



Источник изображения: <https://stock.adobe.com/ru/search/free>

# ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ГЕОТЕХНИЧЕСКИХ ИЗЫСКАНИЯХ

Принята к публикации 13.10.2025

Опубликована 25.11.2025

## МЭЙ Х.

Компания «Институт транспортного проектирования и планирования провинции Шэньси», г. Сиань, провинция Шэньси, Китай  
meihaifeng@163.com

## ЧЖАН В.

Институт искусственного интеллекта Сианьского университета электронных технологий, г. Сиань, провинция Шэньси, Китай  
zhanghuohuo121@163.com

## ГУ Цз.

Институт искусственного интеллекта Сианьского университета электронных технологий, г. Сиань, провинция Шэньси, Китай  
jgu6126@163.com

## АННОТАЦИЯ

Предлагаем вниманию читателей адаптированный перевод статьи китайских исследователей «Применение искусственного интеллекта при геотехнических изысканиях». Эта работа была опубликована в электронном виде в сборнике *Advances in Artificial Intelligence, Big Data and Algorithms* («Достижения в области искусственного интеллекта, больших данных и алгоритмов») международным издательством IOS Press. Статья находится в открытом доступе по лицензии CC BY NC 4.0, которая позволяет копировать и распространять ее, адаптировать, видоизменять и создавать новое, опираясь на нее, но не в коммерческих целях, при указании вида лицензии, типов изменений и ссылки на первоисточник. В данном случае полная ссылка на источник для перевода приведена в конце.

При геотехнических изысканиях широко используются методы динамического зондирования грунтов пробоотборником (SPT) или сплошным наконечником без отбора проб (DPT). Однако в этих случаях при сборе данных могут возникать ошибки, часто связанные с человеческим фактором. Для решения данной проблемы в статье предложено использовать такую разновидность технологии искусственного интеллекта, как усовершенствованный алгоритм YOLOv5 для автоматического подсчета количества ударов молота. Архитектура исходной нейросети YOLOv5 была улучшена следующим образом. Во-первых, была введена функция потерь с фокусировкой для устранения дисбаланса выборки, что обеспечивало более эффективную обработку ударов молотов разных типов. Кроме того, использовалась технология отбора сложных примеров в процессе обучения в режиме онлайн для повышения точности модели (за счет концентрации внимания модели на сложных примерах, наиболее информативных для обучения). Затем улучшенная модель YOLOv5 применялась для выявления ударов молота при испытаниях методами SPT и DPT. Для упрощения обучения модели и ее

работы был создан набор данных (изображений) для детекции молотов разных типов, адаптированный к требованиям конкретных методов геотехнических испытаний. Результаты экспериментов показали высокую эффективность усовершенствованной модели YOLOv5 для выявления и автоматического подсчета ударов молота (на тестовом наборе изображений) на основе использованного набора обучающих данных.

#### КЛЮЧЕВЫЕ СЛОВА:

геотехнические изыскания; полевые динамические испытания грунтов; метод SPT; метод DPT; искусственный интеллект; нейросеть; алгоритм YOLOv5; детекция молота; детекция ударов молота; обучение модели; отбор сложных примеров в режиме онлайн; количество ударов молота; автоматизированный подсчет.

#### ССЫЛКА ДЛЯ ЦИТИРОВАНИЯ:

Мэй Х., Чжан В., Гу Цз. Применение искусственного интеллекта при геотехнических изысканиях (адапт. пер. с англ.) // Геоинфо. 2025. Т. 7. № 3. С. 68–78. DOI:10.58339/2949-0677-2025-7-3-68-78.

# THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN GEOTECHNICAL INVESTIGATION

*Accepted for publication 13.10.2025*

*Published 25.11.2025*

#### MEI H.

Shaanxi Transportation Planning and Design Institute Co. Ltd., Xian, Shaanxi, China  
meihaifeng@163.com

#### ZHANG W.

School of Artificial Intelligence, Xidian University, Xian, Shaanxi, China  
zhanghuohuo121@163.com

#### GU J.

School of Artificial Intelligence, Xidian University, Xian, Shaanxi, China  
jgu6126@163.com

#### ABSTRACT

We present to the readers an adapted translation of the article “The application of artificial intelligence in geotechnical investigation” by Chinese researchers. This work was published electronically in the collection of scientific papers “Advances in Artificial Intelligence, Big Data and Algorithms” by the international publisher “IOS Press”. It is available in open access under the CC BY-NC 4.0 license, which allows copying and redistribution of the article, adaptation, modification, and creation of new works based on it, but not for commercial purposes, provided that the license type, changes made are indicated and the original source is referenced. In our case, the full reference to the original source is provided at the end of the translation.

The standard penetration test (SPT) and dynamic probing test (DPT) are commonly used exploration methods in geotechnical investigation. However, errors can occur during data collection, often attributed to factors such as human error. To mitigate this issue, this paper proposes the utilization of an improved YOLOv5 object detection algorithm, a form of artificial intelligence technology, to automatically count the number of hammer strikes during geotechnical investigations. The proposed approach incorporates several enhancements to the YOLOv5 network architecture. Firstly, a focal loss function is introduced to address sample imbalance, ensuring better handling of different classes of hammer strikes. Additionally, online hard example mining technology is employed to improve model accuracy by focusing on challenging samples that are most informative for training. The improved YOLOv5 model is then applied to detect hammer strikes in SPT and DPT tests. To facilitate training and evaluation, a hammer detection dataset is created, tailored to the specific requirements of geotechnical investigation. Experimental results demonstrate the superior performance of the proposed improved YOLOv5 object detection model on the hammer detection dataset.

#### KEYWORDS:

geotechnical investigation; field dynamic tests of soils; SPT method; DPT method; artificial intelligence; neural network; YOLOv5 algorithm; hammer detection; hammer strikes detection; model training; online hard example mining; number of hammer strikes; automated counting

#### FOR CITATION:

Mei H., Zhang W., Gu J. Primenenie iskusstvennogo intellekta pri geotekhnicheskikh izyskaniyakh (adapt. per. s angl.) [The application of artificial intelligence in geotechnical investigation (adapted translation from English into Russian)] // Geoinfo. 2025. T. 7. № 3. S. 68–78. DOI:10.58339/2949-0677-2025-7-3-68-78 (in Rus.).

## ДЛЯ СПРАВКИ ►

YOLO (You Only Look Once) – это алгоритм на основе глубокого обучения, используемый в области компьютерного зрения для распознавания объектов на изображениях или в видеопотоках. Он был впервые предложен в 2015 году Джозефом Редмоном с соавторами (<https://arxiv.org/abs/1506.02640>) и стал известен своей способностью объединять детекцию и классификацию объектов в один проход нейросети, что значительно ускоряло распознавание по сравнению с предыдущими методами, например R-CNN.

