

Исследование подхода «обнаружение объектов» в задаче фрагментации горных пород на открытых карьерах

К.И. Решетников¹, М.В. Ронкин¹, С.В. Поршнева¹

¹ ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»,
620002, Россия, Екатеринбург, ул. Мира, д. 19

Аннотация

В горнодобывающих работах на открытых карьерах активно используются взрывные технологии. При этом появляется задача оценки качества взрывных работ, которая определяется размерами фрагментов горной породы, полученных в результате взрывов. В связи с этим возникает задача оценки числа фрагментов горной породы и их размеров (задача фрагментации). В настоящее время популярны подходы к решению таких задач на основе систем компьютерного зрения с использованием нейронных сетей семантической или экземплярной сегментации. При этом оказывается, что для их работы требуется существенное привлечение компьютерных ресурсов. В связи с этим использование альтернатив – алгоритмов быстрого обнаружения объектов на цифровых изображениях взорванной породы является актуальным. В статье изучены особенности использования нейронных сетей глубокого обучения с архитектурой YOLO, которая, как ожидалось, будет иметь более высокую скорость обработки видеoinформации. На основе проведенного исследования обоснован выбор в качестве базовой архитектуры использовать YOLOv7x. Для обучения нейронных сетей выбранной архитектуры был использован размеченный авторами набор данных, составленный из цифровых фотографий фрагментов горной породы, образовавшихся в результате взрыва в открытом карьере. Полученные результаты позволили обосновать выбор в качестве метрики оценки качества взрывных работ геометрические размеры описанного прямоугольника вокруг выделенного на изображении фрагмента породы.

Ключевые слова: фрагментация, компьютерное зрение, обнаружение объектов, глубокое обучение нейронных сетей, сверточные нейронные сети.

Цитирование: Решетников, К.И. Исследование подхода «обнаружение объектов» в задаче фрагментации горных пород на открытых карьерах / К.И. Решетников, М.В. Ронкин, С.В. Поршнева // Компьютерная оптика. – 2024. – Т. 48, № 2. – С. 272-281. – DOI: 10.18287/2412-6179-CO-1382.

Citation: Reshetnikov KI, Ronkin MV, Porshneva SV. Investigation of an object-detection approach for estimating the rock fragmentation in the open-pit conditions. Computer Optics 2024; 48(2): 272-281. DOI: 10.18287/2412-6179-CO-1382.

Введение

Добыча полезных ископаемых является сложным техническим процессом, зависящим от условий залегания горных пород. Одним из наиболее распространенных способов добычи ресурсов в горнодобывающей промышленности является открытый способ, который основан на измельчении горной массы с помощью взрыва и ее последующей обработке на горнообогатительной фабрике [1] с целью разделения добываемого вещества и горной породы. Следует отметить, что размеры горной породы, добытой в открытом карьере, напрямую определяют экономические затраты на последующее ее дробление на горнообогатительной фабрике. В этой связи возникает задача автоматизированного контроля размера фрагментов горной породы.

Примеры изображений рабочих площадок открытых карьеров после проведения взрывных работ показаны на рис. 1а–в. Из рисунков видно, что задача автоматизированного анализа цифровых изображений является нетривиальной ввиду наличия большого

числа фрагментов различных размеров и зависит от условий освещенности места съемки. Очевидно, данная задача декомпозируется на две отдельные задачи, решаемые последовательно. Первая из них – это задача нахождения фрагментов горной породы на цифровом изображении; а вторая – оценка максимальных размеров найденных фрагментов горной породы на цифровом изображении [2, 3].

Сегодня одним из наиболее популярных способов решения задачи обнаружения фрагментов горной породы на цифровом снимке и оценивания их геометрических размеров являются системы компьютерного зрения, основанные на использовании нейронных сетей [3, 4, 5, 6], которые могут самостоятельно извлекать признаки объектов и проводить их разделение на цифровом снимке [7]. Анализ вариантов для решения данной задачи позволил выделить несколько подходов, проиллюстрированных для рассматриваемой задачи на рис. 2а–в. Пример обнаружения (детекция) фрагментов горной породы на изображении представлен на рис. 2а. Подход основан на использовании нейронных сетей с глубоким обучением, например

YOLOv3 [8] и YOLOv5 [9]. Пример семантической сегментации изображения фрагментов горной породы представлен на рис. 2б, из которого видно, что при использовании данного подхода выбираются пиксели с заданными свойствами без их отнесения к конкретному фрагменту породы [2, 10, 11]. В этой связи дополнительно требуется привлекать методы разделения экземпляров сегментации, применение которых, в свою очередь, существенно замедляет решение задачи. Наиболее часто для семантической сегментации изображений используется нейронная сеть с архитектурой U-Net [12]. Пример экземплярной сегментации фрагментов горной породы на изображении представлен на

рис. 2в, из которого видно, что при использовании данного подхода находятся не только пиксели с заданными свойствами, но и устанавливается взаимно-однозначное соответствие между каждым пикселем и изображением фрагментов горной породы, которому он принадлежит. В настоящее время среди методов решения обсуждаемой задачи наиболее популярной является сеть с многоступенчатой архитектурой Mask R-CNN, предложенная в 2017 г. [13, 14]. Например, нейронная сеть с архитектурой Mask R-CNN была использована для определения распределения размера фрагментов горной породы на цифровых изображениях открытого карьера Nui Phao, Вьетнам [3].



