производственной среды. Особую роль в этом процессе играют архитектуры интеллектуальных систем, определяющие последовательность и принципы взаимодействия модулей сбора, обработки, анализа и интерпретации данных.

Промышленные мультимедийные данные характеризуются высоким уровнем вариабельности и зашумлённости вследствие нестабильного освещения, вибраций, пыли, различий в фактуре и ориентации объектов, а также наличия движущихся механизмов и персонала. Формально поток данных можно представить как последовательность наблюдений:

$$x_t \in R^{H \times W \times C}, t = 1, \dots, T, \quad (1)$$

где x_t – кадр видеопотока (или многоканальное изображение), H, W – высота и ширина кадра, C – число каналов (например, RGB или RGBD). На каждом шаге времени система должна построить отображение:

$$f_{\theta}: x_t \rightarrow y_t, \quad (2)$$

где y_t – семантическая карта сцены либо иная форма структурированной информации, необходимой для принятия решений. Однако точность отдельного алгоритма f_{θ} сама по себе недостаточна: эффективность системы определяется архитектурой всего конвейера обработки данных:

$$F = F_{ctrl} \circ F_{an} \circ F_{seg} \circ F_{pre} \circ F_{sens}, \quad (3)$$

где

F_{ctrl} – модуль принятия решений и выработки управляющих воздействий;

F_{an} – модуль интеллектуального анализа данных;

F_{seg} – модуль семантической сегментации;

F_{pre} – модуль предварительной обработки и фильтрации шумов;

F_{sens} – модуль сенсорного ввода и синхронизации данных.

Существующие архитектуры обработки изображений включают классические CNN-подходы¹, архитектуры семантической сегментации FCN² и

¹ LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. — 1998. — Vol. 86, No. 11. — P. 2278–2324.

² Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2015. — P. 3431–3440.

U-Net³, а также современные модели на базе Vision Transformers⁴. Однако их прямое применение в условиях производства ограничено высокой вычислительной стоимостью, снижением устойчивости к шумам и отсутствием механизмов адаптации к изменяющимся условиям.

В связи с этим возникает необходимость разработки архитектуры интеллектуальной системы, способной объединить методы семантической сегментации, фильтрации шума, детектирования аномалий и адаптивного обучения в единый программно-аппаратный комплекс, обеспечивающий надёжность, масштабируемость и работу в режиме близком к реальному времени. Такая архитектура должна учитывать специфику промышленной среды, требования по точности и отказоустойчивости, возможности распределённых вычислений и ограниченность ресурсов edge-платформ.

Целью настоящей статьи является разработка и описание архитектуры интеллектуальной системы обработки мультимедийных данных для роботизированных линий производства, обеспечивающей формирование устойчивого, согласованного и адаптивного конвейера обработки информации на основе современных методов компьютерного зрения и глубокого обучения. Представленное решение конкретизирует положения диссертационного исследования и демонстрирует архитектурный подход, необходимый для практического внедрения интеллектуальных систем в промышленную среду.

ОБЗОР СУЩЕСТВУЮЩИХ АРХИТЕКТУР ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Современные архитектуры интеллектуальных систем обработки мультимедийных данных формировались на стыке методов компьютерного

³ Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). — 2015. — P. 234–241.

⁴ Dosovitskiy A., Beyer L., Kolesnikov A., et al. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale // Proceedings of the International Conference on Learning Representations (ICLR). — 2021.

зрения, глубинного обучения и инженерии сложных программно-аппаратных комплексов. Наиболее значимые архитектурные решения можно условно разделить на три крупных направления: классические конвейеры обработки изображений, архитектуры на основе свёрточных нейронных сетей, а также модели нового поколения на базе механизмов self-attention и трансформеров.

Классические конвейеры обработки изображений, используемые ещё до распространения глубоких нейронных сетей, включали в себя цепочки модулей выделения признаков, фильтрации, бинаризации и последующей классификации объектов. Такие системы обладали высокой интерпретируемостью, но ограниченной универсальностью, особенно в условиях сложных производственных сцен, отличающихся высокой динамикой и непредсказуемостью. Эти подходы быстро уступили место глубоким моделям, обеспечившим качественный скачок в точности распознавания.

