Министерство науки и высшего образования Российской Федерации Федеральное государственное автономное образовательное учреждение высшего образования «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина» Институт радиоэлектроники и информационных технологий – РТФ

Школа профессионального и академического образования

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

			Директо	р ШПиАО		
	/	Den		Д.В. Денисов		
	(П	юдпи	сь)	(Ф.И.О.)		
«	03	»	resothe	2024 г.		

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ИССЛЕДОВАНИЕ МЕТОДОВ СЕГМЕНТАЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ФРАГМЕНТАЦИИ ГОРНЫХ ПОРОД

Научный руководитель: Борисов Василий Ильич к.т.н., доцент

Консультант: Решетников Кирилл Игоревич ассистент

Нормоконтролер: Огуренко Егор Владимирович

Студент группы: РИМ-220963 Мишин Илья Сергеевич

подпись

подпись

Екатеринбург 2024 Министерство науки и высшего образования Российской Федерации Федеральное государственное автономное образовательное учреждение высшего образования

«Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий – РТФ Школа профессионального и академического образования Направление подготовки 09.04.01 Информатика и вычислительная техника Образовательная программа 09.04.01/33.03 Инженерия машинного обучения

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студента _____ Мишина Илью Сергеевича _____ группы <u>РИМ-</u>220963 (фамилия, имя, отчество)

1. Тема выпускной квалификационной работы

Исследование методов сегментации для решения задачи фрагментации горных пород и научного руководителя_____

Утверждена распоряжением по институту от «4» декабря 2023 г. № 33.02-05/298

2. Научный руководитель Борисов Василий Ильич, к.т.н., Доцент, Менеджер_

(Ф.И.О., должность, ученая степень, ученое звание)

3. Исходные данные к работе

Набор изображений

4. Перечень демонстрационных материалоз презентация_

5. Календарный план

№ п/п	Наименование этапов выполнения работы	Срок выполнения этапов работы	Отметка о выполнении
1	1 раздел (глава)	до 23.03'.2024 г.	1 OP
2.	2 раздел (глава)	до 29.04.2024 г.	
3.	3—4 раздел (глава)	до 20.05.2024 г.	··· ···
4	ВКР в иелом	до 24.05.2024 г.	

Научный руководитель Борисов Василий Ильич

Ф.И.О.

Студент задание принял к исполнению 23.05.2024

дата

6. Консультанты по проекту (работе) с указанием относящихся к ним разделов*

S PAL CATALANT	THE SECTION FROM THE	Подпись, дата		
Раздел Раздел 1 Раздел 2	Консультант	задание выдал	задание принял	
Разлел 1	Решетников К.И.	Perient	111、11111月1日、「常い」	
Раздел 2	Решетников К.И.	Perherry		
Раздел 3	Решетников К.И.	Janetry	AND	

7. Допустить Мишина Илью Сергеевича к защите выпускной квалификационной работы в экзаменационной комиссии

Директор ШПиАО

Д.В. Денисов

(подпись)

(подпись)

Ф.И.О.

ΡΕΦΕΡΑΤ

Выпускная квалификационная работа 50 с., 16 рис., 1 табл., 33 источника. Ключевые слова: компьютерное зрение, сегментация объектов, обнаружение объектов, сверточные нейронные сети, методы сегментации изображений. Объект исследования: цифровые изображения горных пород проведения взрывных работ. Предмет после исследования: метолы сегментации изображений. Цель работы: изучение и сравнение различных методов сегментации изображения для оценки распределения размеров фрагментов горных пород. В процессе исследования проводились: анализ существующих подходов к сегментации изображений, проведено обучение архитектур моделей нейронных различных сетей ДЛЯ сегментации фрагментов горных пород на изображении, проведена оценка результатов сегментации моделей с помощью метрик: mAP (средняя точность обнаружения), скорость обработки изображений в секундах, кумулятивные функции распределения (Cumulative distribution function (CMD)). В работе проведено сравнение трех моделей различной архитектуры, которые решают задачи сегментирования изображения. Область практического применения: сравнению моделей, которые способны подход К решать задачи сегментирования изображений и выбор наиболее подходящей модели для решения задач сегментирования изображения в области горнодобывающей промышленности.