Рис. 1. Примеры изображений открытых карьеров после проведения взрывных работ: а) Баженовское месторождение, Россия; б) Железная руда, Индия [2]; в) Вольфрамовый карьер Nui Phao [3]

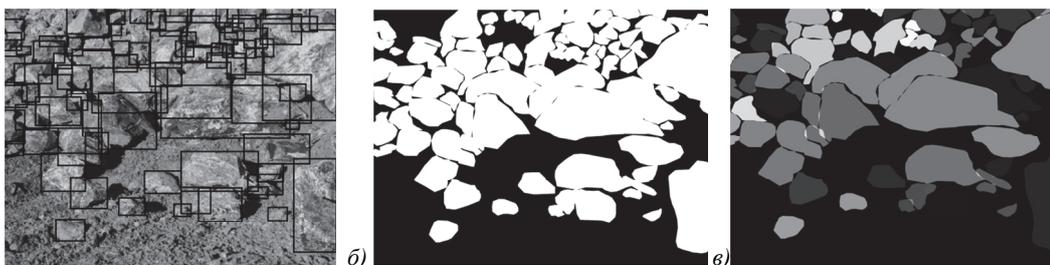


Рис. 2. Иллюстрация изображений в наборе данных для задач: а) обнаружения объектов; б) семантической сегментации; в) объектной сегментации

Для каждого определенного фрагмента породы на цифровом изображении определяется ее максимальный размер. Для этого в классических решениях могут быть использованы простые геометрические подходы на основе диаметра размера вписанного круга или максимального размера вписанного эллипса [15, 16]. Также в ряде современных работ в качестве максимального размера фрагмента горной породы принимают значение максимальной стороны прямоугольника, описанного вокруг фрагмента горной породы (метод описанного прямоугольника) [17]. Анализ вычислительной сложности перечисленных методов определения максимального размера фрагментов горной породы показывает, что наименьшую вычислительную сложность имеет последний метод. В этой связи представляется целесообразным провести проверку возможности его использования в задаче анализа изображений горной породы, получаемой в результате взрывов в открытых карьерах. В статье обсуждаются результаты исследования возможности использования метода описанных прямоугольников в задаче анализа цифровых изображений результатов проведения взрывных работ в открытых карьерах.

1. Обзор известных методов фрагментации

Проведенный анализ современных публикаций в области исследования позволил выделить три основных метода для решения задачи нахождения фрагментов горной породы на цифровом изображении.

Применение семантической сегментации в области горнодобывающей промышленности широко представлено в литературе. Например, в статье [10] предлагается рассматривать задачу определения размера фрагментов пород как задачу сегментации контура (как задачу семантической сегментации). Авторы [11] предлагают слегка модифицированную версию архитектуры Res-U-Net для определения размера руды как в условиях конвейерной ленты, так и в условиях открытого карьера. Кроме того, авторы [10] протестировали комбинацию сети семантической сегментации и водораздела, который показывает сопоставимую точность для изображений на конвейерной ленте в высоком разрешении. Однако такой подход не обеспечивает высокой точности на изображениях открытого карьера.

Второй способ – это задача сегментации экземпляров, которая в компьютерном зрении обычно рас-

считается как задача обнаружения объектов с семантической сегментацией экземпляров внутри ограничивающей рамки [18]. Наиболее распространенной архитектурой для экземплярной сегментации является Mask R-CNN [13]. Эта архитектура имеет многоступенчатую структуру, состоящую из кодировщика карт признаков (такого как ResNet [19]), этапа предварительной оценки предложений набора интересных областей без указания их классов (т.н. анхоров или привязок) и головной части, которая состоит из классификации и регрессии параметров ограничивающей рамки и семантической сегментации для каждого региона-кандидата, полученного на предыдущем этапе. После работы сети алгоритм не максимального подавления (NMS) применяется к окончательным регионам-кандидатам для окончательного отбора результатов [13]. Способ оценки фрагментации с использованием архитектуры Mask R-CNN был предложен в работе [3]. Авторы [3] достигли 92 % для метрики средней точности (average precision) [3]. Также работа показывает, что алгоритмы глубокого обучения работают лучше, чем классические подходы компьютерного зрения для обработки крупномасштабных изображений рабочих площадок. Также в [3] отмечается связь точности классических алгоритмов с внешними условиями, такими как дневное время, достаточная освещенность, сухая погода и так далее. Почти аналогичные результаты были показаны в статье [20], где сравниваются традиционные подходы и глубокое обучение для крупномасштабных изображений рабочих площадок в карьере. Авторы [20] заявляют о преимуществах Mask R-CNN для крупномасштабных изображений. Многие современные авторы утверждают, что традиционные подходы страдают от большого разброса размеров фрагментов породы и большого количества перекрывающихся экземпляров и качества изображения для крупномасштабных изображений, например, в [3, 5, 20].

Третий способ нахождения горных пород на цифровых изображениях – использование методов обнаружения объектов на основе нейронных сетей YOLO. Семейство архитектур «You Look Only Ones» (YOLO) разрабатывалось как попытка решения задачи одноэтапного обнаружения объектов в режиме реального времени. Подход YOLO основан на идее, что входное изображение делится на ячейки одинакового размера, каждая из которых соответствует одному выходному вектору. Такой вектор может включать в себя одну или несколько привязок (анхоров, anchors) с предсказаниями размера ограничивающей рамки, оценкой класса и оценкой объектности. Под оценкой объектности понимается классификация объект/фон в заданной области изображения. Итоговые выходные ячейки, содержащие объекты, определяются с помощью алгоритма не максимального сжатия (NMS) [21].

В рамках наших предыдущих работ был предложен алгоритм оценки производительности горнодо-

бывающих работ на открытых карьерах [22]. Алгоритм включал решение задачи фрагментации. Для этого использовалась архитектура на основе Mask R-CNN. В работе [23] предлагается использовать архитектуру YOLOv5 [9] для определения фрагментов породы и оценки их размера. В этой работе доказывалось, что мы можем рассматривать проблему оценки фрагментации как проблему обнаружения объектов. Благодаря использованию YOLOv5 решение проблем фрагментации изображений открытых карьеров может быть ускорено в 10 раз без существенной потери точности по сравнению с Mask R-CNN.

Развитие семейства архитектур YOLO производилось путем экспериментов с архитектурой и особенностями тренировки как в области кодировщика признаков, так и в области головной и шейной частей. В архитектуре YOLOv2 были предложены только сверточные слои и использованные сети предложений регионов (RPN) в качестве генератора привязок [24]. В архитектуре YOLOv3 в шейной части было предложено использование пирамидального подхода извлечения карт признаков [8]. Архитектура YOLOv4 включила в себя множество специальных приемов обучения («Bag of Tricks»), таких, например, как подбор гиперпараметров, методов аугментации и т.д. и приемов улучшений архитектуры («Bag of Specials») [25]. Архитектура YOLOv5 была создана в фреймворке PyTorch [9] с использованием некоторых небольших улучшений четвертой версии. Версия YOLOv5 стала отраслевым стандартом в области обнаружения объектов. Также отметим, что на основе версий YOLO 3-5 были созданы многие другие работы в области обнаружения объектов в реальном времени [26, 27].