Фундамент современной архитектуры компьютерного зрения был заложен свёрточными нейронными сетями (CNN). Одной из первых успешных архитектур глубокого обучения для анализа изображений стала модель LeNet-5, которая продемонстрировала эффективность многоуровневой свёрточной обработки признаков. Позднее появились более универсальные сети, такие как AlexNet⁵, VGG и ResNet, однако они были ориентированы преимущественно на классификацию и не решали задач пиксельного анализа.

Ключевым архитектурным скачком стала работа Fully Convolutional Networks (FCN) — первой архитектуры, позволяющей преобразовывать изображение в семантическую карту. FCN заложила основу сегментационных моделей: отказ от полносвязных слоёв, использование транспонированных свёрток и прямую возможность работы с изображениями произвольного размера. Практическое значение FCN для промышленных систем заключается в возможности построения полноразмерных карт объектов без их

⁵ Krizhevsky, A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G. E. Hinton // Communications of the ACM. — 2017. — Vol. 60, № 6. — P. 84–90.

предварительного выделения, что делает архитектуру особенно ценной для роботизированных линий.

Существенный вклад в развитие архитектур внесла модель U-Net, изначально разработанная для биомедицинской сегментации. Её принцип encoder–decoder с симметричными skip-connections обеспечил высокую точность локализации объектов при одновременной устойчивости к шумам и частичному исчезновению признаков. Благодаря своей простоте и эффективности U-Net стала фактическим стандартом в задачах сегментации, включая промышленные сценарии.

Новым направлением в архитектурном развитии стали модели на основе трансформеров. Vision Transformers (ViT) продемонстрировали, что механизмы self-attention способны эффективно заменять свёрточные операции, обеспечивая высокий уровень глобального контекстного анализа. Для производственных систем важным преимуществом ViT является способность обрабатывать сцены с большой вариативностью расположения объектов, что особенно актуально для динамических роботизированных линий. Позднее появились более специализированные трансформерные архитектуры для сегментации, такие как SegFormer и Mask2Former, объединяющие преимущества CNN и self-attention.

Однако общим недостатком большинства существующих архитектур является их изолированность: они представляют собой отдельные алгоритмы, которые должны быть интегрированы в более крупную систему. В производственной среде этого недостаточно — требуется комплексная архитектура, объединяющая сбор, фильтрацию, сегментацию, анализ аномалий и принятие решений в реальном времени. Поэтому дальнейшее развитие направлено на переход от отдельных моделей к комплексным программно-аппаратным архитектурам, обеспечивающим стабильность, адаптивность и способность работать в условиях высокой вариативности данных.

ПРЕДЛАГАЕМАЯ АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ

Предлагаемая архитектура интеллектуальной системы обработки мультимедийных данных для роботизированных линий производства основывается на модульном принципе, обеспечивающем устойчивость к шумам, возможность масштабирования и интеграцию различных моделей компьютерного зрения. Архитектура строится как конвейер нескольких последовательно работающих модулей, каждый из которых выполняет уникальную функцию в процессе преобразования данных. В формальном виде систему используя композицию (3).

Такой «поточный» подход соответствует архитектурным принципам современных систем обработки изображений и сигналов⁶ и обеспечивает разделение ответственности между уровнями, что облегчает масштабирование системы и адаптацию к новым условиям производства.

Модуль сенсорного ввода

Сенсорный слой формирует поток данных (1), в котором используются RGB-камеры, стереокамеры, камеры глубины, высокоскоростные промышленные модули. На уровне архитектуры предусмотрены буферизация кадров, синхронизация сенсоров и первичная фильтрация некорректных данных, что является критичным для сценариев реального времени.

Модуль предварительной обработки

Предварительная обработка направлена на подавление промышленных шумов и артефактов (блики, пыль, дрожание камеры). В архитектуру включаются операции нормализации, фильтрации (Gaussian, Median, Bilateral), коррекции освещённости, стабилизации изображения и выравнивания перспективы.

⁶ Szeliski R. Computer Vision: Algorithms and Applications. — Springer, 2022. — 1152 p.

Предобработка выполняется либо на GPU-сервере, либо на edge-устройстве, что снижает нагрузку на последующие модули и уменьшает задержку отклика системы.