К сегодняшнему дню существует уже 13 официальных версий YOLO. Но их последовательность не всегда означает прямое эволюционное усовершенствование.

В ряду YOLOv1–v5 каждая из версий действительно строилась путем добавления к предыдущему варианту улучшений в архитектуре, точности детекции и скорости работы. Однако YOLOv5 официально не является продолжением серии от создателей оригинального YOLO (Джозефа Редмона с коллегами), поскольку был выпущен компанией Ultralytics (в 2020 году) и сильно отличается от YOLOv1–v4 по структуре и кодовой базе.

В ряду YOLOv6–v13 некоторые версии разрабатывались разными командами и не являются прямым продолжением YOLOv1–v5. То есть они используют концепцию YOLO и общие принципы детекции объектов, но могут сильно различаться по архитектуре и возможностям.

В представленной здесь статье рассматривается улучшенный алгоритм глубокого обучения YOLOv5, который совершенно нельзя рассматривать как устаревший для решения задач геотехнических изысканий. Наоборот, он имеет большой потенциал для использования в этой сфере. YOLOv5 активно применяют в разных странах при полевых динамических испытаниях грунтов, но пока только в процессе научных исследований, для экспериментальных проектов. Эта версия еще не стала массово применяемым инструментом при геотехнических изысканиях для строительства.

В России сейчас тоже активно развиваются технологии искусственного интеллекта и их применение в разных областях, но конкретно для полевых динамических испытаний грунтов методами SPT и DPT таких решений в открытых отечественных источниках пока не было найдено. Эти испытания продолжают проводить с «ручной» фиксацией ударов молота, использованием стандартных датчиков и традиционных статистических методов обработки данных.

Чтобы внедрить в стандартную повседневную практику геотехнических изысканий использование YOLOv5 и других подобных технологий, нужны дополнительные усилия: создание специализированных наборов данных, адаптация моделей к специфике испытаний, интеграция искусственного интеллекта с существующим оборудованием, стандартами испытаний, методами анализа данных, а также налаживание процессов сертификации результатов для инженерных расчетов и др.

ных данных и, как следствие, влиять на результаты геотехнических изысканий.

Благодаря развитию технологий искусственного интеллекта (ИИ) различные методы ИИ уже успешно применяются в разных отраслях. В данной статье предлагается использовать передовые методы ИИ для автоматического подсчета числа ударов молота в ходе полевых динамических испытаний грунтов при геотехнических изысканиях, чтобы исключить неточности, вызванные человеческим фактором. Целью работы является решение соответствующих проблем, влияющих на качество и успешность строительства, путем использования более надежной и эффективной альтернативы ручным подсчетам.

Алгоритмы ИИ для детекции объектов нашли широкое применение в разных сферах, включая интеллектуальные системы видеонаблюдения, автономное вождение и аэрофотосъемку с беспилотных летательных аппаратов. Их цели заключаются в автоматической идентификации конкретных объектов на изображениях или в видеопотоках.

Существующие алгоритмы распознавания объектов можно разделить на два основных класса:

1) *двухстадийные методы детекции объектов с предварительным извлечением областей/зон, представляющих интерес*, например:

- R-CNN (Region-based Convolutional Neural Network) – метод детекции объектов, основанный на сверточных нейронных сетях с предварительным выделением зон интереса [1];
- Fast R-CNN – модификация RCNN с более высокой скоростью работы [2];
- Faster R-CNN – модификация Fast R-CNN с еще более высокой скоростью работы по сравнению с Fast R-CNN [3];

2) *одностадийные/сквозные методы детекции объектов*, например:

- YOLO (You Only Look Once – «вы смотрите только один раз») – алгоритм, обрабатывающий все изображение или каждый кадр видеопотока за один проход нейронной сети и работающий в режиме онлайн [4];
- SSD (Single Shot MultiBox Detector – «одностадийный многооблочный детектор объектов») – в отличие от YOLO использует несколько уровней признаков для детекции объектов различных размеров [5].

Применение алгоритмов детекции объектов при геотехнических изысканиях может обеспечить инженеров ценными инструментами для эффективного и точного сбора необходимой инфор-

## ВВЕДЕНИЕ ►

Геотехнические изыскания включают несколько этапов: постановка задачи, планирование проекта, полевые изыскания, лабораторные испытания, обработка данных, проверка и утверждение результатов исследований, подготовка отчетных документов по проекту. Целью является сбор исчерпывающей информации о площадке будущего строительства, поэтому геотехнические изыскания играют ключевую роль в процессе развития проекта.

При этом одной из основных проблем геотехнических исследований является обеспечение точности данных. В частности, для изыскателей критически важен

точный подсчет числа ударов молота при полевых динамических испытаниях грунтов такими методами, как:

- **SPT** (Standard Penetration Testing) – стандартные пенетрационные испытания (пробоотборником при диаметре зонда 50 мм, забиваемого молотом массой 63,5 кг с высоты 760 мм);
- **DPT** (Dynamic Probing Testing) – динамическое зондирование сплошным наконечником без отбора проб (параметры которого зависят от конкретного типа испытания).

Для «ручного» сбора информации при таких испытаниях присуща неопределенность, что может приводить к ошибкам в статистически обработан-

мации, оценки качества строительства и выявления потенциальных опасностей. Поэтому исследование возможностей их использования при изысканиях имеет большое практическое значение.

Алгоритмы детекции объектов за последние годы достигли значительного прогресса, что привело к их широкому применению в различных сферах. Поскольку разные сценарии и типы объектов обуславливают разные уникальные задачи, для их эффективного решения был разработан ряд детекторов.

Одним из первых инструментов для детекции объектов был алгоритм R-CNN, в котором используется селективный (выборочный) поиск в целях извлечения из изображений областей (зон), представляющих интерес, после чего для каждой из них выполнялись классификация и регрессия ограничивающих рамок.

На основе R-CNN были созданы алгоритмы Fast R-CNN и Faster R-CNN, направленные на дальнейшее повышение эффективности и расширение возможностей детекции объектов. В детектор Fast R-CNN был внедрен слой объединения признаков областей интереса, что повысило точность и скорость детекции за счет эффективной обработки этих зон. В архитектуру детектора Faster R-CNN была встроена нейронная сеть RPN (Region Proposal Network – «сеть предложений областей интереса»), что автоматизировало процесс генерации зон-кандидатов и значительно повысило скорость и общую эффективность работы алгоритма.

