

МЕТОД ДЕТЕКТИРОВАНИЯ И КЛАССИФИКАЦИИ АНОМАЛИЙ В ПРОИЗВОДСТВЕННЫХ ДАННЫХ НА ОСНОВЕ ГИБРИДНЫХ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ

A HYBRID DEEP LEARNING METHOD FOR ANOMALY DETECTION IN VISUAL DATA OF ROBOTIC PRODUCTION SYSTEMS

V. Dyomin
O. Romashkova

Summary. The paper addresses the problem of anomaly detection in visual data of robotic production systems. A hybrid method combining convolutional networks, a transformer module, and reconstruction-based analysis is proposed. The approach enables robust detection of both local and global defects and operates without anomaly labels. Experiments demonstrate improved accuracy and robustness to domain shifts compared to existing methods, confirming its applicability to industrial quality control.

Keywords: anomaly detection, machine vision, robotic systems, hybrid models, convolutional networks, transformers; reconstruction, quality control.

Дёмин Владислав Дмитриевич

Аспирант, ГАОУ ВО Московский городской
педагогический университет
neyronich@sofa-invest.ru

Ромашкова Оксана Николаевна

Доктор технических наук, профессор,
Российская академия народного хозяйства
и государственной службы при Президенте РФ, г. Москва
ox-rom@yandex.ru

Аннотация. В работе рассматривается задача детектирования аномалий в визуальных данных роботизированных производственных систем. Предложен гибридный метод, объединяющий сверточные сети, трансформерный модуль и реконструкционный анализ. Метод обеспечивает устойчивое выделение локальных и глобальных дефектов и работает при отсутствии разметки. Эксперименты показывают повышение точности и устойчивости к доменным смещениям по сравнению с существующими подходами, что подтверждает его применимость в промышленном контроле качества.

Ключевые слова: детектирование аномалий, техническое зрение, роботизированные системы, гибридные модели; сверточные сети, трансформеры, реконструкция, контроль качества.

Введение

Задачи детектирования аномалий в визуальных данных роботизированных производственных систем осложняются высокой вариативностью технологических условий, шумами сенсоров и отсутствием достаточного числа размеченных примеров дефектов [1, 2]. Классические статистические методы и отдельные нейросетевые подходы демонстрируют ограниченную эффективность при работе с изображениями высокой размерности и нестабильных условий съёмки [3, 4]. Это определяет необходимость разработки методов, способных объединять локальные и глобальные признаки, обеспечивать устойчивость к доменному смещению и сохранять высокую точность при отсутствии разметки. В настоящей работе предлагается гибридная модель, сочетающая элементы сверточных нейронных сетей, трансформерных архитектур и реконструкционного анализа. Статья включает аналитический обзор существующих методов, формальную постановку задачи, описание предлагаемой архитектуры и результаты экспериментальной оценки на данных, моделирующих реальные условия промышленного производства.

1. Анализ современных методов детектирования аномалий в производственных данных

Классические статистические методы детектирования

Классические статистические методы применяются в задачах промышленного контроля благодаря простоте реализации и интерпретируемости. Они хорошо работают при анализе низкоразмерных данных, однако их эффективность существенно снижается при обработке изображений высокой размерности из-за нелинейной структуры признаков и сильной чувствительности к шумам [3, 5].

Одним из наиболее распространённых критериев статистического отклонения является расстояние Махаланобиса [6]

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}. \quad (1)$$

где x — вектор наблюдения размерности d ;

μ — вектор средних значений распределения нормальных данных;

Σ — ковариационная матрица нормального режима;

Σ^{-1} — обратная ковариационная матрица;

$D_M(x)$ — расстояние Махаланобиса, характеризующее степень отклонения наблюдения от нормы.

При наличии линейной структуры нормальных данных применяется метод главных компонент (PCA) [3], позволяющий выделять подпространство максимальной дисперсии. Разложение матрицы наблюдений задаётся выражением

$$X = USV^T, \quad (2)$$

где X — матрица данных размерности $n \times d$;

U, S, V — матрицы сингулярного разложения.

Восстановление наблюдения через подпространство основных компонент осуществляется по формуле

$$\hat{x} = V_k V_k^T x, \quad (3)$$

где \hat{x} — восстановленное наблюдение;

V_k — матрица первых k главных собственных векторов;

x — исходное наблюдение,

а метрика отклонения вычисляется как

$$e = \|x - \hat{x}\|_2. \quad (4)$$

где e — величина ошибки восстановления;

x — исходное наблюдение;

\hat{x} — восстановленное наблюдение, полученное через подпространство PCA;

$\|\cdot\|$ — евклидова норма.