СОДЕРЖАНИЕ

РЕФ	РЕФЕРАТ					
СОД	СОДЕРЖАНИЕ					
BBE	ВВЕДЕНИЕ					
1.	1. Анализ методов сегментации изображения					
	1.1 Сегментация изображений и ее методы	8				
	1.2 Модель архитектуры YOLO	9				
	1.3 Модель архитектуры Mask R-CNN	13				
2.	. Применение методов сегментации в задаче оценки фрагментаци					
горных по	горных пород					
	2.1 Описание набора данных					
	2.2 Сравнение моделей					
3.	3. Выбор модели для задачи сегментации изображения для оцен					
распределения размеров фрагментов горных пород. Обсуждение результатов.						
		31				
ЗАК	ЗАКЛЮЧЕНИЕ					
БИБ	БИБЛИОГРАФИЧЕСКИЙ СПИСОК					

ВВЕДЕНИЕ

В горнодобывающей промышленности и похожих областях, связанных с изучением горных пород или работой с ними, оценка распределения размеров фрагментов горных пород является одной из задач. Правильное определение размера горной породы дает возможность оценить различные свойства пород, такие как: прочность, плотность, проницаемость и так далее. Оценка перечисленных свойств имеет важное значение для выбора месторождения, выбора способа разработки месторождения, планирования добычи полезных ископаемых и горных пород, обогащение рудного материала.

Задачей сегментации изображений является цифровая обработка изображений, во время которой происходит разделение данных изображений на фон И объекты, которые представляют интерес. В области горнодобывающей промышленности данная технология позволяет автоматизировать и ускорить процесс оценки, а также достичь более точных результатов по обнаружению осколков горных пород на изображениях и их размеров, которые можно воспроизводить.

Объект: цифровые изображения горных пород после проведения взрывных работ.

Предмет: методы сегментации изображений.





B)

Рисунок 1 – Изображения осколков горных пород в открытых карьерах после проведения взрывных работ на изображениях а, б, в и г.

Цель работы: Изучение и сравнение различных методов сегментации изображения для оценки распределения размеров фрагментов горных пород.

Для достижения поставленной цели необходимо выполнить следующие задачи:

• Выполнить анализ существующих подходов к сегментации изображений;

• Провести обучение моделей различных архитектур нейронных сетей для сегментации фрагментов горных пород на изображении;

• Провести оценку результатов сегментации моделей с помощью метрик: mAP (средняя точность обнаружения), скорость обработки изображений в секундах, кумулятивные функции распределения (Cumulative distribution function (CMD).

1 АНАЛИЗ МЕТОДОВ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ

1.1 Сегментация изображений и ее методы

Сегментацией изображений является процесс разделения изображения на различные области и сегменты, которые соответствуют объектам интереса или фону изображения.

Существует несколько главных методов сегментации изображений. Первый метод, это пороговая сегментация (thresholding), быстрый и простой метод, особенно эффективен для контрастных изображений, данный метод основывается на выделении пикселей, значения которых превышают определенный пороговый уровень. Далее, метод сегментации на основе краев (Edge-based Segmentation), который использует алгоритмы обнаружения краев и выделяет границы объектов и разделяет изображение по этим границам. Метод сегментации на основе областей (Region-based Segmentation), данный метод объединяет соседние пиксели со схожими свойствами (текстуры, цвета и.т.д.) в однородные регионы. Четвертый метод, это сегментация на основе кластеризации (Clustering-based Segmentation), который группирует пиксели в кластеры на основе различных характеристик (текстуры, цвета и.т.д.). Метод сегментации на основе нейронных сетей (Deep Learning Segmentation), данный метод использует глубокие сверточные нейронные сети, обученные на наборах данных с размеченными масками, примеры моделей, которые работают на основе данного метода это модели архитектуры Mask R-CNN и YOLO.

Все методы сегментации изображений могут быть использованы как по отдельности, так и в комбинации для достижения более точной и надежной сегментации изображений в зависимости от поставленных задач и особенностей используемых данных.

1.2 Модель архитектуры YOLO

Модель архитектуры YOLO (You Only Look Once) - это метод сегментации изображений, основанный на сверточных нейронных сетях для обнаружения и классификации объектов на изображении. Особенностью модели YOLO архитектуры есть способность в режиме реального времени выполнять классификацию и обнаружение объектов одновременно, Также данная модель являлась одним из первых примеров архитектуры, которая способна выполнять подобное.

Ha данный момент вышло девять версий данной архитектуры: YOLOv1, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, YOLOv8 и YOLOv9. Первая версия модели архитектуры YOLO была представлена в 2016 году. Ее показатели аккуратности mAP (средняя точность обнаружения) на представленных данных были равны 63, а скорость обработки изображения в секундах была равна 155 секундам. Через один год выпускается вторая итерация модели серии YOLO, YOLOv2, которая показала улучшенные показатели как и в средней точности обнаружения, так и в практически увеличенную вдвое скорость обработки изображения за секунду. Еще через год, в 2018 году выходит третья версия модели архитектуры YOLO, YOLOv3. По результатам показателей mAP (средняя точность обнаружения) новая на то время модель архитектуры YOLO проигрывала результатам показателей mAP (средняя точность обнаружения) предыдущей версии модели архитектуры YOLO, YOLOv2. Что касательно скорости, то третья версия серии модели архитектуры YOLO, YOLOv3, также была медленнее, чем предыдущая версия серии модели архитектуры YOLO, YOLOv2. В 2020 году выходит следующая версия модели серии архитектуры YOLO, YOLOv4. Данная версия была по большей части экспериментальной, на которой проверялось и исследовалось множество технических изменений,