Рассмотренная серия алгоритмов R-CNN показала хорошую точность распознавания объектов. Однако они имели такие недостатки, как низкая скорость детекции, сложность в использовании и ограничения при необходимости работы в режиме онлайн, то есть в обработке данных по мере их поступления, например по мере поступления кадров видеопотока, для чего нужна высокая скорость.

Эти проблемы были решены в серии алгоритмов YOLO, которые получили значительно увеличенную скорость работы в режиме онлайн и улучшенную точность детекции объектов.

В алгоритме YOLOv1 [4], предложенном в 2016 году, используются:

- полносвязная сеть, то есть полностью связанные слои, где каждый элемент (узел, нейрон, ячейка) предыдущего слоя соединен с каждым элементом следующего;
- метод детекции объектов на всем изображении (или на каждом кадре ви-

деопотока) за один проход сети, что позволяет быстро обнаруживать множество объектов в режиме онлайн.

Однако относительно неглубокая архитектура сети YOLOv1 (с малым числом слоев) ограничивает общую точность детекции объектов (то есть долю правильных предсказаний от общего количества предсказаний, *assigasy*), делая эту версию менее подходящей для решения задач с высокой точностью положительных предсказаний (то есть долей верных положительных предсказаний от общего числа положительных предсказаний, *precision*).

В последующих версиях, таких как YOLOv2 [6], интегрированы слои нормализации выходных сигналов по пакетам (батчам) и базовые (анкерные, якорные, опорные) рамки для выделения объектов разных размеров, с различными соотношениями сторон и положением на изображении.

В YOLOv3 [7] улучшены скорость и стабильность работы модели за счет реконструкции архитектуры базовой сети DarkNet.

Хотя в серии алгоритмов YOLO постепенно повышалась точность детекции объектов, было замечено постепенное уменьшение скорости их работы. Версии YOLOv4 [8] и YOLOv5 [9] в 2020 году расширили границы как точности, так и скорости, став передовыми алгоритмами в области детекции объектов. На тот момент эти алгоритмы продемонстрировали наивысшую среднюю точность положительных предсказаний (*Average Precision*, AP) на наборе данных для задач компьютерного зрения COCO (*Common Objects in Context* – «общие объекты в контексте» [10]), превзойдя в этом другие алгоритмы, такие как Faster R-CNN и Mask R-CNN [11].

В целом важно находить баланс между точностью и скоростью в зависимости от конкретных требований каждой задачи. Среди множества доступных алгоритмов детекции объектов одним из сильнейших стал YOLOv5 благодаря его очень высокой эффективности. Работа этого алгоритма сочетает высокие скорость и точность, а также возможности точной локализации, что обеспечило ему высокие оценки и широкое применение в последние годы.

В данной работе предлагается расширение возможностей архитектуры YOLOv5 за счет дополнительных улучшений. В частности, авторами были включены функция потерь с фокусировкой и метод отбора сложных примеров в процессе обучения в режиме он-

лайн. Они были применены к задаче детекции ударов молота при полевых динамических испытаниях грунтов методами SPT и DPT, проводимых в рамках геотехнических изысканий. Предлагаемый подход оптимизирует процесс сбора данных благодаря автоматическому подсчету ударов молота во время этих испытаний.

Результаты представленного здесь исследования имеют большое значение для развития технологий применения искусственного интеллекта в инженерных изысканиях, поскольку содержат ценную информацию, на основе которой можно получить рекомендации для будущих исследований и разработок в этой отрасли. Соответствующий вклад данной работы состоит в следующем:

- 1) впервые успешно использован алгоритм детекции объектов в области искусственного интеллекта для распознавания молота при геотехнических изысканиях и предложен метод автоматизированного подсчета числа ударов;
- 2) представлена улучшенная версия YOLOv5, в которой компонент потерь классификации дополнен функцией потерь с фокусировкой, что позволяет модели уделять больше внимания трудно обнаруживаемым и менее многочисленным положительным примерам;
- 3) применен метод отбора сложных примеров в режиме онлайн (*Online Hard Example Mining*, OHEM) для динамической настройки модели, повышения ее точности и улучшения эффективности распознавания труднодетектируемых объектов.

В данной статье рассматривается эволюция серии алгоритмов YOLO, приводится подробное описание предложенного улучшенного метода работы с YOLOv5, представлены и проанализированы результаты экспериментов, подводятся итоги исследований, обсуждаются перспективы их дальнейшего развития.

## ЭВОЛЮЦИЯ YOLO ►

В алгоритме YOLOv1, предложенном Джозефом Редмоном с соавторами, используется единая нейронная сеть для прямого прогнозирования ограничивающих рамок и вероятностей присутствия классов объектов на изображении. Такой одностадийный/сквозной подход позволяет эффективно обрабатывать большое количество данных, обеспечивая высокую скорость и масштабируемость работы алгоритма [12]. Однако YOLOv1 имеет ограничения в точности локализации и может допускать ошибки



в классификации перекрывающихся объектов.

В YOLOv2 применены методы, повышающие общую точность детекции объектов (долю правильных предсказаний от общего количества предсказаний, *assurasy*) и точность положительных предсказаний (долю верных положительных предсказаний от общего числа положительных предсказаний, *precision*). Для улучшения стабильности работы модели использован слой пакетной нормализации (нормализации по батчам), а также применены базовые/анкерные рамки, заданные заранее. Кроме того, YOLOv2 может выполнять детекцию мелких объектов за счет объединения низкоуровневых детализированных признаков с высокоуровневыми сверточными семантическими признаками посредством слияния каналов. Но, несмотря на то что YOLOv2 демонстрирует улучшенную общую точность (*assurasy*) по сравнению с предшествующим YOLOv1, проблемы с локализацией и ошибками классификации объектов в нем остаются.

Алгоритм YOLOv3 еще больше усовершенствован по сравнению с YOLOv2. Благодаря новой базовой сети DarkNet-53, сочетающей элементы архитектур Darknet-19 и ResNet [13], обеспечена совместимость с входными данными разных масштабов и повышена гибкость модели. Также введен метод кластеризации K-means (K-средних) для генерации базовых рамок разных размеров, что улучшает способность обнаруживать мелкие и плотно расположенные объекты.