Среди других версий подхода YOLO отметим следующие. В архитектуре YOLOv7 [28] авторы предлагают новый сверточный блок (расширенные сети агрегации эффективного уровня) как компромисс между быстрыми вычислениями и простым масштабированием. Новейшая архитектура YOLOv8 [29] включает обновленный блок свертки, основанный на идеях, аналогичных YOLOv4, но с гибким коэффициентом масштабирования. Также в головной части использован подход без привязок (без т.н. анхоров). Достоинства этого подхода были доказаны, например, в версии YOLOv6 [30]. Такое решение позволяет значительно снизить вычислительную сложность моделей без существенной потери точности. Отметим, что в современных реализациях архитектуры YOLO v5,7,8 также включают встроенную реализацию необязательной головной части архитектуры для сегментации объектов [3, 9, 29].

Таким образом, результаты анализа литературы в рассматриваемой области показывают, что большинство результатов прикладного глубокого обучения относятся к хорошо зарекомендовавшим себя подходам, таким как U-Net и Mask-R-CNN. Более того, ав-

торы не уделяют большого внимания вычислительной оптимизации предлагаемых алгоритмов (моделей). Однако хорошо известно, что вышеупомянутые архитектуры требуют сравнительно много вычислительных ресурсов и плохо подходят для высокопроизводительных вычислений на конечных устройствах в реальном масштабе времени [13]. Однако решение указанных проблем может быть найдено в использовании таких подходов, как, например YOLO.

2. Фрагментация как задача обнаружения объектов

Собранный набор данных содержит 220 изображений участков открытых карьеров после проведения взрывных работ при добыче асбестосодержащей породы. Изображения собраны в Баженовском месторождении (Свердловская область, г. Асбест). Оптическая система для получения изображений описана в статье [22]. Основная часть изображений была получена на открытых карьерах в течение 5 экспериментов в осеннее и зимнее время года. В каждом эксперименте было от 3 до 5 разрабатываемых площадок открытого карьера, выбранных геологической службой компании как наиболее репрезентативные. Для каждого открытого участка карьера было сделано несколько снимков разных его частей. Каждое изображение содержит порядка 20 – 80 фрагментов горных пород разного размера. Для всех полученных изображений была произведена разметка с помощью инструмента CVAT [31]. Пакет CVAT позволяет произвести объектную (instance) сегментацию, такая разметка также включает получение обрамляющих рамок (bounding boxes), используемых в задачах обнаружения объектов, и возможность создания маски семантической сегментации. Примеры изображений собранных данных приведены на рис. 2а–в для разного типа разметки. Оригинальный размеченный набор данных может быть найден по ссылке DOI:10.17632/pfdbpfygh [21].

Проведен анализ разметки изображений описанного набора данных. Анализ показывает, что большинство рассмотренных объектов имеют ориентацию, позволяющую считать их максимальные размеры соответствующими размерам ограничивающей рамки. Под термином «максимальный размер» тут понимается максимальное расстояние между двумя точками на границе контура отдельного камня. Эта закономерность позволяет предположить возможность заменить задачу оценки распределения размеров фрагментов по результатам сегментации объектов на задачу обнаружения объектов. Отметим, что при этом задача обнаружения объектов имеет гораздо меньшую вычислительную сложность, чем задача сегментации объектов. Для сравнения разметки и ограничивающих рамок были использованы следующие метрики: максимальный размер семантической маски (т.е. максимальное расстояние между двумя

точками на границе контура отдельного фрагмента); ширина ограничивающей рамки; высота ограничивающей рамки; значение максимального размера ограничивающей рамки (т.е. значение максимального размера, выбранного из ширины и высоты ограничивающей рамки); среднее арифметическое значение высоты и ширины ограничивающей рамки; диагональ ограничивающей рамки.

Проведен корреляционный анализ параметра «максимальный размер семантической маски» с перечисленными параметрами, определяющими ограничивающую рамку. Отметим, что параметр маски – максимальный размер выбран, потому что, по нашему мнению, в наибольшей степени характеризует распределение размеров фрагментов. Другие популярные параметры, например, площадь фрагмента, могут быть недостаточно репрезентативными. Фрагмент породы может иметь небольшую площадь на изображении в силу перекрытия другими фрагментами. Для всех размеченных фрагментов был подсчитан коэффициент корреляции Пирсона для выборок выбранных параметров. Значение коэффициентов корреляции для максимального размера рамки – 0,995; высоты рамки – 0,957; ширины рамки – 0,936; диагонали рамки – 0,993; для среднего значения ширины и высоты – 0,990. Видно, что существует линейная связь между максимальным размером маски и размерами ограничивающей рамки. При этом максимальное значение 0,995 критерия Пирсона оказывается в зависимости между максимальным размером семантической маски и максимальным размером ограничивающей рамки.

Далее были исследованы статические свойства выборок составленных значений максимальным размеров семантических масок $L = \cup_{i=1}^N l_i$ каждого из фрагментов породы, найденных на цифровых изображениях и соответствующих максимальным размерам ограничивающих рамок $R = \cup_{i=1}^N r_i$, где N – количество всех фрагментов породы, найденных на всех изображениях используемого набора данных ($N=14087$). На первом этапе была проверена гипотеза о принадлежности исследуемых выборок L и R некоторой генеральной совокупности с помощью критерия типа Колмогорова–Смирнова (гипотеза H_0). Для этого использовалась функция stats.ks_2samp из пакета scipy. Оказалось, что гипотеза H_0 отвергается на уровне значимости 0,05, значение статистики 0,058 и $p_{value} = 2,73 \cdot 10^{-21}$. В этой связи для последовательностей L и R были вычислены аппроксимации функции плотности вероятности методом Розенблатта–Парзена [32] ρ_R и ρ_L соответственно. Далее были вычислены математические ожидания и дисперсии данных выборок:

$$\bar{x} = \int_{x_{min}}^{x_{max}} xp(x) dx, \quad D(x) = \int_{x_{min}}^{x_{max}} (x - \bar{x})^2 \rho(x) dx, \quad (1)$$

где x_{min} , x_{max} – минимальное и максимальное значения, ρ – аппроксимация Розенблатта–Парзена соответствующей выборки, $\sigma = \sqrt{D}$.