С целью лучшего разделения нормальных и аномальных данных был предложен One-Class SVM [4]. Оптимизационная задача формулируется как

$$\min_{w, p, \xi} \left(\frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i - p \right), \quad (5)$$

где w — вектор весов гиперплоскости в пространстве признаков;

p — пороговое значение функции решения;

ξ_i — slack-переменные, допускающие ошибки классификации;

v — параметр, определяющий долю допускаемых ошибок;

n — количество обучающих наблюдений

при ограничениях

$$w^T \phi(x_i) \geq p - \xi_i, \xi_i \geq 0. \quad (6)$$

где $\phi(x_i)$ — отображение признаков в скрытое пространство (обычно RBF-ядро);

x_i — наблюдение из обучающей выборки;

ξ_i — slack-переменные;

p — пороговая величина решения;

w — параметр разделяющей поверхности.

Ещё одно направление применения классических методов — оценка плотности распределения вероятностей (KDE). Плотность определяется выражением

$$p(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad (7)$$

где $p(x)$ — оценённая плотность распределения вероятностей в точке x ;

n — число наблюдений;

h — ширина ядра;

d — размерность пространства признаков;

$K(\cdot)$ — ядровая функция (чаще всего — гауссово ядро);

x_i — наблюдение обучающей выборки.

Однако уже при $d > 20$ модель становится практически неработоспособной (проклятие размерности), а в задачах компьютерного зрения размерность данных достигает $10^5 - 10^6$, что делает KDE неприменимым.

Сравнительные исследования [7, 8] показывают, что традиционные методы работают удовлетворительно только на данных с низкой размерностью и стабильным распределением. Практические исследования показывают, что их качество существенно ниже, чем у современных нейросетевых моделей: на наборах дефектов AUC редко превышает 0,70–0,75, что ограничивает их использование в задачах промышленного визуального контроля.

Глубокие модели сегментации и их роль в задачах детектирования аномалий

Современные методы глубокого обучения позволяют эффективно анализировать изображения высокой размерности и учитывать как локальные, так и глобальные особенности объектов. Сверточные нейронные сети (CNN) обеспечивают извлечение локальных признаков и устойчивы к вариациям мелких структур, тогда как трансформерные архитектуры способны моделировать дальние зависимости за счёт механизма самовнимания [7–9].

Одной из наиболее применимых архитектур сегментации является U-Net, формирующая бинарную карту принадлежности пикселей классу [10]

$$f_{\theta} : R^H \times W \times C \rightarrow \{0,1\}^{H \times W}, \quad (8)$$

где H, W — размеры изображения;

C — число каналов входных данных;

f_{θ} — параметризуемая глубокая модель;

$\{0,1\}^{H \times W}$ — бинарная карта сегментации.

Функция потерь для задач бинарной сегментации часто задаётся через бинарную кросс-энтропию

$$L = -\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W [y_{ij} \ln p_{ij} + (1 - y_{ij}) \ln(1 - p_{ij})]. \quad (9)$$

где y_{ij} — истинная метка пикселя;

p_{ij} — предсказанная вероятность принадлежности классу;

L — функция ошибки, минимизируемая в процессе обучения.

При отсутствии аннотированных данных широко используются реконструкционные модели — автоэнкодеры (AE), вариационные автоэнкодеры (VAE) и их модификации. Они определяют аномалии по величине расхождения между входным и восстановленным изображением (4). В VAE минимизация осуществляется через дивергенцию распределений в скрытом пространстве

$$L_{VAE} = D_{KL}(q_{\phi}(z|x) || p(z)) + E_{q_{\phi}(z|x)} [\|x - \hat{x}\|_2^2]. \quad (10)$$

где $q_{\phi}(z|x)$ — аппроксимация апостериорного распределения;

$p(z)$ — априорное распределение;

D_{KL} — дивергенция Кульбака–Лейблера.

Появление Vision Transformer (ViT) позволило моделировать изображение как последовательность патчей фиксированного размера [7] $x \rightarrow \{p_1, p_2, \dots, p_N\}$. При этом глобальные зависимости описываются механизмом внимания

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right), \quad (11)$$

где Q, K — матрицы запросов и ключей;

d — размерность скрытого пространства;

A — матрица внимания, определяющая взаимосвязи между патчами.

Сравнительные исследования [7, 8] показывают, что CNN-архитектуры превосходят трансформеры в поиске мелких локальных дефектов (царапины, поры, включения), тогда как трансформеры лучше выявляют крупные структурные изменения и глобальные нарушения формы.