Алексей Бочковский с соавторами внесли дальнейшие улучшения в YOLOv3 и в 2020 году представили YOLOv4. В этой версии реализована архитектура частичных межстадийных связей (Cross-Stage-Partial, CSP), обеспечивающая эффективную передачу информации между предыдущими и последующими слоями, что дает повышение общей точности и скорости детекции объектов. Для уточнения карт признаков и, соответственно, увеличения точности детекции и улучшения способности распознавать мелкие объекты внедрены такие механизмы внимания, как модуль пространственного внимания (Spatial Attention Module, SAM) и модуль внимания сверточных блоков (Convolutional Block Attention Module, CBAM). Благодаря введению сети агрегации путей/маршрутов (Path Aggregation Network, PANet [14]) еще больше улучшены возможности извлечения признаков, что позволяет модели лучше

распознавать мелкие и множественные объекты. Кроме того, значительно повышена эффективность работы за счет использования таких алгоритмов и модулей, как блоки пирамидального, то есть многоуровневого, пространственного объединения признаков (SPP-blocks, Spatial Pyramid Pooling blocks) и паноптические пирамидальные сети признаков (panoptic feature pyramid networks).

Алгоритм YOLOv5, предложенный в июне 2020 года, отличается от предыдущих компактной архитектурой сети и введением передовых технологий, таких как плавная сигмоидоподобная функция активации (функция типа Swish) и регуляризация с блоковым дропаутом, при которой случайные блоки нейронной карты признаков временно обнуляются для предотвращения переобучения (метод DropBlock), что дополнительно повышает эффективность детектора. Также введена новая стратегия искусственного расширения обучающей выборки, специально разработанная для повышения точности обнаружения мелких объектов.

## УСОВЕРШЕНСТВОВАННАЯ МОДЕЛЬ YOLOv5 ▶

### Структура сети YOLOv5 ▶

Для YOLOv5 существует четыре версии, различающиеся по глубине и ширине нейронной сети: YOLOv5s, YOLOv5m, YOLOv5l и YOLOv5x [15–17].

Для данной работы была выбрана версия YOLOv5l, которая лучше соответствует специфическим требованиям задач, решаемых при геотехнических изысканиях. Она имеет больший размер модели по сравнению с YOLOv5s и YOLOv5m и при этом сохраняет высокую скорость вывода по сравнению с YOLOv5x, что делает ее подходящей для обнаружения объектов в режиме онлайн.

Ключевыми компонентами YOLOv5 являются (рис. 1):

- входной модуль (input end);
- базовая сеть (backbone), отвечающая за извлечение признаков с разных уровней;
- связующий модуль (neck), или шейка, или модуль агрегации признаков;
- выходной модуль предсказаний (head).

Во входном модуле применяется метод Mosaic («Мозаика»), который заключается в том, что четыре изображения случайным образом обрезаются и объединяются в одно с последующими случайным масштабированием и горизонтальным отражением. Эта техника увеличения объема данных значительно обогащает обучающую выборку и повы-

шает устойчивость и обобщающую способность модели.

В качестве основной архитектуры базовой сети в YOLOv5 используется CSPDarknet-53, куда входят модули:

- CSP (Cross Stage Partial – «частичное соединение между стадиями», «модуль частичного перекрестного соединения»);
- CBS (Cross-Stage-Partial + Batch-Norm + SiLU – «свертка + нормализация пакета + активация с помощью сигмоидно-линейной функции (SiLU)»), который состоит из сверточных слоев, слоев нормализации батча/пакета данных и слоев активации с помощью сигмоидно-линейной функции, что эффективно снижает потерю информации и вычислительные затраты при передаче признаков/данных по сети.

Модуль CSP был изначально представлен в базовой сети YOLOv4 и улучшал возможности извлечения признаков за счет установления связей между различными остаточными блоками (блоками с остаточными связями) и ветвями (потоками) признаков, что повышало точность и скорость детекции объектов. В YOLOv5 используются модули CSP двух типов – CSP1 и CSP2.

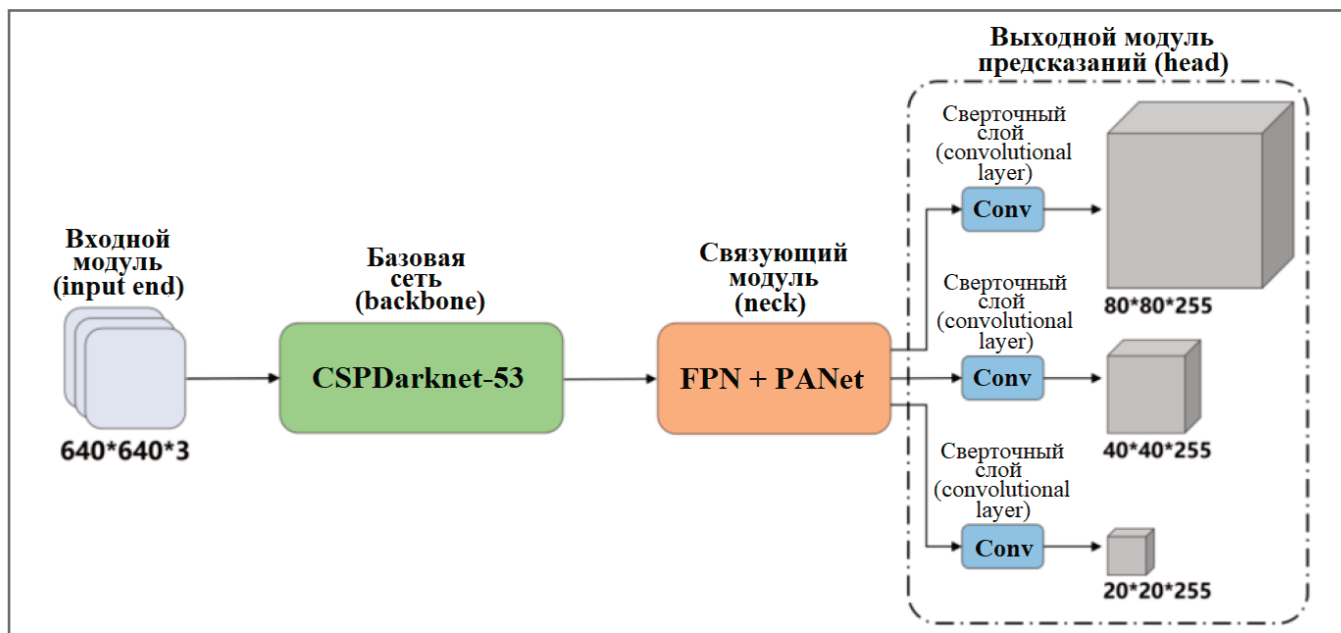
CSP1 имеет два маршрута/пути передачи признаков, один из которых включает модуль CBS и блок с остаточными связями, а другой – только один модуль CBS для согласования (настройки) количества каналов в базовой сети.

CSP2, используемый в связующем модуле, заменяет исходный остаточный блок модулем CBS.

Такая архитектура обеспечивает высокую эффективность при снижении вычислительной сложности, улучшает способность модели представлять признаки и повышает эффективность детекции объектов.