а также влияние этих изменений на точность модели и скорость обработки. В том же 2020 году выходит пятая версия модели серии архитектуры YOLO, YOLOv5, которая была первой модели серии данной архитектуры, которая ввела пре обученный на MS COCO набор данных, а также пятая версия модели серии архитектуры YOLO была реализована в Pytorch. В 2022 году выходит шестая версия модели серии архитектуры YOLO, YOLOv6. В шестой версии серии были исследованы новые подходы к обнаружению объектов на изображении и к сегментированию объектов на изображении, и шестая версия серии показала большую точность в симулировании набора данных СОСО, чем предыдущая пятая версия модели серии архитектуры YOLO, YOLOv5. В том же 2022 году, когда ранее вышла шестая версия модели серии архитектуры YOLO, позднее, через несколько месяцев, выходит седьмая версия модели серии архитектуры YOLO, YOLOv7. На момент 2022 года, седьмая версия модели оказалась самой точной и быстрой показав наибольший показатель точности mAP (средняя архитектурой, точность обнаружения), а также наименьшие затраты по времени на обработку одного изображения в секундах. Данная седьмая версия модели серии архитектуры YOLO, включала в себя наиболее продвинутые на момент 2022 года методы и техники глубокого обучения нейронных сетей (deep learning). В 2023 году выходит восьмая версия модели серии архитектуры YOLO, VOLOv8. Данная версия модели была разработана Ultralytics, которая использовала последние достижения в технологии для достижения увеличения универсальности модели, а также увеличения показателей точности и скорости обработки изображений. В данной версии модели серии YOLO V Ultralytics успешно архитектуры получилось развить универсальность модели, во фреймворке которой можно легко переключаться между предыдущими версиями моделей серии архитектуры YOLO, а также исследовать результаты используемых моделей, сравнивать их точности и

скорости обработки. Тем самым восьмая версия модели серии архитектуры YOLO позволила одновременно использовать и самые продвинутые метода модели серии архитектуры YOLO для задач обнаружения объектов и сегментирования изображений, так и запускать все предыдущие версии YOLO. модели серии архитектуры Благодаря усовершенствованной технологии восьмая версия модели серии архитектуры YOLO не только стала первой моделью серии по уровню универсальности, но наблюдается также и значительное повышение в эффективности, увеличение параметра точности mAP (средняя точность обнаружения), как и увеличение параметра скорости обработки каждого изображения. Последняя на сегодняшний момент версия модели серии архитектуры YOLO, это девятая версия модели, в которой провели опыты на решение проблем с потерей памяти, от которой страдают глубокие нейронные сети. Решение данной проблемы позволит повысить эффективность обучения моделей, что должно повысить показатели точности и скорости обработки изображений.

Так как Архитектура модели YOLO способна обнаруживать и классифицировать объекты в режиме реального времени, то она представляет собой эффективный и высокопроизводительный метод. Данная архитектура обладает способностью обрабатывать сцены с множеством объектов и глобальным контекстом, что делает ее полезной в различных приложениях, распознавание лиц или видеонаблюдение.

Структура архитектуры модели YOLO выглядит следующим образом: Вackbone, сверточные нейронные сети, которые выступают в роли основного инструмента извлечения признаков из входных изображений; второй модуль в архитектуре - это Neck, данный модуль получает признаки от blackbone и занимается их преобразованием в более эффективный формат для выполнения следующих задач. YOLO использует Path Aggregation Network (PAN), структура, которая для более точного обнаружения объектов сочетает

признаки нескольких уровней самой сети; третий модуль архитектуры - это Head, который на выходе генерирует данные в виде Detection head, Segmentation head или Pose head; в четвертых специализированные функции потерь, используемые для всех задач и техники оптимизации; последней частью архитектуры модели серии YOLO является аргументация и масштабирование, чтобы найти объекты, имеющие разные пропорции.



