

УДК 004.932.2:004.85

EDN WMTVBA

ПОЭТАПНОЕ ОБУЧЕНИЕ ДЛЯ РОБАСТНОГО ОБНАРУЖЕНИЯ ОБЪЕКТОВ В СИСТЕМАХ ПРЕДИКТИВНОГО МОНИТОРИНГА

В.Н. СичкарORCID: 0000-0001-9825-0881 e-mail: valik123@gmail.com

Национальный исследовательский университет ИТМО

Санкт-Петербург, Россия

Д.И. МуромцевORCID: 0000-0002-0644-9242 e-mail: d.muromtsev@gmail.com

Национальный исследовательский университет ИТМО

Санкт-Петербург, Россия

Построена методология обучения системы предиктивного мониторинга, повышающая робастность обнаружения к перекрытию без увеличения вычислительной нагрузки в режиме инференса, что критически важно для развертывания на бортовых и периферийных вычислительных устройствах. Предлагается фреймворк из трех согласованных стратегий: поэтапного обучения с информационно-теоретической метрикой сложности, динамического адаптивного взвешивания функции потерь по байесовской оценке неопределенности и прогрессивной синтетической аугментации перекрытий. Компоненты действуют исключительно на этапе обучения и не модифицируют итоговую сеть. Эксперименты на наборе данных *MS COCO-Occluded* демонстрируют рост метрики *Fragment-Aware mAP* с 81,7 до 89,3 % при одновременном сокращении числа эпох обучения на 38 %. Предложенный подход обладает свойством переносимости: согласованный прирост 8,0-9,1 % абсолютной точности воспроизводится на детекторах YOLOv8-x, YOLOv9-c и YOLOv10-m, наиболее распространенных в задачах предиктивной диагностики. Практическая применимость метода для систем мониторинга в режиме реального времени и снижения эксплуатационных затрат при переобучении моделей под новые условия съемки обоснована полученными результатами.

Ключевые слова: техническое зрение; обнаружение объектов; частичное перекрытие; робастность модели; поэтапное обучение; предиктивное обслуживание; нейронные сети YOLO; глубокое обучение; адаптивное взвешивание потерь; байесовская неопределенность.

ДЛЯ ЦИТИРОВАНИЯ: Сичкар, В.Н. Поэтапное обучение для робастного обнаружения объектов в системах предиктивного мониторинга / В.Н. Сичкар, Д.И. Муромцев // Труды НГТУ им. П.Е. Алексеева. 2026. № 3. С. 24-33. EDN WMTVBA

CURRICULUM LEARNING FOR ROBUST OBJECT DETECTION IN PREDICTIVE MONITORING SYSTEMS

V.N. SichkarORCID: 0000-0001-9825-0881 e-mail: valik123@gmail.com

ITMO University

Saint-Petersburg, Russia

D.I. MouromtsevORCID: 0000-0002-0644-9242 e-mail: d.muromtsev@gmail.com

ITMO University

Saint-Petersburg, Russia

Abstract. This paper presents a methodology for training a predictive monitoring system that improves the occlusion robustness of detection without increasing the inference-time computational cost, which is critical for deployment on onboard and edge devices. A framework of three coordinated strategies is proposed: curriculum learning driven by an information-theoretic difficulty metric, dynamic adaptive loss weighting based on Bayesian uncertainty, and pro-

gressive synthetic occlusion augmentation. The components operate only during the training stage and do not modify the final network. Experiments on the MS COCO-Occluded dataset demonstrate an increase in the Fragment-Aware mAP metric from 81.7 to 89.3 % while simultaneously reducing the number of training epochs by 38%. The proposed approach exhibits transferability: a consistent improvement of 8.0 to 9.1% in absolute accuracy is reproduced on YOLOv8-x, YOLOv9-c and YOLOv10-m detectors, which are the most common in predictive diagnostics tasks. The obtained results substantiate the practical applicability of the method for real-time monitoring systems and the reduction of operational costs when retraining models for new imaging conditions.

Keywords: computer vision; object detection; partial occlusion; model robustness; curriculum learning; predictive maintenance; YOLO neural networks; deep learning; adaptive loss weighting; Bayesian uncertainty.