Аппроксимации Розенблатта–Парзена для соответствующих плотностей распределения представлены на рис. 3. Оказалось, что $\bar{L} = 0,0995$, $D_L = 6,398$ и $\bar{R} = 0,0924$, $D_R = 5,782$. Таким образом, оказывается, что показатель L отличается от R не более чем на 7%, а показатель D_L отличается от D_R не более 9%, в то же время на каждом изображении содержится в среднем по 64 фрагмента камней. При этом время вычисления \bar{R} на порядок меньше аналогичной величины \bar{L} . Аналогичные вычисления были проведены для каждого изображения из датасета. При этом оказалось, что максимальное отклонение для среднего значения не превышает 17% (7 фрагментов камней), а для дисперсии – не более 30% (5 фрагментов камней). Последующий анализ показал, что высокая ошибка измерений связана с вариативностью изображений, используемых в наборе данных. Наибольшая ошибка возникает на тех изображениях, которые были получены при фокусировке камеры на отдельном камне. Таким образом, для получения уменьшения погрешности измерений \bar{R} и D_R целесообразно использовать короткие фокусные расстояния в сравнении с размером области, в которой находятся горные породы.

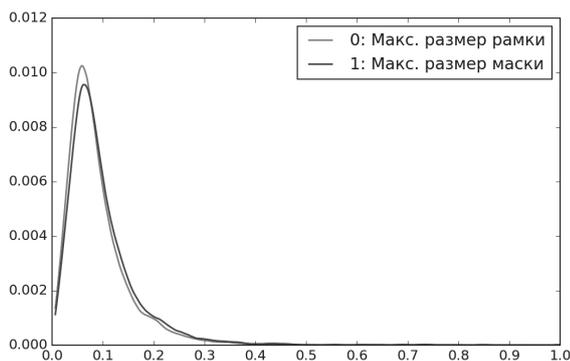


Рис. 3. Оценка плотности функций распределения методом Розенблатта–Парзена

Эмпирически обнаружено, что точность оцениваемых характеристик фрагментов породы на цифровых изображениях зависит от расстояния до участка съемки и количества объектов, которые находятся на изображении. Чем больше объектов на цифровом изображении, тем лучше средняя оценка размера каждого объекта. Отметим также, что при изменении фокусного расстояния видеокамеры можно ожидать увеличения числа фрагментов породы на цифровом изображении, что, в свою очередь, будет приводить к более точной оценке среднего размера фрагментов. В то же время увеличение размера фотографируемого участка карьера может приводить к ухудшению качества обнаружения мелких фрагментов породы и, соответственно, к увеличению погрешности аппроксимации Розенблатта–Парзена плотности распределения случайных последовательностей, составленных из размеров максимальной стороны прямоугольников, описанных вокруг каждого из фрагментов горной породы. Следовательно, можно ожидать, что су-

ществует некоторое оптимальное расстояние, зависящее от расстояния до места съемки, фокусного расстояния объектива видеокамеры и числа фрагментов породы, находящихся на цифровых изображениях. Нахождение данного расстояния является направлением дальнейшего исследования.

3. Исследование подхода YOLO

В качестве модели для решения задачи обнаружения объектов на изображениях было предложено использовать архитектуру семейства YOLO. Как показал анализ литературы, семейство данных моделей является наиболее современным и популярным решением для задач обнаружения объектов в «реальном масштабе времени» (см., например, [33]).

Целью экспериментов было сравнить качество восстановления распределения размеров фрагментов горной породы на изображениях и скорость обработки изображений на собранном наборе данных. В качестве архитектур для обучения были выбраны YOLOv5x, YOLOv6l, YOLOv7x и YOLOv8x [3, 9, 29, 30]. Выбранные модели имеют наибольшее число внутренних параметров для каждой серии моделей YOLO, а также являются наиболее современными архитектурами в семействе YOLO. При проведении экспериментов изображения исходного набора данных сжимались до размеров 512×512 пикселей. Это необходимо, так как исходный набор изображений имеет слишком высокое разрешение по сравнению с предполагаемым при разработке архитектур. Выбранный размер изображений соответствует рекомендациям к использованию выбранных нейронных сетей. Также ожидается, что снижение размера изображения позволяет оптимизировать работу моделей для использования на портативных устройствах. Также было произведено сравнение с широко популярной в литературе Mask R-CNN моделью. Была обучена Mask R-CNN модель с архитектурой кодировщика ResNeXt-101-FPN из фреймворка Detectron2 [34].

Для обучения и тестирования моделей использовался подход кросс-валидации методом k-блоков (k-fold). Этот принцип в проведенных экспериментах сводился к разбиению всей выборки на 4 не пересекающихся блока данных. Для каждых 3 из полученных 4 блоков производилось обучение (тренировочная часть), тогда как четвертая часть выбиралась в качестве валидационной части. Обучение проводилось 4 раза со сменой валидационной части. Оценка качества работы модели вычислялась как арифметическое среднее соответствующих значений для каждого из экспериментов. Модели оценивались с помощью таких метрик, как точность (Precision), полнота (Recall) и средняя точность (mean Average Precision, mAP), а также оценивалось время работы модели на одно изображение. В отношении используемых метрик использованы определения из работы [35].