Рисунок 2 - структура архитектуры модели YOLO

Архитектура модели YOLO состоит из нескольких слоев. Сначала, входное изображение с определенными параметрами высоты и широты подается на входной слой нейронной сети. Сверточные слои являются последовательность сверточных слоев обрабатывает входное изображение с целью извлечения признаков. Чем глубже слой, тем больше абстрактные представления он извлекает. В модели архитектуры YOLO используется предварительно обученная сеть, для получения более точных и обобщенных признаков. Далее, после нескольких сверточных слоев используются слои уменьшения размерности, такие как pulling или уменьшение размера карты чтобы сократить количество признаков, параметров И сохранить репрезентативность изображения. После сверточных слоев И слоев уменьшения размерности следуют полносвязные слои, которые преобразуют признаки в вектора вероятностей классов и координаты прямоугольников, описывающих обнаруженные объекты. Выход нейронной сети (выходной слой) - это массив, содержащий предсказанные классы объектов и их координаты. Координаты объектов представлены в виде нормализованных значений, относительных к размеру входного изображения. Для устранения множественных предсказаний одного объекта и улучшения точности обнаружения применяется алгоритм подавления немаксимума. Алгоритм удаляет дубликаты предсказанных объектов и оставляет только наиболее уверенные предсказания с наибольшей вероятностью и наименьшим перекрытием. В завершении, модель архитектуры YOLO обучается на размеченных наборах данных, где происходит оптимизация параметров нейронной сети с использованием методов функции потерь и градиентного спуска (среднеквадратичная ошибка или кросс-энтропия).

1.3 Модель архитектуры Mask R-CNN

Модель R-CNN (Region-Convolutional Neural Network) была представлена в 2014 году. Данная модель была одна из первых моделей, которая для задачи обнаружения объектов использовала сверточные нейронные сети (Convolutional Neural Networks), на ранних этапах развития архитектуры R-CNN показывала неплохие результаты по точности, но уступала по скорости обработки изображений другим моделям.

В 2015 году, через год, после того, как вышла модель R-CNN (Region-Convolutional Neural Network), вышла обновленная версия модели Fast R-CNN, целью данного обновления была повысить скорость обработки изображений.

В 2017 после Faster R-CNN выходит модель Mask R-CNN (Mask Region-Convolutional Neural Network), в данной модели стал использоваться новый подход к задаче по обнаружению объектов и сегментации изображений, который позволил выполнять данные задачи одновременно в одной сети. Модель Mask R-CNN (Mask Region-Convolutional Neural Network) является одной из самых известных моделей в области сегментации изображений и обнаружении объектов.

После появления модели Mask R-CNN (Mask Region-Convolutional Neural Network), которая улучшила показатели точности и скорости обработки, были предприняты попытки создать новые версии модели на основе той же архитектуры, но до сих пор модель Mask R-CNN (Mask Region-Convolutional Neural Network) остается одной из наиболее распространенной и самой используемой моделью в области компьютерного зрения для задач обнаружения объектов и сегментирования.

Модель Mask R-CNN (Mask Region-Convolutional Neural Network) - это метод сегментации изображений, основанный на архитектуре Faster R-CNN (Region-Convolutional Neural Network) и расширенный для извлечения масок сегментов объектов на изображении. Модель Mask R-CNN стала одним из ведущих методов сегментации, обеспечивая высокую точность и детализацию сегментации объектов.

Модель Mask R-CNN (Mask Region-Convolutional Neural Network) - это один из метод сегментации изображений, основанный на архитектуре Faster R-CNN (Region-Convolutional Neural Network) и расширенный для извлечения масок сегментов объектов на изображении. Модель Mask R-CNN стал одним

из самых популярных методов сегментации, который обеспечивает высокую детализацию и точность объектов при сегментировании.

Функция региона предложений RPN (Region Proposal Network, RPN) часть архитектуры Faster R-CNN, которая генерирует кандидаты ЭТО которые могут содержать объекты на изображениях. RPN регионов, просматривает различные масштабы и соотношения сторон, затем данная функция вычисляет предложения регионов с использованием сверточной карты признаков и применяет классификацию и регрессию для определения наилучших регионов. Модель Mask R-CNN использует предварительно обученные сверточные нейронные сети, такие как ResNet или VGGNet, для извлечения признаков из входного изображения. После серии сверточных слоев получается признаковая карта с высокоуровневыми признаками. Далее, RoI-Align (Региональная сверточная сеть), модифицированная версия сети RoI Pooling, которая позволяет более точное выравнивание признаков объектов в регионах. Rol-Align устраняет дискретизацию и искажения, обеспечивая точные и пропорциональные выравнивания между объектами и признаковыми картами. Recognition Head (функция распознавания) после RoI-Align применяется функция распознавания, включающая полносвязные слои и классификаторы для определения присутствия объекта определенного класса в каждом регионе предложения. Recognition Head также выполняет регрессию для точного ограничения обнаруженных объектов. Mask Head (Функция маскирования) на основе регионов предложений через предсказания маски каждого объекта производит сегментацию объектов. Mask Head использует сверточные слои для создания плотных масок объектов, соответствующих их положению и форме. Для обучения модели Mask R-CNN применяется несколько потерь, включая потери такие, как классификации, обнаружения, регрессии И сегментации. Потери соответствуют компоненту каждому сети, оптимизируются с И

использованием градиентного спуска и алгоритма обратного распространения ошибки.