FOR CITATION: Sichkar V.N. and Mouromtsev D.I. Curriculum Learning for Robust Object Detection in Predictive Monitoring Systems. Transactions of NNSTU n.a. R.E. Alekseev. 2026. № 3. Pp. 24-33. EDN WMTVBA

Введение

Эффективность современной системы предиктивного мониторинга определяется не только качеством сенсорики, но и скоростью реакции вычислительного модуля, который превращает изображение в управляющий сигнал. В сценариях осмотра линий электропередачи беспилотным летательным аппаратом, в роботизированных линиях контроля упаковки и при автоматическом сканировании готовой продукции на конвейере детектор объектов функционирует как центральный элемент киберфизической цепочки: данные с камеры (физический мир) преобразуются алгоритмом в решение (кибер-мир), а решение немедленно становится действием – отбраковкой, оповещением оператора или планированием технического обслуживания. Поэтому даже единичная ошибка детектирования влечет реальную потерю – выпуск дефектной продукции, пропуск повреждения опоры или ложный вызов выездной бригады.

Серьезным барьером для применения детекторов в подобных задачах является устойчивое к перекрытию обнаружение. По различным оценкам, от 60 до 80 % наблюдаемых объектов в полевых условиях частично скрыты [1]: опорные конструкции перекрывают изоляторы, растительность экранирует элементы трубопроводов, а смежные детали на конвейере взаимно закрываются. При этом большинство публикуемых решений сосредоточено на архитектурных модификациях [2-4]: на добавлении модулей внимания, выделении контекстных признаков или ансамблировании. Процесс обучения – стратегия сэмплирования, схема взвешивания компонент потерь, тип аугментации обычно остается стандартным, что ограничивает потенциал даже самых современных архитектур.

Цель настоящей работы – предложить методологию обучения, повышающую устойчивость детектирования к перекрытиям без увеличения вычислительной нагрузки на инференс, что особенно важно для размещения моделей на бортовых вычислителях беспилотников и периферийных устройствах. Новизна работы определяется тремя компонентами: предложенной информационно-теоретической мерой сложности, количественно описывающей восстанавливаемое содержание перекрытого объекта; правилом адаптивного взвешивания потерь, использующим байесовскую оценку неопределенности через *Monte Carlo Dropout*; и трехэтапной прогрессивной аугментацией, согласованной с расписанием обучения и приближающей синтетические перекрытия к статистике реальных полевых сцен. Все три компонента применяются только во время обучения и оставляют конечную развернутую модель без изменений, что обеспечивает прямую совместимость с существующими стеками инференса. Прирост качества и сокращение вычислительных затрат достигаются переносом нагрузки с архитектуры на методологию обучения – направление, которое в литературе до сих пор изучено заметно слабее.

Постановка задачи и формальная модель

Базовый сценарий, в котором рассматривается работа детектора, имеет следующую структуру. Сенсор – бортовая камера беспилотника или стационарная камера над производственной зоной – формирует поток изображений; вычислительный модуль выполняет инференс детектора с фиксированной частотой (как правило, не выше 30:40 мс на кадр); полученные детекции поступают в управляющую логику, формирующую сигнал предиктивного обслуживания или решение о действии. Такие частотные рамки оставляют этап обучения единственным резервом для повышения качества без изменения вычислительных ресурсов.

Введем формальное описание задачи. Пусть I – входное изображение размера $H \times W \times 3$, M – бинарная маска перекрытия того же пространственного размера, Y – множество разметок объектов на изображении. Детектор f_θ с параметрами θ формирует множество детекций $D = f_\theta(I)$. Для каждого истинного объекта o с рамкой b_o^{gt} введем коэффициент видимости как долю незатененных пикселей внутри этой рамки, а степень перекрытия – как дополнение видимости до единицы:

$$v(o) = \text{Area}(b_o^{gt} \cap \{M = 0\}) / \text{Area}(b_o^{gt}), \quad \alpha(o) = 1 - v(o), \quad (1)$$

где $v(o) \in [0, 1]$ – коэффициент видимости объекта o , $\alpha(o) \in [0, 1]$ – степень его перекрытия. Стандартная процедура обучения минимизирует математическое ожидание функции потерь при равномерном сэмплировании из совместного распределения $P(I, Y, M)$. Это предположение в полевых условиях нарушается: распределение $\alpha(o)$ смещено в область высоких значений, а сложность обучения нелинейно зависит от $\alpha(o)$. Задача состоит в том, чтобы заменить равномерный сэмплер адаптивным распределением P_t , изменяющимся по эпохам t , и ввести динамические веса $\lambda_t(I, Y, M)$ для компонент составной функции потерь.