В связующий модуль (модуль агрегации признаков) в YOLOv5 входят: модуль SPPF; модуль CSP; комбинация модулей FPN и PAN.

Модуль SPP (Spatial Pyramid Pooling – «пирамидальное/многоуровневое пространственное объединение признаков») повышает способность сети обрабатывать изображения различных размеров за счет использования операции «объединение по максимуму» (max pooling), то есть операции формирования подвыборки признаков, благодаря которой на выходе сохраняется максимальное значение для каждой области входного тензора. Это позволяет сети работать с изображениями произвольных масштабов. А модуль



**Рис. 1.** Общая архитектура YOLOv5. Названия модулей: CSPDarknet-53 (Cross Stage Partial Darknet-53) – улучшенная архитектура базового модуля/сети по сравнению с Darknet-53, используемая в YOLO начиная с четвертой версии; FPN (Feature Pyramid Network – «сеть пирамид признаков») в комбинации с PANet (Path Aggregation Network – «сеть агрегации путей/маршрутов», «сеть объединения информационных потоков»). Под схемой входного модуля числами указан размер входного тензора изображения: ширина изображения в пикселях\*высота изображения в пикселях\*количество каналов (обычно RGB, то есть красный, зеленый и синий). В схеме выходного модуля признаков числами для каждого из трех уровней указан размер выходного тензора, то есть ширина карты признаков\*высота карты признаков\*количество каналов, где ширина и высота измеряются в ячейках карты признаков, а число каналов соответствует предсказаниям для всех базовых рамок на данном уровне

SPPF (Spatial Pyramid Pooling Fast – «ускоренное пирамидальное/многоуровневое пространственное объединение признаков») является облегченной версией SPP, которая обеспечивает более высокую скорость работы при сохранении эффективности обработки признаков.

Модуль CSP (Cross Stage Partial – «частичное соединение между стадиями», «модуль частичного перекрестного соединения») был подробнее рассмотрен ранее.

Модуль FPN (Feature Pyramid Network – «сеть пирамид признаков») объединяет признаки разных уровней, применяя операцию увеличения разрешения (upsampling – «масштабирование вверх»), то есть пространственное разрешение карты признаков увеличивается на каждом уровне, чтобы выполнить объединение с признаками более низкого уровня и в итоге создать богатую и согласованную карту признаков. В свою очередь, модуль PAN (Path Aggregation Network – «сеть агрегации путей/маршрутов», «сеть объединения информационных потоков») обеспечивает передачу информации о расположении признаков/объектов снизу вверх по пирамиде признаков (начиная с нижних уровней). Такое сочетание позволяет одновременно задействовать детальную информацию (низкоуровневые призна-

ки) и семантическую (смысловую) информацию высокого уровня, способствуя распознаванию разнообразных объектов. Это усиливает способность сети обучаться и извлекать более широкий спектр признаков.

В выходном модуле прогнозирования для вычисления потерь при локализации ограничивающих рамок используется функция потерь *GloU* (Generalized Intersection over Union – «обобщенное пересечение по объединению», то есть доля площади пересечения областей по отношению к площади их объединения). Эта функция является модифицированной метрикой *IoU*. В ней учитывается не только доля площади пересечения предсказанной и базовой рамок, как в *IoU*, но и расстояние между их центрами и соотношение сторон. Благодаря учету этих дополнительных факторов функция потерь *GloU* обеспечивает более точный расчет потерь (ошибку) ограничивающих рамок, что повышает точность детекции объектов моделью.

### Обучение сети с выбором сложных примеров в режиме онлайн ▶

Набор данных (изображений) по молотам был создан специально для этого исследования. Он на 90% состоял из простых примеров. Однако преобладание такого рода примеров при обучении

модели ограничивает ее способность эффективно выявлять более сложные случаи. Для решения данной проблемы применяется метод обучения сети с отбором сложных примеров в режиме онлайн (Online Hard Example Mining, ОНЕМ [19]). Этот метод динамически регулирует процесс обучения, выбирая трудные примеры на основе величины потерь и включая их в последующие итерации (шаги обновления весов модели с обработкой одного пакета/батча данных, то есть одной мини-выборки). Активное использование сложных примеров позволяет модели лучше учиться и повышает эффективность детекции труднораспознаваемых объектов.

Как показано на рисунке 2, применение метода ОНЕМ на основе YOLOv5 включало следующие шаги:

- 1) для каждой мини-выборки входные данные подавались в сеть, которая генерировала предсказанные координаты ограничивающих рамок и вероятности принадлежности каждого объекта к каждому классу;
- 2) рассчитывалась функция потерь как мера различий между предсказанными и базовыми ограничивающими рамками с учетом координат рамок и вероятностей принадлежности объектов к классам;
- 3) сортировались значения потерь и выбиралась подвыборка рамок с наиболь-

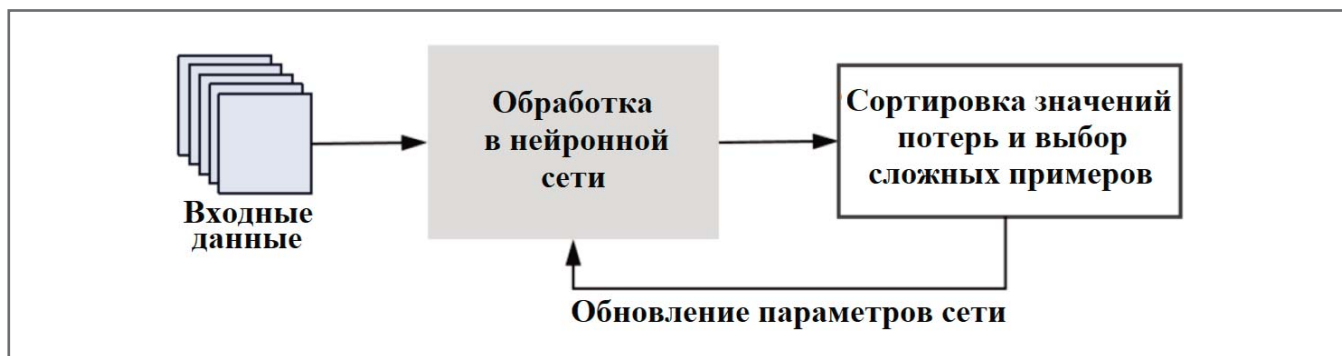


Рис. 2. Схема процесса обучения модели с использованием метода ОНЕМ

шими потерями в качестве трудных примеров для дальнейшего обучения; 4) эти трудные примеры использовались для обновления параметров сети с помощью алгоритма обратного распространения ошибки [20].