Структура архитектуры модели Mask R-CNN выглядит следующим образом: во первых Backbone, сверточная нейронная сеть, которая извлекает визуальные признаки. Второй модуль - это Feature Pyramid Network, который дает возможность обнаруживать объекты наиболее эффективным способом вне зависимости от размеров объектов на изображении. Третья часть архитектуры модели Mask R-CNN - это сеть Region Proposal Network, которая для потенциальных объектов, которые могут представлять интерес предлагает разметку (bounding boxes). Далее модель Mask R-CNN использует методы Rol Align и Rol Head, первый метод используется для для того, чтобы извлекать признаки из областей, которые были предложены, а второй метод используется для классификации ограничивающих рамок и для сегментации объектов. В заключении функции потерь, используемые для задач классификации сегментации объектов, И базовые инструменты трансформации, которые поворачивают, масштабируют, отражают.



Рисунок 3 - структура архитектуры модели Mask R-CNN

Модель Mask R-CNN способна достигать высокой детализации сегментации объектов на изображении благодаря включению функции маскирования и использованию сверточных нейронных сетей с высокоуровневыми признаками.

Так же как и модель YOLO, модель Mask R-CNN широко применяется в таких областях, как видеонаблюдение, обработка изображений и во многих других областях, требующих точной и высококачественной сегментации объектов.

2 ПРИМЕНЕНИЕ МЕТОДОВ СЕГМЕНТАЦИИ В ЗАДАЧЕ ОЦЕНКИ ФРАГМЕНТАЦИИ ГОРНЫХ ПОРОД

2.1 Описание набора данных

Набор данных состоит из 253 изображений с разрешением 2592 × 2048 формата bmp открытых карьеров асбестосодержащей породы после взрывных работ. Фотографии сделаны в Баженовском месторождении при различных погодных условиях (во время дождливых дней, во время солнечных дней, во время тумана, во время, когда идет снег и так далее) и при различное время года (осенью, зимой, весной и летом). Все фотографии имеют разметку в файлах формата json для выполнения задачи сегментации изображения и для задачи обнаружения объектов. Разметка была сделана с обеспечения CVAT. Данное помощью программного программное обеспечение дает возможность выполнять задачи по обнаружению объектов с помощью обрамляющих рамок, так и сегментацию объектов. На рисунке ниже приведены примеры разметок.





Рисунок 4 – Примеры обнаружения объектов на изображениях а, б, в и г.

Анализ разметки из примеров на рисунке демонстрирует, что максимальная дистанция между двумя точками на гранях соотносится с размером разметки, что позволяет выдвинуть положение о том, что в данной работе есть возможность решить менее ресурсозатратную задачу по обнаружению объектов, чем задачу по сегментации изображений.

Для каждого участка карьера было сделано несколько фотографий, на которых находятся от двадцати до восьмидесяти фрагментов горных пород, отличающихся по размеру.

Для каждого эксперимента было рассмотрено от трех до пяти участков открытых карьеров, которые были выбраны самой геологической службой, как наиболее репрезентативные участки. Каждый участок был снят с различных ракурсов разных его площадей. Система, с помощью которой были сделаны фотографии, используемые в данной работе выглядит следующим образом: электрически регулируемая платформа поворотного стола, которая была предварительно установлена на треноге, камера для компьютерного зрения «Dalsa Genie Nano M2590NIR» с полутоновой матрицей в один дюйм и разрешением 2590 х 2048 пикселей, также имеющей повышенную чувствительность в ближнем инфракрасном диапазоне; объектив камеры «LMZ25300M3P-IR» с электрорегулировкой, имеющий двеннадцатикратное увеличение и повышенной чувствительностью в ближнем инфракрасном диапазоне. Подробнее о наборе данных и сам набор данных можно найти на сайте DOI:10.17632/pfdbfpfygh [30].

2.2 Сравнение моделей.

Задача данной работы является сравнение различных архитектур для задачи сегментации изображений для фрагментации горных пород, чтобы узнать, насколько качественно и быстро работает каждая архитектура. Модели, которые были выбраны, являются восьмая версия модели серии архитектуры YOLO, YOLOv81, девятая версия модели серии архитектуры YOLO, YOLOv9c и модель архитектуры Mask-RCNN. Модель архитектуры YOLOv81 и модель архитектуры Mask-RCNN являются одними из самых популярных архитектур для задачи сегментации изображений, а модель архитектуры YOLOv9c самая новая на данный момент версия из серии моделей архитектуры YOLO.

обучения моделей использовались изображения разрешения Для 420x420. Использование больших форматов изображений, например 1080Х1920 может помочь в обучении моделей сегментирования изображений, увеличить качество обучения и четкость результатов. Но большие форматы изображений повышают вычислительные затраты и требуют большей памяти. Поэтому, чтобы оптимизировать процесс обучения моделей был выбран формат 420Х420. Данный формат изображения позволил найти баланс между затрат времени обучения и сложностью точностью моделей, ПО В компьютерных вычислениях.