Поэтому глубокие модели сегментации рассматриваются как ключевой компонент в построении гибридных методов детектирования аномалий, которые объединяют локальный анализ CNN, глобальный контекст трансформеров и устойчивость реконструкционных моделей для повышения точности контроля качества в условиях высокой вариативности производственных данных.

2. Проблемы и ограничения существующих решений на роботизированных линиях производства

Применение существующих методов детектирования аномалий в роботизированных производственных системах осложняется высокой вариативностью условий съёмки, технологическими шумами и нестабильностью распределений данных [7, 8]. Наблюдаемое изображение можно представить в виде

$$x_t = x_t^* + \eta_t, \quad (12)$$

где x_t^* — идеальное (шумосвободное) изображение;

η_t — совокупный технологический шум.

Шум η_t имеет сложную структуру и включает несколько компонент

$$\eta_t = \eta_t^{illum} + \eta_t^{motion} + \eta_t^{sensor}, \quad (13)$$

где η_t^{illum} — шумы освещения;

η_t^{motion} — смещения и размытость из-за вибраций;

η_t^{sensor} — шумы сенсора видеокамеры.

Такой тип шумов существенно затрудняет применение моделей, предполагающих стационарность распределения данных (как в статистических методах) или чистоту входа (как в реконструкционных моделях).

Ещё одним важным ограничением является редкость аномалий. В реальных производственных условиях соотношение «аномалия/норма» часто составляет менее 1 %: $\frac{|A|}{|N|} < 0.01$. Такой дисбаланс приводит к переобучению нейросетевых моделей на нормальные паттерны и снижению чувствительности к редким видам дефектов.

Значительное влияние оказывает и смещение распределения данных (domain shift), возникающее при изменении типа сырья, условий освещения, конфигурации оборудования и других параметров производственного процесса. В таких условиях выполняется

$$p_t(x) \neq p_{t+1}(x), \quad (14)$$

где $p_t(x)$ — распределение данных в момент времени t ;

$p_{t+1}(x)$ — распределение данных после изменения технологических условий.

Большинство моделей глубокого обучения чувствительны к таким смещениям и требуют дополнительного дообучения, что затруднительно в условиях ограниченных вычислительных ресурсов промышленного оборудования.

В совокупности эти факторы приводят к тому, что существующие методы детектирования аномалий недостаточно устойчивы к производственным шумам, плохо адаптируются к изменениям технологических условий и не обеспечивают требуемой скорости обработки. Это делает необходимым создание гибридных подходов, объединяющих локальный анализ CNN, глобальный контекст трансформеров и устойчивость реконструкционных моделей.

3. Постановка задачи и разработка гибридного метода

Рассматривается множество визуальных наблюдений $X = \{x_1, x_2, \dots, x_n\}$, тогда каждое наблюдение может относиться либо к нормальному классу, либо к классу аномалий

$$x_i \in \begin{cases} N, \text{нормальный режим,} \\ A, \text{аномалия.} \end{cases}$$

При этом для производственных данных характерна сильная несбалансированность классов: $|A| \ll |N|$, что

затрудняет использование традиционных идеологий обучения. Наблюдение подвержено технологическим шумам, структура которых ранее была представлена в формуле (12), что приводит к вариативности распределений данных и усложняет выявление дефектов.

Задача заключается в построении функции отклонения $D(x)$, обеспечивающей разделение классов

$$D(x) > \tau \Rightarrow x \in A, \quad (15)$$

где $D(x)$ — функция отклонения, формируемая моделью;

τ — порог детектирования, определяемый статистически или на основе реконструкционной ошибки.

Для оценки качества работы метода используются стандартные метрики бинарной классификации

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, \quad (16)$$

где TP — число истинно положительных срабатываний;

FP — ложноположительные срабатывания;

FN — ложноотрицательные решения.

Таким образом, требуется построить устойчивую метрику отклонения $D(x)$, учитывающую шум, нестационарность и ограниченность разметки.

Архитектура предлагаемой гибридной модели

Предлагаемая архитектура объединяет локальные признаки, извлекаемые CNN, глобальные зависимости, моделируемые трансформером, и реконструкционный модуль для восстановления нормального состояния изображения. CNN-encoder обеспечивает обработку мелких локальных структур, тогда как трансформер формирует представление на уровне патчей, взаимодействие которых определяется матрицей внимания, приведённой ранее (11).