Применение метода ОНЕМ в процессе обучения позволяет модели сосредоточиться на сложных примерах, благодаря чему она лучше усваивает их отличительные черты, что приводит к повышению точности и надежности детекции ею объектов.

### Функция потерь с фокусировкой ▶

Функция потерь в YOLOv5 может быть выражена следующим образом:

$$L = \lambda_{coord} L_{coord} + \lambda_{obj} L_{obj} + \lambda_{cls} L_{cls} + \lambda_{giou} L_{giou}, \quad (1)$$

где  $L_{coord}$  – потеря точности локализации объекта;  $L_{obj}$  – потеря уверенности в обнаружении объекта;  $L_{cls}$  – потеря точности классификации;  $L_{giou}$  – потеря точности локализации с учетом площади пересечения и охвата рамок (GloU-потеря). Всё это гиперпараметры, используемые для балансировки вклада каждой составляющей функции потерь.

Из-за значительного дисбаланса между положительными (есть удар молота) и отрицательными (нет удара молота) примерами в собственном наборе данных использование стандартной функции потерь, вычисляемой на основе кросс-энтропии (метрики расхождения между истинными и предсказанными вероятностями) привело бы к тому, что модель сосредотачивалась бы на многочисленных отрицательных примерах. А это могло бы негативно сказаться на эффективности детекции положительных примеров. Поэтому была введена функция потерь с фокусировкой, то есть функция, ориентированная на трудные примеры (Focal Loss Function [21]), которая решает проблему несбалансиро-

ванных выборок, динамически регулируя веса отрицательных примеров с помощью настраиваемого параметра и тем самым стимулируя модель фокусировать внимание на положительных примерах. Когда указанный настраиваемый параметр равен нулю, функция потерь с фокусировкой сводится к функции потерь, рассчитываемой на основе бинарной кросс-энтропии (при предсказаниях для двух классов). Однако с увеличением значения данного параметра эта функция эффективно снижает веса легко классифицируемых примеров, тем самым направляя внимание модели на сложные положительные примеры, которых относительно мало [22]. Формула функции потерь с фокусировкой (Focal Loss,  $FL$ ) записывается следующим образом:

$$FL(p_i) = -\alpha_i (1 - p_i)^\gamma \lg(p_i),$$

$$\text{где } p_i = \begin{cases} p, & \text{если } y = 1 \\ (1 - p) & \text{иначе} \end{cases}, \quad (2)$$

где  $p_i$  – вероятность, предсказанная моделью для положительных примеров в наборе данных (для истинных ударов молота);  $p$  – вероятность принадлежности к положительному классу, которую модель присваивает конкретному примеру; значение  $y=1$  соответствует положительному классу,  $y=-1$  – отрицательному классу (фону);  $\alpha_i$  – весовой коэффициент, равный  $\alpha$  для положительных примеров и  $(1-\alpha)$  для отрицательных;  $(1-p_i)^\gamma$  – модулирующий множитель/коэффициент, нужный для того, чтобы подавить вклад легких примеров, в которых модель уже уверена, и оставить значимым или даже подчеркнуть вклад трудных для классификации примеров, в которых модель мало уверена или не уверена;  $\gamma$  – настраиваемый (задаваемый) параметр фокусировки, который управляет силой этой модуляции.

Функция потерь с фокусировкой предназначена для компенсации дисба-

ланса между положительными и отрицательными примерами из набора данных. В улучшенной модели YOLOv5, представленной в данной работе, эта функция применяется для вычисления потерь классификации  $L_{cls}$  по следующей формуле:

$$L_{cls} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C [y_i = c] (1 - p_{i,c})^\gamma \lg(p_{i,c}), \quad (3)$$

где  $N$  – количество обучающих примеров;  $C$  – функция локализационных потерь;  $y_i$  – истинный класс  $i$ -го примера,  $p_{i,c}$  – предсказанная моделью вероятность того, что  $i$ -й пример принадлежит классу  $c$ .

Использование функции потерь с фокусировкой делает обучение модели детекции объектов более эффективным, тем самым улучшая точность обнаружения сложных положительных примеров. Кроме того, эта функция дополнительно повышает способность модели к обобщениям в процессе обучения, поскольку позволяет динамически корректировать веса примеров.

## РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ И ИХ АНАЛИЗ ▶

### Формирование набора данных ▶

Авторами представленной работы был сформирован специализированный набор данных, предназначенный для обнаружения молотов при динамических испытаниях грунтов методами SPT и DPT в процессе геотехнических изысканий, с целью предоставления исходных данных для дальнейших исследований. Он включал изображения молотов двух разных типов с учетом ряда влияющих факторов, таких как различные условия освещения и ракурсы съемки, что позволяло имитировать сложности, характерные для реальных условий. На рисунке 3 приведены примеры использованных типов молотов (типов 1 и 2).



Для формирования указанного набора данных на площадке проведения динамических испытаний грунтов методами SPT и DPT применялась съемка высокоразрешающими видеокамерами. Затем полученные видеозаписи обрабатывались путем извлечения отдельных кадров, которые впоследствии отбирались для получения множества изображений. Для обеспечения точной разметки применялся специальный инструмент под названием LabelImg, с помощью которого аннотировалось наличие молота на каждом изображении. Всего набор данных включал 2400 обучающих и 800 тестовых (для проверки работы модели) изображений, при этом на каждом из них присутствовал только один детектируемый молот.

### Состав экспериментальной системы ►

В состав экспериментальной системы входили следующие компоненты:

- 1) *аппаратная часть*: использовалась видеокарта NVIDIA GeForce RTX 1060s с 6 ГБ видеопамяти;
- 2) *программная среда*: эксперименты проводились в операционной системе Windows 11 с применением такой интегрированной среды разработки на языке программирования Python, как PyCharm 2022.1, и такой платформы с набором инструментов для параллельных вычислений на видеокартах (GPU) компании NVIDIA, как CUDA 11.8;
- 3) *среда (фреймворк) глубокого обучения*: использовалась библиотека для глубокого обучения PyTorch 2.0.0 (которая предоставляет инструменты для построения, обучения и оценки нейросетей).

Во время обучения модели использовался адаптивный алгоритм оптимизации Adam (Adaptive Moment Estimation – «Адаптивная оценка моментов»), который использует скользящие оценки среднего и дисперсии градиента. Размер пакета данных (батча) составлял 32 примера/изображения, и обучение модели проводилось в ходе 100 итераций. Параметр фокусировки  $\gamma$  в функции  $FL$  (см. формулу (2)) был задан равным 1,5.