Информационно-теоретический фреймворк обучения

Предлагаемая методология включает три согласованные стратегии, применяемые исключительно на этапе обучения детектора: поэтапное обучение с информационно-теоретической метрикой сложности, динамическое адаптивное взвешивание функции потерь и прогрессивную синтетическую аугментацию. Каждая стратегия адресует сторону проблемы перекрытия – временную структуру обучения, относительный вклад компонент потерь и характер аугментаций, а в совокупности они образуют целостный фреймворк.

Поэтапное обучение с информационно-теоретической метрикой сложности

Идея поэтапного обучения, впервые сформулированная Bengio и соавторами [5] и систематизированная в обзоре [6], заключается в предъявлении сети обучающих примеров в порядке возрастания сложности. Для задачи перекрытия «сложность» нельзя свести к одному параметру: она зависит как от величины $\alpha(o)$, так и от семантики объекта, и от информативности окружающего его контекста. Чтобы корректно описать эту многомерность, введем информационно-теоретическую меру сложности.

Пусть F_{vis} – признаки, извлекаемые из видимой части объекта, F_{ctx} – признаки контекстного окружения, O – полное представление объекта. Восстановимая информация определяется как сумма взаимной информации видимых признаков с объектом и условной взаимной информации контекста с объектом при заданных видимых признаках. Через условную энтропию это представление эквивалентно записи $I_{rec}(o) = H(O) - H(O | F_{vis}, F_{ctx})$. Нормированная мера сложности тогда:

$$D(o) = 1 - I_{rec}(o) / H(O) = H(O | F_{vis}, F_{ctx}) / H(O), \quad (2)$$

где $H(\cdot)$ – дифференциальная энтропия, оценка которой выполняется по плотности k -ближайших соседей в признаковом пространстве [6]. Значение $D(o) \in [0,1]$ показывает, насколько недостаточно видимой и контекстной информации для надежной реконструкции объекта. На основании этой меры определяется эмпирирующее распределение $\pi_t(D)$, которое выбирает примеры со сложностью, близкой к текущему порогу $D_{max}(t)$:

$$D_{max}(t) = D_{init} + (1 - D_{init}) \cdot \sigma(t / T), \quad (3)$$

где $D_{init} = 0,3$ – начальный порог сложности, T – общее число эпох обучения, σ – сигмоидальная функция. Сигмоидальный профиль обеспечивает плавный переход от легких примеров к трудным и согласуется с эмпирическими наблюдениями [6]. Постепенное расширение границы $D_{max}(t)$ реализует последовательность обучения, которая критически важна в задачах полевого мониторинга: на ранних эпохах сеть формирует устойчивые фрагментарные признаки на легких примерах, а в дальнейшем учится восстанавливать объекты с сильным перекрытием, не теряя ранее усвоенных представлений.

Динамическое адаптивное взвешивание функции потерь на основе байесовской оценки неопределенности

Совокупная функция потерь в задаче обнаружения перекрытых объектов включает несколько компонент: классификационную L_{cls} , локализационную L_{loc} , потерю внимания L_{att} , потерю восстановления признаков L_{comp} и потерю согласованности L_{cons} . Применение фиксированных весовых коэффициентов игнорирует и эмпл-специфические особенности (для сильно перекрытых объектов критичнее восстановление признаков), и временные фазы обучения (на ранних эпохах важнее классификация, на поздних – точная локализация). В [7] показано, что взвешивание потерь через гомоскедастическую неопределенность обеспечивает ощутимый прирост качества; в [8] родственная идея реализована в фокальной потере для балансировки легких и сложных примеров. Развивая эту линию исследований для составных потерь при частичной наблюдаемости, предлагается следующее правило.