Для объединения скрытых представлений используется взвешенное смешивание

$$z = \alpha z_{cnn} + (1 - \alpha) z_{tr}, \quad (17)$$

где z_{cnn} — локальные признаки, извлечённые CNN;

z_{tr} — глобальные признаки трансформера;

α — коэффициент, определяющий вклад локального и глобального анализа.

Реконструкционный модуль восстанавливает \hat{x} , а отклонение определяется через ошибку реконструкции (см. формулу (4)). Такая схема обеспечивает устойчивость к шумам и доменным смещениям, а также чувствительность к локальным и глобальным дефектам.

Функция потерь и критерии оптимизации

Для повышения точности детектирования используется комбинированная функция потерь, включающая реконструкционную и признаковую составляющие. Ошибка в пространстве признаков задаётся следующим выражением

$$L_{emb} = \|z_{cnn} - \hat{z}_{cnn}\|_2^2 + \|z_{tr} - \hat{z}_{tr}\|_2^2, \quad (18)$$

где z_{cnn}, z_{tr} — локальные и глобальные признаки исходного изображения;

$\hat{z}_{cnn}, \hat{z}_{tr}$ — признаки реконструкции;

$\|\cdot\|$ — евклидова норма.

Итоговая функция потерь имеет вид

$$L = \lambda_1 L_{rec} + \lambda_2 L_{emb}, \quad (19)$$

где L_{rec} — реконструкционная ошибка (4);

L_{emb} — ошибка в пространстве признаков (18);

λ_1, λ_2 — коэффициенты, регулирующие вклад каждого компонента.

Выбор коэффициентов позволяет регулировать чувствительность модели к локальным или глобальным искажениям.

Практическая интеграция в производственный цикл

При внедрении метода в промышленные системы учитываются ограничения по времени обработки, нестабильность условий съёмки и ограниченность ресурсов встраиваемого оборудования. Для обеспечения работы в режиме, близком к реальному времени, применяются методы оптимизации: квантование параметров, уменьшение размерности признаков и упрощение операций самовнимания.

Для компенсации доменных смещений используются адаптивные обновления статистики нормального режима и периодическое дообучение отдельных блоков. Решения модели интегрируются в систему управления производством, где требуется минимизация ложных срабатываний и обеспечение интерпретируемости.

4. Экспериментальная оценка метода

Эксперименты проводились на данных, моделирующих условия функционирования роботизированных производственных линий. Использовались изображения поверхностей металлических и композиционных материалов, содержащие нормальные образцы и аномалии различных типов. Набор данных включал как реальные производственные снимки, так и синтетически добавленные дефекты для контроля структуры аномальных событий.

Обучение модели выполнялось только на нормальных изображениях, что соответствует промышленному сценарию с редкими аномалиями. Для оценки точности использовались метрики AUC-ROC, Precision, Recall и F1 (16). Перед обучением выполнялась нормализация интенсивности, масштабирование изображений и разбиение на обучающую и тестовую части. Для повышения объективности результатов проводилась перекрёстная проверка с несколькими разбиениями данных.

Такой экспериментальный протокол позволяет достоверно оценить способность модели работать в условиях реального производственного процесса при наличии технологических шумов, вариативности освещения и доменных смещений.

Сравнение с существующими методами

Для сравнения эффективности гибридной модели были использованы статистические методы (PCA, One-Class SVM, KDE), реконструкционные нейросетевые модели (AE, VAE), а также сегментационные и трансформерные архитектуры [3, 7, 8]. Классические методы демонстрируют низкую устойчивость к дефектам малой площади и заметно уступают современным нейросетям. Реконструкционные архитектуры обеспечивают более высокую точность, но склонны сглаживать мелкие аномалии. Трансформерные модели хорошо выявляют крупные структурные нарушения, однако имеют высокую вычислительную сложность и недостаточную устойчивость к шумам.

Для количественного сравнения использовались метрики AUC-ROC, Precision, Recall и F1-мера. Результаты приведены в таблице 1.

Предложенная архитектура демонстрирует лучшие результаты по основным метрикам, что подтверждает преимущества объединения локальных и глобальных признаков.

Результаты и анализ эффективности

Для детального анализа оценивалась эффективность выявления различных типов аномалий: мелких поверх-

Таблица 1.
Результаты сравнительного анализа методов

Метод	AUC-ROC	Precision	Recall	F1
PCA	0,71	0,52	0,48	0,50
One-Class SVM	0,74	0,55	0,57	0,56
KDE	0,63	0,49	0,41	0,45
AE	0,85	0,71	0,66	0,68
VAE	0,88	0,74	0,69	0,71
U-Net (без разметки аномалий)	0,82	0,69	0,61	0,65
ViT	0,90	0,78	0,72	0,75
Гибридная модель	0,95	0,86	0,83	0,84

ностных повреждений, текстурных искажений, разрывов структуры и крупных деформаций. В таблице ниже приведены значения F1-меры для различных классов дефектов. Таблица 2 отражает сравнительные значения F1-меры для различных типов дефектов.