### Показатели оценки ►

В данной работе для оценки качества и эффективности модели использовалась средняя точность положительных предсказаний, усредненная по всем классам детектируемых объектов (mean Average Precision,  $mAP$ ), в отличие от Average Precision,  $AP$  – средней точно-



Рис. 3. Примеры типов молотов (1 и 2)

сти положительных предсказаний для одного конкретного класса). Для определения этой метрики нужны следующие показатели.

- $TP$  (True Positive) – истинно положительные предсказания, то есть количество истинно положительных примеров, которые модель правильно отнесла к положительным (распознала удары молота там, где они действительно были).  $TP$  соответствует числу ограничивающих рамок, определенных моею, как положительные, которые достаточно точно совпадают с базовыми/истинными рамками (то есть с долей площади пересечения предсказанной и базовой рамок  $IoU$  не меньше заранее заданного порога).
- $FP$  (False Positive) – ложноположительные предсказания, то есть количество отрицательных примеров, которые модель ошибочно классифицировала как положительные («увидела» удары молота там, где их в действительности не было).
- $FN$  (False Negative) – ложноотрицательные предсказания, то есть число положительных примеров, которые модель ошибочно определила как отрицательные (не распознала реальные удары молота).
- $TN$  (True Negative) – истинно отрицательные предсказания, то есть количество отрицательных примеров, которые модель правильно определила как отрицательные («не увидела» ударов молота там, где их действительно не было).

На основе этих показателей можно вычислить такие метрики, как точность положительных предсказаний  $P$  (precision) и полнота  $R$  (recall) детекции молота для каждого класса.

Точность положительных предсказаний  $P$  определяется как отношение количества истинно положительных сра-

батываний ( $TP$ ) к сумме истинно положительных и ложноположительных оценок ( $TP+FP$ ).

Полнота  $R$  определяется как отношение количества истинно положительных предсказаний ( $TP$ ) к сумме истинно положительных и ложноотрицательных срабатываний ( $TP+FN$ ).

Если речь идет о точности  $P$  и полноте  $R$  для распознавания моделью конкретного класса детектируемых объектов  $k$ , то используют соответствующие обозначения с нижним индексом  $k$ :

$$P_k = \frac{TP_k}{TP_k + FP_k}, \quad (4)$$

$$R_k = \frac{TP_k}{TP_k + FN_k}, \quad (5)$$

Могут быть вычислены разные значения точности  $P_k$  и полноты  $R_k$  для разных порогов уверенности модели. Кривая «точность  $P_k$  – полнота  $R_k$ » (по горизонтальной оси –  $R_k$ , по вертикальной –  $P_k$ ) может наглядно представить взаимосвязь между этими двумя показателями при разных порогах. Площадь под этой кривой соответствует средней точности положительных предсказаний  $P_k$  для конкретного класса  $k$  детектируемых объектов ( $AP$ , Average Precision).

Набор данных, самостоятельно собранный авторами, включал два типа/класса молотов. Средняя точность положительных предсказаний, усредненная по обоим этим классам объектов ( $mAP$ ), дает обобщенную оценку качества работы модели при распознавании рассматриваемых молотов:

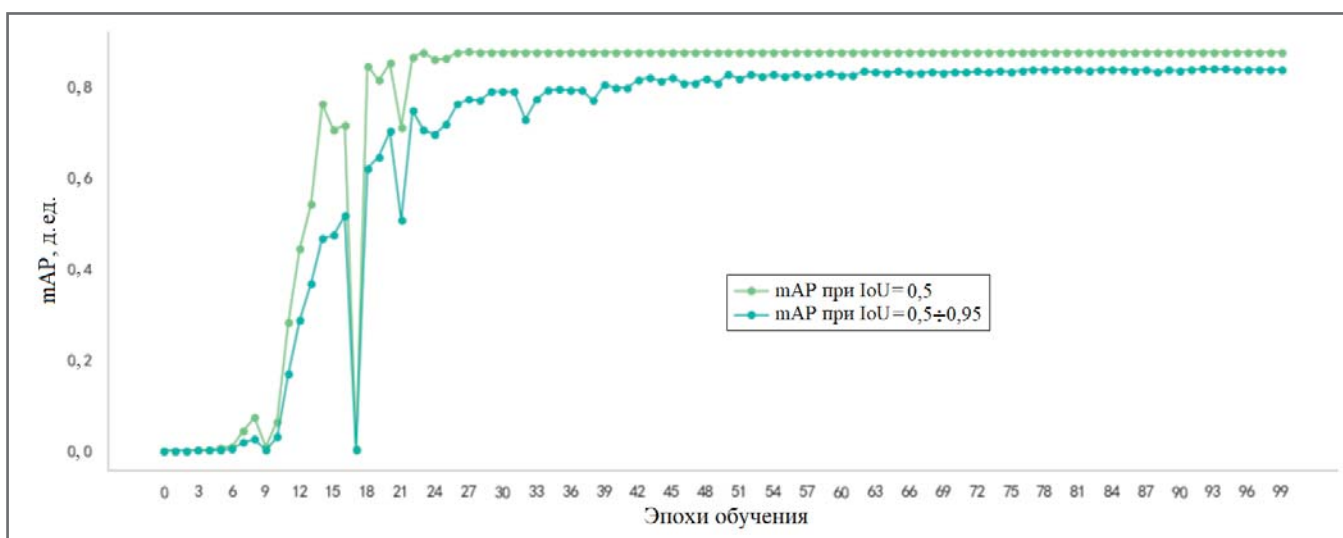
$$mAP = \frac{1}{2} \sum_{k=1}^2 AP_k, \quad (6)$$

где  $AP_k$  – средняя точность положительных предсказаний при детекции конкретного класса  $k$ , то есть молота типа 1 или молота типа 2 (см. рис. 3)





**Рис. 4.** Примеры результатов детекции молотов. Полученные значения средней точности положительных предсказаний, усредненной по обоим классам молотов ( $mAP$ ), указаны над соответствующими фотографиями, сделанными на площадке испытаний



**Рис. 5.** Динамика качества работы модели на тестовом наборе данных

### Результаты экспериментов ►

В данном исследовании для обучения модели распознаванию молотов использовалась описанная выше модифицированная версия алгоритма YOLOv5. Оценка работы модели проводилась на тестовом наборе данных с использованием различных порогов коэффициентов пересечения предсказанной и базовой рамок ( $IoU$ ), а именно: 0,50; 0,55; 0,60; 0,65; 0,70; 0,75; 0,80; 0,85; 0,90; 0,95. Для комплексной оценки качества (надежности и стабильности) модели на

различных уровнях перекрытия спрогнозированных и базовых ограничивающих рамок применялась метрика  $mAP$ . На рисунке 4 представлены примеры результатов детекции на тестовом наборе данных.