Неопределенность каждой компоненты k оценивается методом *Monte Carlo Dropout*, теоретически обоснованным в [9] как байесовская аппроксимация. На каждой итерации детектор прогоняется $S=10$ раз с активным *dropout*, и выборочная дисперсия u_k предсказаний принимается за оценку неопределенности. Адаптивный коэффициент вычисляется как:

$$\lambda_k(t) = \lambda_{k,0} \cdot \exp(-u_k / \tau), \quad k \in \{cls, loc, att, comp, cons\}, \quad (4)$$

где $\lambda_{k,0}$ – начальный весовой коэффициент компоненты, u_k – оценка ее выборочной дисперсии по S dropout-проходам, $\tau = 0,5$ – температурный параметр, регулирующий чувствительность правила. Высокая неопределенность указывает на плохо освоенную компоненту, и она получает увеличенный вес на следующем шаге градиентного спуска. После каждого обновления коэффициенты нормируются так, чтобы их сумма оставалась постоянной – это предотвращает вырождение составной функции потерь в одну компоненту, что особенно важно при длинных циклах переобучения, типичных для систем предиктивного мониторинга.

Прогрессивная синтетическая аугментация перекрытий

Применяемые на практике аугментации перекрытий – это прямоугольные маски, *random erasing* [10] и аналогичные формируют распределение синтетических искажений, плохо согласованное со статистикой реальной полевой сцены, что приводит к разрыву «синтетика – реальность». Особенно остро это проявляется в задачах БПЛА-инспекции, где реальные окклюдеры имеют сложную органическую форму, а не прямоугольную. Предлагается трехэтапная прогрессивная схема, в которой реалистичность и сложность маскирующих объектов синхронизированы с расписанием обучения. На стадии 1 (эпохи 1-100) применяются прямоугольные маски; на стадии 2 (эпохи 101-200) – маски с органическими очертаниями,

имитирующие соседние конструкции; на стадии 3 (эпохи 201-320) – маски, синтезируемые условным GAN-генератором, обученным на парах «чистое изображение – реальная маска перекрытия». Итоговая маска формируется как взвешенная комбинация трех типов:

$$M_{final}(t) = w_1(t) \cdot M_{rect} + w_2(t) \cdot M_{organic} + w_3(t) \cdot M_{GAN}, \quad (5)$$

где веса $w_i(t)$ задаются косинусным профилем, обеспечивающим плавный переход между стадиями. Синхронизация PSOА с расписанием поэтапного обучения обеспечивает синергический эффект: легкие примеры из расписания получают простые прямоугольные маски и используются для формирования базы фрагментарных признаков, тогда как трудные примеры аугментируются максимально реалистичными GAN-перекрытиями и приближают обучающее распределение к статистике реальной полевой сцены, которая регистрируется бортовой камерой беспилотника при инспекции инфраструктурных объектов.

Методика и условия эксперимента

Экспериментальная проверка проведена на наборе данных MS COCO-Occluded – расширении классического MS COCO [11], в котором к 118 тыс. обучающих и 5 тыс. валидационных изображений добавлены пиксельные маски перекрытия трех уровней сложности (легкое, умеренное, сильное). Выбор продиктован тем, что MS COCO является общепринятым промышленным эталоном, что обеспечивает воспроизводимость и сопоставимость с другими исследованиями, расширенная аннотация покрывает 847 тыс. экземпляров, достаточно для статистически устойчивых выводов, а разнообразие категорий объектов делает выводы переносимыми на различные сценарии предиктивного мониторинга.

В качестве базовых детекторов используются три однопроходные модели семейства YOLO: YOLOv8-x, YOLOv9-c [12] и YOLOv10-m [13]. Выбор продиктован практикой развертывания на периферийных вычислительных устройствах: именно YOLO-семейство массово применяется в полевых системах машинного зрения благодаря балансу точности и скорости инференса 25:30 кадров в секунду. Здесь YOLOv8-x представляет современный CNN-канон, YOLOv9-c реализует концепцию программируемой градиентной информации, а YOLOv10-m исключает процедуру подавления немаксимумов, что упрощает интеграцию в управляющий контур. Эксперименты выполнены на четырех ускорителях NVIDIA A100 80 ГБ при размере мини-батча 64, с оптимизатором *AdamW*, косинусным расписанием темпа обучения и смешанной точностью FP16. Итоговая валидация выполнена методом 5-блочной перекрестной проверки. Кроме стандартной метрики mAP@0,5 [11] применяются специализированные метрики, введенные в [3]. Основной является *Fragment-Aware mAP* (FA-mAP) – взвешенное по степени перекрытия среднее значение точности, в котором сильно перекрытые объекты получают больший вклад:

$$FA - mAP = \frac{\sum_o w(\alpha(o)) \cdot AP(o)}{\sum_o w(\alpha(o))}, \quad w(\alpha) = 1 + \gamma \cdot \alpha, \quad (6)$$