Таблица 2.
Точность выявления различных типов аномалий (F1-мера)

Тип дефекта	PCA	AE	VAE	ViT	Гибридная модель
Мелкие поверхностные дефекты	0,42	0,61	0,64	0,68	0,81
Локальные разрывы / царапины	0,47	0,66	0,69	0,72	0,85
Сложные текстурные отклонения	0,50	0,58	0,62	0,71	0,84
Крупные геометрические деформации	0,67	0,78	0,80	0,89	0,92
Шумы/артефакты освещения	0,39	0,55	0,60	0,74	0,88

Дополнительно проведён анализ зависимости времени обработки кадра от размера входного изображения. Результаты представлены в таблице 3.

Таблица 3.
Среднее время обработки одного кадра (мс)

Метод	256×256	512×512	768×768
AE	4,1	7,8	12,4
VAE	5,3	9,6	15,7
ViT	18,5	34,1	61,3
Гибридная модель	9,4	16,8	28,9

Таким образом, гибридная архитектура достигает оптимального баланса между точностью, устойчивостью и вычислительной производительностью, что делает её особенно подходящей для внедрения в роботизированные линии промышленного контроля качества.

Несмотря на высокую точность, гибридный метод остаётся чувствительным к изменению статистики нормального режима и требует периодического обновления эталонных данных при существенном доменном смещении. При неправильной настройке коэффициентов функции потерь (19) возможны как пропуски слабовыраженных дефектов, так и избыточные ложные срабатывания, особенно при нестабильном освещении и шуме сенсора. В условиях ограниченных ресурсов скорость обработки зависит от степени оптимизации трансформерного модуля и реконструкционного блока, однако даже при этих ограничениях метод демонстрирует устойчивость к технологическим шумам и вариативности структуры поверхности, что позволяет применять его в реальных роботизированных системах контроля качества.

Заключение

В работе предложен гибридный метод детектирования аномалий, объединяющий сверточные сети, трансформерный модуль и реконструкционный анализ. Использование локальных и глобальных признаков позволило повысить точность и устойчивость к шумам и доменным смещениям по сравнению с классическими и современными нейросетевыми подходами. Экспериментальная оценка подтвердила эффективность модели при работе с визуальными данными производственных систем и показала, что метод обеспечивает требуемое время обработки для применения в роботизированных линиях контроля качества. Дальнейшее развитие связано с улучшением адаптации к изменяющимся условиям и внедрением методов самообучения в условиях дефицита размеченных данных.

ЛИТЕРАТУРА

1. Kagermann H., Wahlster W., Helbig J. Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group. Acatech — National Academy of Science and Engineering, 2013. 78 p.
2. Lee J., Bagheri B., Kao H.-A. A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 2015, vol. 3, pp. 18–23.
3. Jolliffe I.T. *Principal Component Analysis*. Springer Series in Statistics. 2nd ed. New York: Springer, 2002. 488 p.
4. Новикова А.С., Ромашкова О.Н. Интеграция нейросетей в информационные системы розничных торговых сетей: прогнозирование и управление распределением ресурсов // *Современная наука: актуальные проблемы теории и практики*. Серия: Естественные и технические науки. 2024. № 1-2. С. 49–52.
5. Mahalanobis P.C. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 1936, vol. 2(1), pp. 49–55.
6. Гурарий С.А. Методы анализа изображений в автоматизированных системах контроля качества. М.: Машиностроение, 2013. 248 с.
7. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI 2015, LNCS*, vol. 9351, Springer, pp. 234–241.
8. Dosovitskiy A. et al. An image is worth 16×16 words: Transformers for image recognition at scale. *ICLR 2021*. arXiv:2010.11929.
9. Ломовцев Р.С., Ромашкова О.Н., Пономарева Л.А. Алгоритм интеллектуальной поддержки управленческих решений для региональной образовательной системы // *Вестник Брянского государственного технического университета*. 2018. № 10 (71). С. 35–43.
10. Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox T., Ronneberger O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *MICCAI 2016, LNCS*, vol. 9901, Springer, pp. 424–432.

© Дёмин Владислав Дмитриевич (neuronich@sofa-invest.ru); Ромашкова Оксана Николаевна (ox-rom@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»