Обучение моделей семейства YOLO проводилось со стандартными параметрами, рекомендуемыми разработчиками [29]. В т.ч. скорость обучения – 0,01, число эпох обучения – 100, размер батча равен 8. Обучение модели проводилось на системе с видеокартой Tesla V100-SXM2 32Гб, Intel Xeon Silver 4216, 94Гб RAM. Во время обучения для обучающих изображений использовались стандартные приемы динамической аугментации из пакета YOLO, в том числе mosaic augmentation. Параметры аугментаций: HSV-оттенок – 0,015; HSV-насыщенность – 0,7; HSV-яркость – 0,4; масштабирование изображения – 0,5; вероятность поворота изображения слева направо – 0,5. Для тестирования использовалась наилучшая модель, полученная на эпохе обучения с наилучшим значением взвешенного показателя mAP (mean Average Precision) на валидационном наборе. Взвешенный показатель mAP вычисляется согласно:

$$mAP = mAP(0,5) \cdot 0,1 + mAP(0,5:0,95) \cdot 0,9, \quad (2)$$

где $mAP(0,5)$ – значение, подсчитанное на валидационном наборе для IOU, равное 0,5, и $mAP(0,5:0,95)$ – усредненное значение mAP при значениях IoU, равном 0,5, 0,55, ..., 0,9, 0,95.

Для обучения Mask R-CNN использовались следующие параметры: скорость обучения – 0,001, число эпох – 3000, размер батча – 2, размер изображения был изменен на 800×800. Для обучения использовался набор аугментаций, отражающий пространственные преобразования и изменение цвета со следующими параметрами из фреймворка Detecton2: случайный масштаб в диапазоне [0,7...1,3] и сдвигом 0,1, случайное изменение яркости в диапазоне [0,7...1,5], случайное изменение контраста в диапазоне [0,7...1,4], случайное изменение насыщенности в диапазоне [0,7...1,4], случайный поворот на угол в диапазоне [–90,90].

В табл. 1 показаны полученные результаты метрик: точность (P), полнота (R), интерполированная средняя точность (mean Average Precision) для значения пересечения объектов IOU 0,5, усредненное значение mAP при значениях IoU, равных 0,5; 0,55; ... 0,9; 0,95, а также скорость обработки одного изображения в миллисекундах (t). Метрики, указанные в табл. 1, являются средними значениями на валидационных подвыборках для блоков метода k-fold.

Табл. 1. Результаты метрик работы моделей YOLO на валидационных данных

Модель	P	R	mAP@ 0,5	mAP@ 0,5..0,95	t, мс
YOLOv8x	0,71	0,66	0,72	0,47	13,1
YOLOv7x	0,70	0,67	0,72	0,46	10,9
YOLOv6l	0,70	0,66	0,63	0,43	15,1
YOLOv5x	0,70	0,65	0,70	0,44	19,0
Mask R-CNN	-	-	0,51	0,30	189,1

Из табл. 1 видно, что наилучшие значения метрик точности показывает модель YOLOv8x. Время обработки одного изображения для YOLOv7x меньше всех. Наихудшие результаты точности получены для модели YOLOv6l. Также для этой модели получено максимально высокое время обработки изображения. Модель Mask R-CNN значительно проигрывает моделям семейства YOLO по точности обнаружения. Интерполированная средняя точность в задаче обнаружения для модели Mask R-CNN равна 51%. При использовании Mask R-CNN в режиме сегментации не наблюдается улучшения качества в обнаружении фрагментов. Разница значения mAP составляет около 1%. Время обработки одного изображения для Mask R-CNN больше в среднем в 10–17 раз.

Типичные результаты работы моделей для валидационных данных приведены на рис. 4 для моделей YOLOv8x и YOLOv7x. Представленные результаты получены для значения параметра минимального порога уверенности сети (conf_thres) 0,25. Визуальный анализ показывает сопоставимость результатов обнаружения для всех моделей. Различия в результатах видны в случаях относительно небольших фрагментов горных пород.

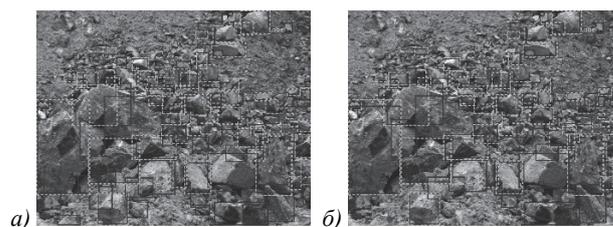


Рис. 4. Результат обнаружения фрагментов на валидационном изображении для модели YOLO8× слева и YOLO7× справа. Предсказания сетей показаны с помощью сплошной обрамляющей рамки

4. Исследование точности в задаче фрагментации

Проведена оценка точности результатов восстановления распределения размеров фрагментов горной породы на изображениях для анализируемых моделей. Оценка проведена путем сравнения схожести распределений предсказанного максимального размера фрагментов с максимальным размером ограничивающих рамок для разметки. Для проведения указанных оценок проведен анализ кумулятивных функций распределения максимального размера для предсказанных и размеченных ограничивающих рамок по каждому валидационному блоку. Примеры результатов для валидационных блоков под номером один и три показаны на рис. 5.

Из визуального анализа данного рисунка видно, что для всех моделей максимальный размер ограничивающих рамок близок к максимальному размеру для размеченных ограничивающих рамок. Результат модели YOLOv8x визуально наиболее близок к разметке максимального размера ограничивающих ра-

мок. Для модели YOLOv6l полученный максимальный размер имеет наибольшее отклонение от максимального размера разметки.

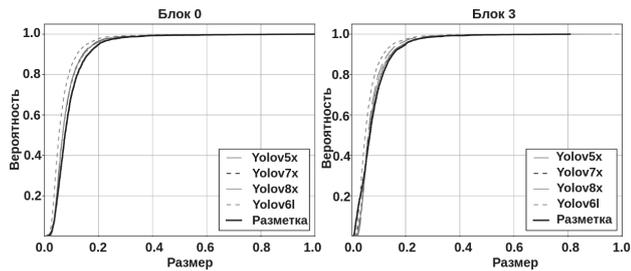


Рис. 5. Иллюстрации кумулятивных функций распределения восстановленного размера фрагментов скальной породы на изображениях на примере валидационных блоков №0 и №3

Для численной оценки схожести распределений проведена оценка расстояния Васерштейна. Расстояние Васерштейна можно интерпретировать как минимальную стоимость энергии, которую необходимо совершить, чтобы переместить массу из одного распределения, чтобы заполнить другое распределение. Если F и G – функции распределения для ν и μ соответственно, тогда расстояние Васерштейна определяется как [36]:

$$d_W(\mu, \nu) = \left(\int_{-\infty}^{\infty} |F(x) - G(x)|^2 dx \right)^{\frac{1}{2}} dt^{\frac{1}{2}} = \left(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt \right)^{\frac{1}{2}} \quad (3)$$

где F^{-1} и G^{-1} – квантильные функции. Чем меньше расстояние между двумя распределениями, тем выше их близость.