где $\gamma = 1,5$ – коэффициент усиления, эмпирически калиброванный по согласованности с экспертной оценкой; $AP(o)$ – средняя точность для объекта o ; $\alpha(o)$ – степень его перекрытия.

В качестве дополнительной метрики используется коэффициент устойчивости (Robustness Score, RS), определяемый как отношение FA-mAP при сильном перекрытии к FA-mAP при отсутствии перекрытия и характеризующий сохранение производительности модели при росте α .

Результаты и их обсуждение

В табл. 1 приведено сравнение базовой архитектуры ACRAFD [3], обученной по стандартной процедуре, с полным предложенным фреймворком, примененным поверх той же архитектуры ACRAFD, а также вклад каждой из стратегий по отдельности. Архитектура между

не изменяется – изменяется только процедура обучения. Все значения получены на валидационной части MS COCO-Occluded.

Таблица 1.
Влияние адаптивных стратегий обучения на качество и эффективность обучения детектора на MS COCO-Occluded

Table 1.
Influence of adaptive learning strategies on the quality and efficiency of detector training on MS COCO-Occluded

Метод обучения	FA-mAP, %	mAP@0,5, %	RS, %	Эпохи
ACRAFD (базовое обучение)	81,7	81,7	94,8	350
ACRAFD + поэтапное обучение	86,1	86,4	97,2	217
ACRAFD + динамические веса	84,6	84,9	96,1	315
ACRAFD + прогрессивная аугментация	84,3	84,6	95,7	328
ACRAFD + полный фреймворк	89,3	89,7	98,4	217
Прирост к базовому ACRAFD	+7,6	+8,0	+3,6	–38 %

Наибольший индивидуальный вклад вносит поэтапное обучение (+4,4 % FA-mAP при сокращении числа эпох на 38 %), что согласуется с теоретическими ожиданиями [5, 6]. Адаптивное взвешивание добавляет еще +2,9 % FA-mAP за счет более рационального использования градиентной информации. Прогрессивная аугментация дает +2,6 %, причем ее вклад особенно заметен в зоне сильных перекрытий. Совокупный эффект (+7,6 %) превышает сумму индивидуальных вкладов (+4,7 %), что подтверждает наличие синергии между стратегиями: расписание обучения, неопределенность и реалистичность аугментаций взаимно усиливаются. Дополнительный интерес представляет анализ качества по уровням перекрытия (табл. 2), позволяющий понять, в каком именно режиме фреймворк дает наибольший эффект.

Таблица 2.
Стратификация качества обнаружения (FA-mAP, %) по уровням перекрытия объектов

Table 2.
Stratification of detection quality (FA-mAP, %) by object occlusion levels

Метод обучения	Легкое (10-30 %)	Умеренное (30-50 %)	Сильное (50-70 %)	Среднее
ACRAFD (базовое обучение)	91,8	87,3	78,9	86,0
ACRAFD + поэтапное обучение	95,2	91,7	84,1	90,3
ACRAFD + динамические веса	93,7	89,6	81,4	88,2
ACRAFD + прогрессивная аугментация	93,4	89,2	82,7	88,4
ACRAFD + полный фреймворк	96,8	93,4	87,2	92,5

Полученные данные показывают, что наибольший абсолютный прирост (+8,3 процентных пункта) обеспечивается именно в режиме сильного перекрытия, который и составляет основную трудность для систем предиктивного мониторинга. В этих условиях стан-

дартно обученная модель пропускает значительную долю дефектных или поврежденных объектов, тогда как предложенный фреймворк целенаправленно компенсирует именно этот режим работы. Для оценки возможностей переноса предложенного подхода на различные архитектуры YOLO в табл. 3 приведены значения метрики $mAP@0,5$, полученные при применении одной процедуры обучения. Базовое обучение здесь означает стандартную процедуру с равномерным сэмплированием и фиксированными весами потерь.