Для обучения использовался датасет из 253 изображений горных пород. Данные изображения были разбиты на три группы: 1 - группа фотографий для обучения, 2 - группа фотографий для тестирования, 3 - группа фотографий для валидации. В первой группе было 68% от датасета, во второй группе 22% от датасета, в третьей группе 10%. С помощью средних значений производится оценка качества моделей. Метрики, с помощью которых проводилась оценка моделей: точность (Box(P), Mask(P)), полнота (Box(R), Mask(R)), взвешенная средняя точность(Box (mAP50), Box (mAP50-95), Mask (mAP50), Mask (mAP50-95)), скорость (Speed).

Параметры, с которыми проходило обучение моделей были следующими: число эпох (epochs) равно 100, размер батча было равно 16. Для YOLO моделей параметры аугментации были следующие: оттенок (hsv_h) = 0.015, насыщенность (hsv_s) = 0.7, яркость (hsv_v) = 0.4, вероятность поворота изображения (fliplr) = 0.5, масштабирование изображения (scale) = 0.5. Наилучшая модель (best fit model) использовалась для тестирования.

Выбранные для данной работы модели сравниваются в соответствии с параметрами средней точности обнаружения (mean average precision (mAP)), скорости обработки изображения в секундах и отклонение от максимального размера разметки на графиках кумулятивных функциях распределения (Cumulative distribution function (CMD)).

Параметр средней точности обнаружения (mean average precision (mAP)) часто используется для того, чтобы оценить модели на производительность в области компьютерного зрения для задач обнаружения объектов и сегментирования. Данный параметр позволяет измерить точность обнаружения.

Вычисление средней точности обнаружения (mean average precision (mAP)) происходит в несколько этапов. Сначала для каждого класса объектов строится кривая precision-recall, данная кривая показывает соотношение между полнотой (recall) и точностью (precision) на разных порогах обнаружения. На следующем этапе вычисляется площадь под кривой precision-recall (Average Precision) для указанного класса. Average Precision является интегралом кривой precision-recall и измеряет среднюю точность для данного класса на различных уровнях полноты. На третьем этапе повторяется тот же процесс, что был на втором этапе, но только для всех классов объектов, которые имеются в наборе данных. mean Average Precision (mAP) является средней величиной Average Precision по всем классам.

Математическое определение средней точности обнаружения (mean average precision (mAP)) выглядит следующим образом:

$$mAP = (1 / N) * \Sigma APi$$

где N — количество классов, APi — значение Average Precision для каждого отдельного класса.

Показатель средней точности обнаружения (mean average precision (mAP)) принимает значения от нуля до единицы, где единица является идеальной точностью. Таким образом, чем ближе показатель средней точности обнаружения (mean average precision (mAP)) ближе к единице, тем лучше модель справляется с задачей обнаружения, а чем параметр средней точности обнаружения (mean average precision (mAP)) ближе к нулю, то данная модель с задачей обнаружения справляется хуже.

Параметр средней точности обнаружения (mean average precision (mAP)) является одной из самых известных метрик, так как кроме учитывания точности, данный параметр учитывает и полноту обнаружения для классов. Подобная особенность дает возможность провести всестороннюю оценку модели, особенно в случаях, когда классы имеют разное распределение в наборе данных. Параметр средней точности обнаружения (mean average precision (mAP)) широко используется в соревнованиях и исследованиях по обнаружению объектов, таких как СОСО.

Второй параметр, который используется для визуального анализа это графики кумулятивных функциях распределения (Cumulative distribution function (CMD)).

Функциях распределения (Cumulative distribution function (CMD)) - это одно из основных базовых понятий теории вероятности и статистики, которая

показывает вероятность, что случайная величина будет равна, больше или меньше заданному значению.

Функция распределения (Cumulative distribution function (CMD)) используются во многих областях, в финансовой математике, статистическом анализе и так далее. Функция распределения (Cumulative distribution function (CMD)) позволяет находить вероятности того, как попала случайная величина в определенный интервал.

Функция распределения (Cumulative distribution function (CMD)) используют для построения доверительных интервалов, проведения и определения статистических тестов и статистических гипотез.

Функция распределения (Cumulative distribution function (CMD)) принимает различные формы в зависимости от того, какой тип распределения: непрерывное или дискретное.