Табл. 2. Расстояние Васерштейна между распределениями для валидационных блоков

Модель	Расстояние Васерштейна				Среднее
	Блок				
	1	1	2	3	
YOLOv8x	0,016	0,014	0,016	0,011	0,014
YOLOv7x	0,009	0,013	0,014	0,006	0,011
YOLOv6l	0,024	0,029	0,029	0,019	0,025
YOLOv5x	0,010	0,015	0,010	0,012	0,012

Расстояние Васерштейна между распределениями для обученных моделей для каждого валидационного набора данных показано в табл. 2. Из табл. 2 видно, что для моделей YOLOv7x и YOLOv5x вычисленное расстояние Васерштейна между распределениями для предсказанного и размеченного размера минимально. Для модели YOLOv6l имеет место наибольшее расстояние между распределениями для предсказанного и размеченного максимального размера. Таким образом, согласно расстоянию Васерштейна, наиболее близкое распределение размеров фрагментов камней к разметке удастся получить с помощью моделей YOLOv7x (0,011) и YOLOv5x (0,012).

5. Использование расстояния Васерштейна для повышения качества работы моделей

Анализируемый параметр – расстояние Васерштейна может быть использован для подбора параметра минимального порога уверенности сети (confidence threshold) во время обучения. Обученная нейронная модель для каждого обнаруженного объекта определяет координаты обрамляющей рамки и вероятность нахождения объекта внутри рамки t_x, t_y, t_h, t_w, t_o [25]. Предсказания, у которых вероятность нахождения объекта ниже порога $t_o < conf_{thres}$, отбрасываются из окончательного результата. Используя расстояние Васерштейна, можно подобрать параметр минимального порога уверенности сети, при котором расстояние между распределением максимального размера для предсказанных и размеченных ограничивающих рамок будет минимально.

Рассмотрим зависимость расстояния Васерштейна между распределениями максимального размера от параметра минимального порога уверенности сети. Для этого был проварьирован параметр минимального порога уверенности сети от 0,2 до 0,8 с шагом 0,05. Для каждого значения параметра было подсчитано расстояние Васерштейна между распределением предсказанного максимального размера ограничивающей рамки и распределением максимального размера ограничивающей рамки для разметки. На рис. 6 показаны значения расстояния Васерштейна для выбранных значений параметра минимального порога уверенности сети для тренировочных выборок.

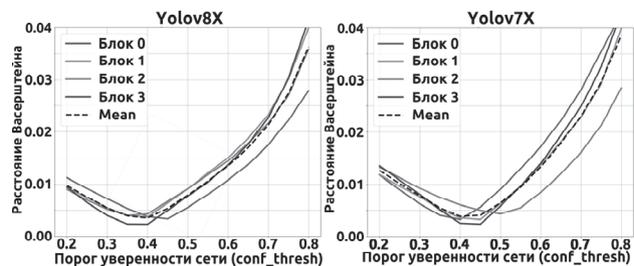


Рис. 6. Расстояние Васерштейна между распределением разметки и распределением, полученным обученной сетью для моделей YOLOv8x и YOLOv7x. Каждая графическая линия обозначает отдельный обучающий блок

Из рис. 6 видно, что при значении минимального порога уверенности сети в пределах от 0,4 до 0,5 удастся получить минимальное значение расстояния между распределениями. Иными словами, для значений минимального порога уверенности сети в интервале от 0,4 до 0,5 распределение размеров, полученное моделью, наиболее близкое к распределению размеров для разметки. Таким образом, использование расстояния Васерштейна позволяет выбрать значение минимального порога уверенности сети, лучше подходящее для оценки фрагментации горной породы, то есть для задачи оценки распределения размеров фрагментов горной породы, образованных в результате взрывных работ.

6. Обсуждение результатов

Полученные результаты показывают возможность проведения оценки фрагментации горных пород на открытых карьерах при помощи архитектур быстрого обнаружения объектов. Полученные точности восстановления распределения (расстояние Васерштейна 0,011) и результаты визуального анализа свидетельствуют о соответствии полученного и заданного распределений. Однако в данной работе мы не рассматриваем критерии оценки качества работ для специалистов геологических служб. Полагается, что дальнейшая обработка распределений выходит за рамки статьи.

Новизна работы заключается в разработанном методе оценки фрагментации и его апробации, а также в апробации предложенного оригинального метода оценки качества решения задач фрагментации при помощи анализируемых моделей. Анализируемый метод основан на использовании быстрых архитектур обнаружения объектов. Суммарно метод может быть описан следующим образом. На этапе тренировки модель YOLO обучается на размеченных изображениях результатов взрывных работ. В качестве метрики используется mAP. Для тренированной модели на валидационном наборе данных подбирается наилучшее значение порога уверенности для алгоритма NMS. На тестовых данных производится оценка фрагментации пород как оценка распределения максимального размера (среди длины и ширины ограничивающей рамки).

В качестве основной модели для данной задачи рекомендуется взять модель YOLOv7x. Данная модель показала наилучшие показатели по соотношению между временем работы и точностью восстановления распределений, а также сопоставимые метрики точности. Также вклад работы в область исследований заключается в том, что впервые проведен анализ задачи фрагментации на открытых карьерах как задачи обнаружения объектов. Предложен и апробирован метод оценки фрагментации при помощи быстрых архитектур глубокого обучения для обнаружения объектов. Проанализированы наиболее современные достижения в области обозначенных архитектур, в том числе основные современные модели семейства YOLO. А также предложено использование расстояния Васерштейна в качестве метрики оценки качества восстановления распределений в данном типе задач.