Таблица 3.
Перенос предложенного фреймворка на различные архитектуры YOLO.
Метрика – $mAP@0,5$, %

Table 3.
Transfer of the proposed framework to various YOLO architectures. Metric: $mAP@0,5$, %

Архитектура	Базовое обучение	Адаптивное обучение	Прирост, %
YOLOv8-x	71,2	79,5	+8,3
YOLOv9-c	71,8	80,9	+9,1
YOLOv10-m	70,5	78,8	+8,3
Базовая модель ACRAFD	81,7	89,7	+8,0

Значения приростов в диапазоне 8,0:9,1 % при среднеквадратичном отклонении менее половины процента показывают, что предложенные стратегии не привязаны к конкретной архитектуре и могут быть применены к разным моделям, совместимой с градиентной оптимизацией. Это важно для практики, так как парки бортовых вычислителей беспилотников и периферийных устройств обычно стандартизованы под единое семейство детекторов, и предложенный подход может быть внедрен без замены существующего инференс-стека.

Заключение

В работе предложен и экспериментально проверен фреймворк адаптивных стратегий обучения для робастного обнаружения объектов при частичном перекрытии в системах предиктивного мониторинга на основе технического зрения. Он объединяет три согласованные компоненты: поэтапное обучение с информационно-теоретической мерой сложности, динамическое адаптивное взвешивание функции потерь на основе байесовской оценки неопределенности и прогрессивную синтетическую аугментацию перекрытий. Формализована сложность перекрытия через информационно-теоретическую меру восстановимой информации, построены схема взвешивания, опирающаяся на байесовскую неопределенность, и иерархическая схема прогрессивной аугментации, согласованная с расписанием обучения. Применение фреймворка к ранее предложенной архитектуре ACRAFD на наборе MS COCO-Occcluded [3] повышает метрику *Fragment-Aware mAP* с 81,7 до 89,3 % при одновременном сокращении числа эпох обучения на 38 %. Эксперименты переносимости подтверждают сохранение прироста в диапазоне 8,0:9,1 % на детекторах YOLOv8-x, YOLOv9-c и YOLOv10-m.

Полученные результаты позволяют рекомендовать предложенную методологию к практическому применению в системах предиктивного мониторинга на основе технического зрения, в задачах БПЛА-инспекции инфраструктурных объектов и автоматического контроля качества на периферийных вычислительных устройствах. Особенно важно, что все компоненты фреймворка действуют исключительно на этапе обучения и оставляют развернутую модель неизменной: это означает совместимость с существующими стеками инференса, отсутствие увеличения задержки и потребления памяти на бортовых вычислителях, а также возможность развертывания модели без изменения существующих эксплуатационных регла-

ментов. Дальнейшие исследования будут направлены на интеграцию протокола мета-переноса для быстрой адаптации к новым условиям съемки и на валидацию фреймворка на отраслевых наборах данных систем предиктивного обслуживания.