Таблица 1 – Результаты обучения модели архитектуры YOLOv8l, модели архитектуры YOLOv9c и модели архитектуры Mask R-CNN

Модель	Точность		Полнота		mAP50		mAP50-95		mc
									per
									imag
									e
	Box(P)	Box (R)	Mask(P)	Mask(R)	Box(mAP50)	Mask(mAP 50)	Box(mAP50- 95)	Mask(mAP 50-95)	
YOLO8	0.788	0.70 7	0.772	0.654	0.751	0.71	0.533	0.401	13.3
YOLO9	0.795	0.70 4	0.752	0.658	0.743	0.687	0.527	0.38	19.5
R-CNN					0.48		0.37		167.3

Изучая показатели параметров из таблицы результатов видим, что худшие значения показателей у Mask R-CNN. Несмотря на то, что в серии моделей YOLO появилась новая девятая версия модели, но результаты восьмой версии лучше, что также делает модель архитектуры YOLOv8c моделью с наилучшими показателями. Показатель mAP (средняя точность обнаружения) для модели архитектуры YOLOv8l равен 0.751, на втором месте модель архитектуры YOLOv9c со значением 0.743, и на последнем месте по значению показателя mAP (средняя точность обнаружения) идет модель архитектуры Mask R-CNN, со значением показателя равному 0.48. По времени обработки каждого изображения больше всего времени требуется модели архитектуры Mask R-CNN, а именно 167.3 миллисекунд на обработку одного изображения, затем по затрате времени идет модель архитектуры YOLO9c, где затрата по времени равна 19.5 миллисекунд на обработку одного изображения и самая быстрая скорость обработки изображения у модели архитектуры YOLO81, которая равна 13.3 миллисекунд на обработку одного изображения.



Рисунок 5 – Пример результата обнаружения обнаружения осколков горных пород в открытых карьерах после проведения взрывных работ из валидационного набора данных для модели архитектуры YOLOv81



Рисунок 6 – Пример результата обнаружения обнаружения осколков горных пород в открытых карьерах после проведения взрывных работ из валидационного набора данных для модели архитектуры YOLOv9c



Рисунок 7 – Пример результата обнаружения обнаружения осколков горных пород в открытых карьерах после проведения взрывных работ из валидационного набора данных для модели архитектуры Mask R-CNN

Из рисунков с седьмой по восьмой можно увидеть, что наиболее явные различия в работе выбранных моделей прослеживаются на изображениях с большим количеством небольших осколков горных пород. Модель на основе архитектуры Mask R-CNN имеет тенденцию к сегментированию объектов в тех местах, где осколков горных пород не наблюдается, при наличии большого количества осколков на изображении.

Далее описан визуальный анализ результатов сегментирования изображений на вопрос точности результата восстановления распределения размеров осколков горных пород в открытых карьерах после проведения взрывных работ для выбранных для анализа моделей. Были отдельно построены для каждой модели кумулятивные функции распределения (Cumulative distribution function (CMD)).

Из визуального анализа можно увидеть, что лучший результат показывает модель архитектуры YOLOv8l, график максимального размера которой наиболее близок к графику максимального размера разметки. Модель архитектуры Mask-RCNN показала наихудший результат, так как показатели максимального размера для данной модели имеют наибольшее отклонение от графика показателей максимального размера разметки. Модель архитектуры YOLOv9c, визуально имеет график максимального размера близок к графику модели архитектуры YOLOv8l, но отклонение от максимального размера разметки выглядит больше, чем у модели архитектуры YOLOv8l, но ближе, чем показатель максимального размера для модели архитектуры Mask-RCNN.



Рисунок 8 – График кумулятивной функции распределения (Cumulative distribution function (CMD)) для модели архитектуры YOLOv81



Рисунок 9 – График кумулятивной функции распределения (Cumulative distribution function (CMD)) для модели архитектуры YOLOv9c



Рисунок 10 – График кумулятивной функции распределения (Cumulative distribution function (CMD)) для модели архитектуры Mask R-CNN

3 ВЫБОР МОДЕЛИ ДЛЯ ЗАДАЧИ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ ДЛЯ ОЦЕНКИ РАСПРЕДЕЛЕНИЯ РАЗМЕРОВ ФРАГМЕНТОВ ГОРНЫХ ПОРОД. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ.

В данной работе проведена оценка нескольких современных моделей машинного обучения для задачи фрагментации изображения. Для анализа были выбраны следующие модели: YOLOv8l, YOLOv9c и Mask-RCNN. Подготовка изображений, условия проведения обучения каждой модели были одинаковы. Сам процесс обучения был разделен в три этапа: этап тренировки моделей на изображениях горных пород после проведения взрывных работ, которые были предварительно размечены; этап валидации с установленным одинаковым коэффициентом уверенности (confidence); третий этап тестирования моделей на изображениях горных пород после проведения взрывных работ, выбрания моделей на изображениях горных пород после проведения.

В соответствии с результатами обучения моделей лучше всех себя показала восьмая версия модели серии YOLO, так как она имеет самые высокие коэффициенты показателей, а также самую быструю скорость обработки отдельно взятого изображения.