Модуль семантической сегментации

Ключевым компонентом архитектуры является модуль семантической сегментации, преобразующий исходное изображение в карту классов:

$$F_{seg}(x_t) \rightarrow y_t, \quad (4)$$

где $y_t \in R^{H \times W \times C}$ – карта сегментации с K классами объектов.

В зависимости от требований к точности и скорости могут использоваться различные типы моделей:

- Облегчённые архитектуры для работы в реальном времени: MobileNet-DeepLab, Fast-SCNN, SegFormer-B0;
- высокоточные архитектуры: U-Net, FCN, DeepLabV3+.

SegFormer, как модель нового поколения, объединяет преимущества трансформерных механизмов с высокой вычислительной эффективностью, что делает её подходящей для роботизированных систем реального времени.

Модуль интеллектуального анализа

Аналитический модуль отвечает за интерпретацию карты сегментации и выделение значимых объектов и событий. Основные функции:

- определение геометрических параметров объектов;
- отслеживание положения и траекторий;
- сопоставление параметров изделий с эталонами производственного контроля;
- обнаружение аномалий в виде дефектов, отклонений или сбоев в процессе⁷.

⁷ Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. — 2009. — Vol. 41, No. 3. — P. 1–58.

Этот модуль может использовать как классические ML-модели, так и механизмы глубокого обучения, включая сети для классификации, ранжирования или анализа временных последовательностей.

Модуль принятия решений

Финальным элементом является модуль управления, формирующий управляющее действие:

$$a_t = F_{ctrl}(y_t, z_t), \quad (5)$$

где z_t – дополнительные данные от производственного оборудования (скорость конвейера, состояние манипуляторов и т.д.). Архитектура предусматривает работу в жёстких временных рамках $t_{proc} \leq 50$ мс, что соответствует требованиям большинства промышленных циклов. Управляющие воздействия включают корректировку траекторий, сортировку изделий, автоматическую отбраковку или остановку линии при нарушениях.

РЕАЛИЗАЦИЯ АРХИТЕКТУРЫ В УСЛОВИЯХ ПРОМЫШЛЕННОГО ПРОИЗВОДСТВА

Реализация предложенной архитектуры в условиях реального производства требует выбора оптимальной программно-аппаратной платформы, обеспечивающей как высокую производительность обработки мультимедийных данных, так и устойчивость работы в условиях промышленной среды. На практике применяются гибридные конфигурации, сочетающие edge-вычисления (NVIDIA Jetson, Intel Movidius), локальные GPU-сервера и промышленные контроллеры, взаимодействующие по протоколам OPC UA или EtherCAT⁸. Такое распределение нагрузки позволяет выполнять предварительную обработку и часть сегментационных операций локально на периферийных устройствах, а ресурсоёмкие этапы анализа или дообучения моделей — на центральных вычислительных узлах.

⁸ OPC Foundation. OPC Unified Architecture Specification. — Version 1.05. — OPC Foundation, 2020.

Применение edge-вычислений снижает задержку передачи данных, минимизирует зависимость от сетевой инфраструктуры и обеспечивает устойчивость системы к возможным сбоям связи. Для критически важных операций (например, отбраковка изделий или контроль безопасности) реализуется механизм локального аварийного контура, позволяющий выполнять минимально необходимую аналитику без обращения к внешним вычислительным ресурсам.

Передача данных и коммуникационная инфраструктура

Одним из ключевых элементов промышленной реализации является организация надёжной коммуникационной подсистемы. Для передачи мультимедийных данных высокой частоты применяются интерфейсы GigE Vision, USB3 Vision и Camera Link HS. На уровне логики взаимодействия модулей используется модель обмена сообщениями publisher–subscriber и протоколы с гарантированной доставкой.

Для интеграции системы в существующую производственную инфраструктуру применяются стандарты промышленного обмена:

- *OPC UA* – для обмена параметрами оборудования;
- *MQTT* — для передачи асинхронных событий;
- *REST/gRPC* — для взаимодействия с серверными компонентами.

Использование формальных протоколов исключает необходимость разработки специфичных интерфейсов и облегчает масштабирование.