Библиографический список

1. **Wang A.** Robust object detection under occlusion with context-aware compositional nets / A. Wang, Y. Sun, A. Kortylewski, A. Yuille // *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR 2020)*. – Seattle, USA, 2020. – P. 12642–12651.
2. **Carion N.** End-to-end object detection with transformers / N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko // *Computer vision – ECCV 2020. Lecture notes in computer science*, vol. 12346. – Springer, Cham, 2020. – P. 213-229.
3. **Sichkar V.N.** Advanced contextual reconstruction architecture for robust object detection under partial occlusion // *Problems of artificial intelligence*. – 2026. – Vol. 40. – P. 25-40.
4. **Finn C.** Model-agnostic meta-learning for fast adaptation of deep Networks / C. Finn, P. Abbeel, S. Levine // *Proceedings of the 34th International conference on machine learning (ICML 2017)*. PMLR, vol. 70. – Sydney, Australia, 2017. – P. 1126-1135.
5. **Bengio Y.** Curriculum learning / Y. Bengio, J. Louradour, R. Collobert, J. Weston // *Proceedings of the 26th Annual international conference on machine learning (ICML 2009)*. – Montreal, Canada, 2009. – P. 41-48.
6. **Soviany P.** Curriculum learning: a survey / P. Soviany, R.T. Ionescu, P. Rota, N. Sebe // *International journal of computer vision*. – 2022. – Vol. 130. – P. 1526-1565. – Springer.
7. **Kendall A.** Multi-task learning using uncertainty to weigh losses for scene geometry and semantics / A. Kendall, Y. Gal, R. Cipolla // *Proceedings of the IEEE Conference on computer vision and pattern recognition (CVPR 2018)*. – Salt Lake City, USA, 2018. – P. 7482-7491.
8. **Lin T.-Y.** Focal loss for dense object detection / T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár // *Proceedings of the IEEE International conference on computer vision (ICCV 2017)*. – Venice, Italy, 2017. – P. 2980-2988.
9. **Gal Y.** Dropout as a bayesian approximation: representing model uncertainty in deep learning / Y. Gal, Z. Ghahramani // *Proceedings of the 33rd International conference on machine learning (ICML 2016)*. PMLR, vol. 48. – New York, USA, 2016. – P. 1050-1059.
10. **Yun S.** CutMix: regularization strategy to train strong classifiers with localizable features / S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo // *Proceedings of the IEEE/CVF International conference on computer vision (ICCV 2019)*. – Seoul, Korea, 2019. – P. 6023-6032.
11. **Lin T.-Y.** Microsoft COCO: common objects in Context / T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick // *Computer Vision – ECCV 2014. Lecture notes in computer science*, vol. 8693. – Springer, Cham, 2014. – P. 740-755.
12. **Wang C.-Y.** YOLOv9: learning what you want to learn using programmable gradient information / C.-Y. Wang, I.-H. Yeh, H.-Y.M. Liao // *Computer vision – ECCV 2024. Lecture notes in computer science*, vol. 15089. – Springer, 2025. – P. 1-21.
13. **Wang A.** YOLOv10: Real-time end-to-end object detection / A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding // *Advances in neural information processing systems*. – 2024. – Vol. 37. – P. 107984-108011.
14. **He K.** Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // *Proceedings of the IEEE Conference on computer vision and pattern recognition (CVPR 2016)*. – Las Vegas, USA, 2016. – P. 770-778.
15. **Ren S.** Faster R-CNN: towards real-time object detection with region proposal networks / S. Ren, K. He, R. Girshick, J. Sun // *Advances in neural information processing systems*. – 2015. – Vol. 28. – P. 91-99.
16. **Vaswani A.** Attention is All You Need / A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin // *Advances in neural information processing systems*. – 2017. – Vol. 30. – P. 5998-6008.
17. **Zhang S.** Occlusion-aware R-CNN: detecting pedestrians in a crowd / S. Zhang, L. Wen, X. Bian, Z. Lei, S.Z. Li // *Proceedings of the european conference on computer vision (ECCV 2018)*. Lecture notes in computer science. – Springer, Cham, 2018. – P. 637-653.

18. **Zhao Y.** DETRs Beat YOLOs on real-time object detection / Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, J. Chen // Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR 2024). – Seattle, USA, 2024. – P. 16965-16974.
19. **Oks S.J.** Cyber-physical systems in the context of industry 4.0: a review, categorization and outlook / S.J. Oks, M. Jalowski, M. Lechner, S. Mirschberger, M. Merklein, B. Vogel-Heuser, K.M. Möslin // Information systems frontiers. – 2024. – Vol. 26, no. 5. – P. 1731-1772. – Springer.
20. **Carvalho T.P.** A systematic literature review of machine learning methods applied to predictive maintenance / T.P. Carvalho, F.A.A.M.N. Soares, R. Vita, R.P. Francisco, J.P. Basto, S.G.S. Alcalá // Computers and industrial engineering. – 2019. – Vol. 137. – Article 106024. – Elsevier. – DOI: 10.1016/j.cie.2019.106024 (в DOI нуждается в проверке).
21. **DeVries T.** Improved regularization of convolutional neural networks with cutout [Электронный ресурс] / T. DeVries, G.W. Taylor. – arXiv preprint arXiv:1708.04552. – 2017. – URL: <https://arxiv.org/abs/1708.04552> (дата обращения: 17.04.2026).
22. **Krizhevsky A.** ImageNet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // Advances in neural information processing systems. – 2012. – Vol. 25. – P. 1097–1105.