После 25 эпох обучения (полных проходов модели по всему обучающему набору данных с пересчетом потерь,  $mAP$  и других метрик) качество работы модели на тестовом наборе данных вышло на стабильный уровень, как показано на рисунке 5. Следует отметить,

что итоговая модель дала высокое значение  $mAP$ , составившее 0,8753 при пороге  $IoU$ , равном 0,5. А величина  $mAP$ , усредненная для диапазона порога  $IoU$  от 0,5 до 0,95, составила 0,8383. Эти результаты говорят о высокой точности распознавания моделью ударов молота.

После детекции положения молота определение того, совершался ли удар (было ли это движением вниз), выполнялось на основе изменения координаты центра молота. Критерий был сле-

дующим: если смещение центра молота вниз превышало 10 единиц по сравнению с предыдущим кадром и это происходило не менее 4 раз за 8 последовательных кадров, то фиксировался один удар (одно нисходящее движение). Чтобы исключить повторный подсчет одного и того же удара, после каждого зарегистрированного удара вводилась пауза длиной в 15 кадров, в течение которой подсчет не выполнялся. Процесс считался завершенным, если в течение 10 минут не было зафиксировано ни одного движения вниз.


Экспериментальные испытания подтвердили эффективность предложенного метода: он может обеспечить точный автоматический подсчет числа ударов молота во время динамических испытаний грунтов методами SPT и DPT при геотехнических изысканиях, позволяя

избегать ошибок в регистрации данных, возникающих при ручном подсчете.

### **ЗАКЛЮЧЕНИЕ И ПЛАНЫ НА БУДУЩЕЕ ►**

В представленном исследовании был усовершенствован алгоритм YOLOv5. В него были введены функция потерь с фокусировкой и метод обучения на сложных примерах в режиме онлайн для использования в процессе обучения нейросети. Этот модифицированный алгоритм был обучен и протестирован на собранном авторами наборе данных по динамическим испытаниям грунтов методами SPT и DPT, в который входили изображения молотов (кадры видеопотоков), что позволило создать надежную модель для детекции этих объектов при подсчете количества ударов.

Полученные экспериментальные результаты продемонстрировали высокую эффективность улучшенного алгоритма YOLOv5. Средняя точность положительных предсказаний, усредненная по двум исследованным классам молотов (mean Average Precision, mAP), достигла величины 0,838.

Поскольку благодаря постоянно проводимым геотехническим изысканиям продолжают поступать новые видеоданные с площадок испытаний, существующий набор данных по молотам постоянно пополняется и уточняется. Это позволит проводить дальнейшие исследования по оптимизации модели детекции молотов при полевых динамических испытаниях грунтов, а также увеличить охват и эффективность применения методов распознавания разных объектов при геотехнических изысканиях. 

### **Источник для перевода ► (Source for the translation) ►**

Mei H., Zhang W., Gu J. The application of artificial intelligence in geotechnical investigation // *Frontiers in Artificial Intelligence and Applications*. Vol. 373. *Advances in Artificial Intelligence, Big Data and Algorithms*. Amsterdam, Netherlands: IOS Press, 2023. P. 263–273. DOI:10.3233/FAIA230818.

### **Список литературы, использованной авторами переведенной статьи ► (References used by the authors of the translated paper) ►**

1. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation // *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*. 2014. P. 580–587.
2. Girshick R. Fast R-CNN // *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. 2015. P. 1440–1448.
3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks // *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV 2015)*. 2015. P. 91–99.
4. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*. 2016. P. 779–788.
5. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A.C. SSD: Single shot multibox detector // *Proceedings of the 2016 European Conference on Computer Vision (ECCV 2016)*, 2016. P. 21–37.
6. Redmon J., Farhadi A. YOLO9000: Better, faster, stronger // *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017. P. 6517–6525.
7. Redmon J., Farhadi A. YOLOv3: An incremental improvement // *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*. 2018. P. 6517–6525.
8. Bochkovskiy A., Wang C.Y., Liao H.Y.M. YOLOv4: Optimal speed and accuracy of object detection // *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*. 2020. P. 10934–10944.
9. Wong B., AbdSalam R., Wong S.H. YOLOv5: A better, faster, stronger object detector // *Proceedings of the 2020 International Conference on Neural Information Processing (ICONIP 2020)*. 2020. P. 1078–1090.
10. Lin T.Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick C.L. Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science* // *Proceedings of the 2014 European Conference on Computer Vision (ECCV 2014)*. Part V. *Lecture Notes in Computer Science*. Switzerland: Springer International Publishing, 2014. Vol. 8693. P. 740–755.
11. He K., Gkioxari G., Dollar P., Girshick R. Mask R-CNN // *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV 2017)*. 2017. P. 2980–2988.
12. Liu X., Wang Z., He Y., Liu Q. Research on small target detection based on deep learning // *Tactical Missile Technology*. 2019. Vol. 1. P. 100–107.

13. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). 2016. P. 770–778.
14. Liu S., Qi L., Qin H., Shi J., Jia J. Path aggregation network for instance segmentation // Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018). 2018. P. 8759–8768.
15. Lisi W., Yu Z. Glass bottle mouth defect detection based on YOLOv5 // Yangtze River Information Communication. 2023. Vol. 36. № 1. P. 9–11 (in Chinese).
16. Jiang L., Cui Y. Small object detection based on YOLOv5 // Computer Knowledge and Technology. 2021. Vol. 17. № 26. P. 131–133.
17. Song Y.X., Zhao Y., Zhang J.Y., Zhu W.P., Yang Z.H., Zhang Q. Design of sitting posture monitoring system based on YOLOv5 // FEMT (Frontiers of Electronic Materials). 2023. Vol. 19. № 8. P. 22–25.
18. Lin T.Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). 2017. P. 936–944.
19. Shrivastava A., Gupta A., Girshick R. Training region-based object detectors with Online Hard Example Mining // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). 2016. P. 761–769.
20. Shi F., Qiu Z., Han Q., Li J., Qian H., Xiang W. Improved faster R-CNN algorithm based on variable weight loss function and hard example mining module // Computer and Modernization, 2020. Vol. 8. P. 56–62 (in Chinese).
21. Lin T.Y., Goyal P., Girshick R., He K., Dollar P. Focal loss for dense object detection // Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018). 2018. P. 2980–2988.
22. Huang J., Zhang G. A Review of object detection algorithms based on deep convolutional neural networks // Computer Engineering and Applications. 2020. Vol. 56. № 17. P. 12–23.

# Независимый электронный журнал ГеоИнфо

С 2022 года журнал «ГеоИнфо»  
выходит в формате \*PDF.  
4 выпуска в 2025 году



WWW.GEOINFO.RU