В качестве ограничений работы мы хотим отметить следующие. В работе мы использовали только модели глубокого обучения семейства YOLO. Мы полагаем, и это подтверждается литературным исследованием, что подход на основе моделей сверточных нейронных сетей глубокого обучения является основным в решении задач компьютерного зрения. Все рассмотренные архитектуры этого семейства относятся к 2020–2023 годам. Известно, что подход

YOLO является основным при решении указанных задач в настоящее время [27]. Более того, отметим, что выбор именно подхода быстрого решения задач обнаружения объектов важен с практической точки зрения, так как именно он позволяет реализовывать итоговые модели на низкопроизводительных конечных устройствах в реальном масштабе времени. Также отметим, что целью данной статьи не был выбор наименьшей из моделей каждого семейства. Эта задача, по нашему мнению, представляет в первую очередь практический интерес и должна быть связана только с конкретным выбором конечного устройства. Более того, рекомендованная модель YOLOv7 имеет только одну версию, которая сопоставима по числу параметров с остальными проанализированными моделями. По нашему мнению, указанные ограничения не снижают общую научную ценность работы.

Заключение

В работе исследован подход обнаружения объектов с использованием глубокого обучения нейронных сетей для задачи оценки фрагментации на открытых карьерах. Для оценки изображений фрагментации горных пород, образованных в результате взрыва, использовались сверточные нейронные сети семейства YOLO. Проведено сравнение моделей YOLO с помощью интерполированной средней точности (mAP) на собранном наборе изображений участков открытых карьеров после взрывных работ. Для обучения и тестирования моделей использовался подход кросс-валидации методом четырех блоков. Согласно mAP лучший показатель демонстрирует модель YOLOv8x со значением 72% и YOLOv7x со значением 72%. Показано превосходство моделей YOLO над архитектурой Mask R-CNN по точности и скорости обработки при обнаружении фрагментов горной породы по результатам взрыва. Предложен подход к выбору лучшей модели для оценки распределения размеров фрагментов на основе расстояния Васерштейна. Таким образом, данная модель может быть рекомендована в качестве основной при дальнейших исследованиях данной темы. Также предложен метод выбора порогового уровня уверенности моделей в рабочем режиме на основе расстояния Васерштейна. Такой выбор позволяет наилучшим образом регулировать число обнаруживаемых камней на изображениях для последующего восстановления распределений. Согласно оценке расстояния по Васерштейну лучшей моделью, восстанавливающей распределение размеров фрагментов, является YOLOv7x.

Благодарности

Исследование выполнено за счет совместного гранта Российского научного фонда и Правительства Свердловской области № 22-21-20051, <https://rscf.ru/project/22-21-20051/>.

References

- [1] Luzin V. Complex studies of longitudinal-fiber chrysotile asbestos of the Bazhenov deposit [In Russian]. Source: <http://resources.krc.karelia.ru/krc/doc/publ2011/miner_tech_ocenka_118-126.pdf>.
- [2] Shrivastava S, Bhattacharjee S, Debasis D. Segmentation of mine overburden dump particles from images using Mask R CNN. *Sci Rep* 2023; 13: 2046. DOI: 10.1038/s41598-023-28586-0.
- [3] Vu T, Bao T, Hoang Q, Drebenstedt C, Hoa P, Thang H. Measuring blast fragmentation at Nui Phao open-pit mine, Vietnam using the Mask R-CNN deep learning model. *Mining Technology* 2022; 130(4): 232-243. DOI: 10.1080/25726668.2021.1944458.
- [4] Mohammad B, Mohammad A, Farhang S, Farzad S, Sadjad M. A new framework for evaluation of rock fragmentation in open pit mines. *J Rock Mech Geotech Eng* 2019; 11(2): 325-336.
- [5] Bamford T, Esmacili K, Schoellig A. A deep learning approach for rock fragmentation analysis. *Int J Rock Mech Min Sci* 2021; 145: 104839.
- [6] Jung D, Choi Y. Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. *Minerals* 2021; 11(2): 148.
- [7] Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT Press; 2016.
- [8] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. *arXiv Preprint*. 2018. Source: <<https://arxiv.org/abs/1804.02767>>.
- [9] ultralytics/yolov5. 2020. Source: <<https://github.com/ultralytics/yolov5>>.
- [10] Wang W, Li Q, Xiao C, Zhang D, Miao L, Wang L. An improved boundary-aware U-Net for ore image semantic segmentation. *Sensors* 2021; 21(8): 2615. DOI: 10.3390/s21082615.
- [11] Li H, Pan C, Chen Z, Wulamu A, Yang A. Ore image segmentation method based on U-Net and Watershed. *Comput Mater Contin* 2020, 65(1), 563-578.
- [12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In *Book: Medical image computing and computer-assisted intervention – MICCAI 2015. Pt III*. Cham, Heidelberg: Springer International Publishing Switzerland; 2015: 234-241.
- [13] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2017 IEEE Int Conf on Computer Vision (ICCV) 2017: 2980-2988. DOI: 10.1109/ICCV.2017.322.
- [14] Ramesh S, Kumar V. A review on instance segmentation using mask R-CNN (December 24, 2020). *Proc Int Conf on Systems, Energy & Environment (ICSEE)* 2021: 1-4. DOI: 10.2139/ssrn.3794272.
- [15] CivilNode. Gold Size 2.0 Download. Source: <<https://civilnode.com/download-software/10159053855788/gold-size-20>>.
- [16] Fitzgibbon A, Pílu M, Fisher RB. Direct least square fitting of ellipses. *IEEE Trans Pattern Anal Mach Intell* 1999; 21(5): 476-480. DOI: 10.1109/34.765658.
- [17] Li M, Wang X, Yao H, Saxén H, Yu Y. Analysis of particle size distribution of coke on blast furnace belt using object detection. *Processes* 2022; 10(10): 1902.
- [18] Gu W, Bai S, Kong L. A review on 2D instance segmentation based on deep neural networks. *Image Vis Comput* 2022; 120: 104401.
- [19] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 770-778.
- [20] Schenk F, Tscharf A, Mayer G, Fraundorfer F. Automatic muck pile characterization from UAV images. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2019; IV-2/W5: 163-170.
- [21] Redmon J, Divvala S, Girshick R, Farhadi F. You only look once: Unified, real-time object detection. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 779-788.
- [22] Zyuzin V, Ronkin M, Porshnev S, Kalmykov A. Automatic asbestos control using deep learning based computer vision system. *Appl Sci* 2021; 11(22): 10532. DOI: 10.3390/app112210532.
- [23] Mendeley Data. openpits asbestos. Source: <<https://data.mendeley.com/datasets/pfdbfpfygh/2>>.
- [24] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2017: 6517-6525. DOI: 10.1109/CVPR.2017.690.
- [25] Bochkovskiy A, Wang C, Liao H. Yolov4: Optimal speed and accuracy of object detection. *arXiv Preprint*. 2020. Source: <<https://arxiv.org/abs/2004.10934>>.
- [26] Diwan T, Anirudh G, Tembhurne JV. Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimed Tools Appl* 2022; 82: 9243-9275.
- [27] Jiang P, Ergu D, Liu F, Cai Y, Ma B. A review of Yolo algorithm developments. *Procedia Comput Sci* 2022; 199: 1066-1073.
- [28] Wang C, Bochkovskiy A, Liao H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv Preprint*. 2022. Source: <<https://arxiv.org/abs/2207.02696>>.
- [29] Jocher G. What is YOLOv8? The ultimate guide. 2023. Source: <<https://blog.roboflow.com/whats-new-in-yolov8/>>.
- [30] Li C, et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv Preprint*. 2022. Source: <<https://arxiv.org/abs/2209.02976>>.
- [31] CVAT. Open Data Annotation Platform. Source: <<https://cvat.ai/>>.
- [32] Racine JS. An introduction to the advanced theory and practice of nonparametric econometrics. Cambridge University Press; 2019.
- [33] Real-time object detection. 2023. Source: <<https://paperswithcode.com/task/real-time-object-detection>>.
- [34] facebookresearch/detectron2. 2019. Source: <<https://github.com/facebookresearch/detectron2>>.
- [35] Hui J. mAP (mean Average Precision) for object detection. Source: <<https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>>.
- [36] Ramdas A, Garcia N, Cuturi M. On wasserstein two sample testing and related families of nonparametric tests. *arXiv Preprint*. 2015. Source: <<https://arxiv.org/abs/1509.02237>>.