References

1. Wang A., Sun Y., Kortylewski A., Yuille A. Robust object detection under occlusion with context-aware compositional nets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020). Seattle, USA; 2020. p. 12642-12651. (in English).
2. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-end object detection with transformers. In: Computer Vision – ECCV 2020. Lecture Notes in Computer Science, vol. 12346. Springer, Cham; 2020. p. 213-229. (in English).
3. Sichkar V.N. Advanced contextual reconstruction architecture for robust object detection under partial occlusion. *Problemy iskusstvennogo intellekta* [Problems of Artificial Intelligence]. 2026;40:25-40. (in English).
4. Finn C., Abbeel P., Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML 2017). PMLR, vol. 70. Sydney, Australia; 2017. p. 1126-1135. (in English).
5. Bengio Y., Louradour J., Collobert R., Weston J. Curriculum learning. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML 2009). Montreal, Canada; 2009. p. 41-48. (in English).
6. Soviany P., Ionescu R.T., Rota P., Sebe N. Curriculum learning: a survey. *International Journal of Computer Vision*. 2022;130:1526-1565. Springer. (in English).
7. Kendall A., Gal Y., Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018). Salt Lake City, USA; 2018. p. 7482-7491. (in English).
8. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017). Venice, Italy; 2017. p. 2980-2988. (in English).
9. Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning (ICML 2016). PMLR, vol. 48. New York, USA; 2016. p. 1050-1059. (in English).
10. Yun S., Han D., Oh S.J., Chun S., Choe J., Yoo Y. CutMix: regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2019). Seoul, Korea; 2019. p. 6023-6032. (in English).
11. Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., et al. Microsoft COCO: common objects in context. In: Computer Vision – ECCV 2014. Lecture Notes in Computer Science, vol. 8693. Springer, Cham; 2014. p. 740-755. (in English).
12. Wang C.-Y., Yeh I.-H., Liao H.-Y.M. YOLOv9: learning what you want to learn using programmable gradient information. In: Computer Vision – ECCV 2024. Lecture Notes in Computer Science, vol. 15089. Springer; 2025. p. 1-21. (in English).
13. Wang A., Chen H., Liu L., Chen K., Lin Z., Han J., et al. YOLOv10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*. 2024;37:107984-108011. (in English).

14. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, USA; 2016. p. 770-778. (in English).
15. Ren S., He K., Girshick R., Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*. 2015;28:91-99. (in English).
16. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., et al. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017;30:5998-6008. (in English).
17. Zhang S., Wen L., Bian X., Lei Z., Li S.Z. Occlusion-aware R-CNN: detecting pedestrians in a crowd. In: Proceedings of the European Conference on Computer Vision (ECCV 2018). Lecture Notes in Computer Science. Springer, Cham; 2018. p. 637-653. (in English).
18. Zhao Y., Lv W., Xu S., Wei J., Wang G., Dang Q., et al. DETRs beat YOLOs on real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024). Seattle, USA; 2024. p. 16965-16974. (in English).
19. Oks S.J., Jalowski M., Lechner M., Mirschberger S., Merklein M., Vogel-Heuser B., et al. Cyber-physical systems in the context of industry 4.0: a review, categorization and outlook. *Information Systems Frontiers*. 2024;26(5):1731-1772. Springer. (in English).
20. Carvalho T.P., Soares F.A.A.M.N., Vita R., Francisco R.P., Basto J.P., Alcalá S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering*. 2019;137:106024. Elsevier. doi: 10.1016/j.cie.2019.106024. (in English).
21. DeVries T., Taylor G.W. Improved regularization of convolutional neural networks with cutout [Internet]. arXiv preprint arXiv:1708.04552. 2017. Available from: <https://arxiv.org/abs/1708.04552> [accessed 2026 Apr 17]. (in English).
22. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 2012;25:1097-1105. (in English).

*Дата поступления
в редакцию: 16.05.2026*

*Дата принятия
к публикации: 17.07.2026*