Оптимизация вычислений

Промышленные задачи требуют обработки данных в режиме, близком к реальному времени. В связи с этим архитектура включает ряд оптимизационных механизмов:

1. *Квантизация и компрессия моделей* - снижение разрядности весов нейронных сетей (FP32 → FP16 → INT8) позволяет увеличить скорость вычислений при минимальной потере точности⁹.
2. *Пайплайнинг и параллелизм* — выполнение операций предобработки, сегментации и аналитики в отдельных потоках снижает общую задержку конвейера.
3. *Оптимизация памяти* — использование zero-copy буферов и pinned memory уменьшает накладные расходы на передачу данных между CPU и GPU.
4. *Профилирование и автоматическая настройка параметров* — позволяет адаптировать частоту кадров, глубину сегментации и набор фильтров к текущей загрузке системы.

Эффективность оптимизаций особенно важна при использовании облегчённых edge-платформ.

Интеграция в роботизированные линии

Интеграция интеллектуальной системы в роботизированную линию требует соблюдения трёх ключевых условий:

- согласование частоты обработки с тактом производственного цикла;
- устойчивость к вариациям внешних условий (смена освещения, вибрации, изменение скорости конвейера);
- гарантированное время реакции.

Для большинства производственных сценариев критически важно, чтобы полный цикл обработки, включая семантическую сегментацию и анализ, выполнялся в пределах $t_{proc} \leq 50 - 70$ мс, что обеспечивает работу в реальном времени даже при вариациях сложности сцены.

В результате интеграции система позволяет выполнять контроль качества изделий, автоматизировать сортировку, обеспечивать безопасность человеко-машинного взаимодействия и корректировать траектории манипуляторов в

⁹ Jacob B., Kligys S., Chen B., et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2018. — P. 2704–2713.

зависимости от расположения объектов. Опыт промышленных внедрений⁸ показывает, что подобные системы способны снизить количество дефектов на 25–40 %, повысить эффективность линий и уменьшить время реагирования оператора.

Заключение

В работе представлена архитектура интеллектуальной системы обработки мультимедийных данных, ориентированная на применение в роботизированных производственных линиях. Предложенный конвейер объединяет модули сенсорного ввода, предобработки, семантической сегментации, аналитики и принятия решений, обеспечивая устойчивую работу в условиях шумов и ограниченного вычислительного бюджета. Формальная модель системы позволяет описать последовательность обработки данных и определить ключевые требования к времени отклика, надежности и распределению вычислений между периферийными и серверными компонентами.

Особое внимание уделено вопросам практической реализации архитектуры: выбору аппаратных платформ, организации передачи данных, методам оптимизации вычислений и интеграции с промышленными коммуникационными протоколами. Показано, что применение современных моделей компьютерного зрения в сочетании с edge-вычислениями и механизмами адаптации позволяет улучшить качество контроля, снизить количество производственных ошибок и повысить эффективность взаимодействия роботизированных систем с динамичной производственной средой.

Представленная архитектура служит универсальной основой для дальнейшего развития промышленных систем компьютерного зрения. Перспективы включают внедрение методов непрерывного обучения на производстве, применение более эффективных трансформерных архитектур, расширение использования мультимодальных данных и интеграцию с другими компонентами киберфизических систем.

СПИСОК ЛИТЕРАТУРЫ

1. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // *Proceedings of the IEEE*. — 1998. — Vol. 86, No. 11. — P. 2278–2324.
2. Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. — 2015. — P. 3431–3440.
3. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. — 2015. — P. 234–241.
4. Dosovitskiy A., Beyer L., Kolesnikov A., et al. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale // *Proceedings of the International Conference on Learning Representations (ICLR)*. — 2021.
5. Krizhevsky, A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G. E. Hinton // *Communications of the ACM*. — 2017. — Vol. 60, № 6. — P. 84–90.
6. Szeliski R. *Computer Vision: Algorithms and Applications*. — Springer, 2022. — 1152 p.
7. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // *ACM Computing Surveys*. — 2009. — Vol. 41, No. 3. — P. 1–58.
8. OPC Foundation. *OPC Unified Architecture Specification*. — Version 1.05. — OPC Foundation, 2020.
9. Jacob B., Kligys S., Chen B., et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. — 2018. — P. 2704–2713.
10. Xie E., Wang W., Yu Z., et al. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers // *Advances in Neural Information Processing Systems (NeurIPS)*. — 2021.