Сведения об авторах

Решетников Кирилл Игоревич, 1995 года рождения, аспирант ИРИТ-РТФ УрФУ по направлению 09.06.01 «Информатика и вычислительная техника». Область научных интересов включает компьютерное зрение, машинное обучение и искусственный интеллект. E-mail: ki.reshetnikov@urfu.ru

Ронкин Михаил Владимирович, 1987 года рождения, кандидат технических наук, доцент ИРИТ-РТФ УрФУ. Область научных интересов включает компьютерное зрение, методы обработки сигналов, машинное обучение и искусственный интеллект. E-mail: m.v.ronkin@urfu.ru

Поршнев Сергей Владимирович, 1961 года рождения, доктор технических наук, профессор, профессор Учебно-научного центра «Информационная безопасность» ИРИТ-РТФ, УрФУ. Область научных интересов включает методы обработки и анализа данных. E-mail: s.v.porshnev@urfu.ru

ГРНТИ: 28.23.00

Поступила в редакцию 28 июня 2023 г. Окончательный вариант – 19 сентября 2023 г.

Investigation of an object-detection approach for estimating the rock fragmentation in the open-pit conditions

K. Reshetnikov¹, M. Ronkin¹, S. Porshnev¹

¹ Federal State Autonomous Educational Institution of Higher Education
«Ural Federal University named after the first President of Russia B.N.Yeltsin»,
620002, Ekaterinburg, Russia, Mira str. 19

Abstract

Optimization of open-pit mining is one of significant tasks to date, with the blasting quality estimation being a key factor. The blasting quality is determined through evaluating the number of fragments and block size distribution, the so-called fragmentation task. Currently, computer vision-based methods using instance or semantic segmentation approaches are most widely applied in the task. However, in practice, such approaches require a lot of computational resources. Because of this, the use of alternative techniques based on algorithms for the real-time object detection is highly relevant. The paper studies the use of YOLO family architectures for solving the task of the blasting quality assessment. Based on the research results, YOLOv7x architecture is proposed as a baseline model. The proposed neural network architecture was trained on a dataset selected by the present authors from digital images of blasted open-pit block fragments, which consisted of 220 images. The obtained results also allow one to suggest the geometrical size of rock chunks as a measure of blasting quality.

Keywords: fragmentation, deep learning, object detection, computer vision, open-pit, blast quality estimation.

Citation: Reshetnikov KI, Ronkin MV, Porshnev SV. Investigation of an object-detection approach for estimating the rock fragmentation in the open-pit conditions. *Computer Optics* 2024; 48(2): 272-281. DOI: 10.18287/2412-6179-CO-1382.

Acknowledgements: This research was financially supported by the Russian Science Foundation and Government of Sverdlovsk region under joint grant No 22-21-20051, <https://rscf.ru/en/project/22-21-20051/>.

Authors' information

Kirill Reshetnikov, (b. 1995), PhD student in URFU by direction Unformation Technologies. Research interests: computer vision, machine learning, deep learning. E-mail: ki.reshetnikov@urfu.ru

Mikhail Ronkin, (b. 1987), PhD, Associated Proff, in UrFU. Research interests: computer vision, machine learning, deep learning, advanced signal processing, time series analysis. E-mail: m.v.ronkin@urfu.ru

Sergey Porshnev, (b. 1961), PhD, full Proff, in UrFU. Research interests: computer vision, advanced signal processing, time series analysis. E-mail: s.v.porshnev@urfu.ru

Received June 28, 2023. The final version – September 19, 2023.