Интересным для дальнейшего изучения представляется тот факт, что более новая модель серии YOLO показала хуже результаты, чем предыдущая к ней версия модели этой серии. Подобный случай уже можно было наблюдать на презентации третьей версии серии моделей архитектуры YOLO, YOLOv3 в 2018 году, где данная версия модели показала результаты показателей mAP (средняя точность обнаружения) хуже, чем результаты предыдущей версии модели архитектуры YOLO, YOLOv2, а также оказалось, что третья версия серии модели архитектуры YOLO, YOLOv3, была

медленнее, чем предыдущая версия серии модели архитектуры YOLO, YOLOv2.

Также, для продолжения изучения вопроса данной работы предлагается изучить и сравнить другие модели, которые способны выполнять задачи сегментации изображений, такими моделями, которые могут представлять интерес для дальнейшего изучения вопроса сегментации изображения на примере изображений с осколками горных пород в открытых карьерах после проведения взрывных работ могут быть такие модели, как SAM (segment anything model), выпущенной в 2023 году, Detection2, которая была разработана той же компанией, модель Detic, выпущенная в 2022 году, модель сегментации в реальном времени YOLACT.

Segment anything model представляет интерес из-за своей универсальности, так как модель была обучена на огромном количестве данных.

SAM (Segment Anything Model) является примером архитектуры сверточной нейронной сети, которая разработана для задач сегментации объектов на изображениях. Архитектура SAM (Segment Anything Model) состоит из нескольких главных частей: первый компонент является энкодер изображения, который при обработке входного изображения извлекает визуальные признаки по средствам сверточной нейронной сети (CNN), такой как ResNet или EfficientNet. После формируется тензор признаков из полученных признаков, который будет использоваться в будущих этапах. Следующий компонент- это энкодер маски, данный компонент обрабатывает входную маску, а данная входная маска является разметкой объекта небольшого размера, которая в свою очередь была задана пользователем. Входная маска проходит через полносверточную сеть небольшого размера, а та в свое время извлекает признаки маски. Третьим компонентом является кросс-энкодер, данный компонент посредством механизма трансформера с

перекрестным вниманием объединяет маски и признаки изображения. Кросс-энкодер позволяет модели принимать во внимание взаимное влияние между признаками маски и изображения, что может быть основным фактором в обобщении на новые категории объектов. Маска декодер: Этот компонент генерирует окончательную маску сегментации объекта на основе Он результатов кросс-энкодера. состоит ИЗ последовательности полносверточных слоев, которые преобразуют признаки из кросс-энкодера в пиксельную сегментацию. Пятый компонент - это вспомогательные головки, для модели SAM (Segment Anything Model) данный компонент решает дополнительные задачи, различные например такие задачи как классификация объектов или предсказание границ объекта. Данный компонент улучшают производительность основной задачи сегментации.

SAM (Segment Anything Model) имеет одной из ключевых особенностей способность к обобщению на невидимые ранее моделью новые категории объектов. Подобное было достигнуто с помощью обучения со сквозным вниманием (end-to-end prompt learning), в процессе данного обучения модель SAM (Segment Anything Model) учится извлекать релевантные признаки из небольшой входной маски, а затем применять их к сегментации объектов.

SAM (Segment Anything Model) показывает высокий уровень производительности и универсальность для задач в области компьютерного зрения, сегментирования изображения, что позволяет соревноваться в точности и скорости с другими современными моделями, как Mask R-CNN, которая была описана в данной работе ранее.

SAM (Segment Anything Model) используют в промышленных и медицинских областях, а также в различных программных обеспечениях.



Рисунок 11- структура архитектуры модели SAM

Модель архитектуры Detectron2, которая до этого имела название Facebook AI Research. Модель архитектуры Detectron2, которая представляет собой модульную библиотеку, которая выполняет задачи по обнаружению объектов на изображениях и сегментации объектов на изображениях.

В основу архитектуры модели Detectron2 основана на двух популярных на данный момент моделях архитектуры R-CNN: Mask R-CNN и Faster R-CNN.

Модель архитектуры Detectron2 состоит из нескольких компонентов. Первый компонент модели архитектуры Detectron2 является Backbone. Данный компонент является глубокой сверточной нейронной сетью (CNN), используемые для извлечения визуальных признаков из входных изображений. Второй компонент модели архитектуры Detectron2 является Feature Pyramid Network (FPN). Данный компонент является эффективной структурой, которая строит пирамиду из признаков на разных масштабах, что позволяет обнаруживать объекты различных размеров. Третий компонент модели архитектуры Detectron2 является Region Proposal Network (RPN). Данный компонент является сетью, которая предлагает потенциальные области объектов (bounding boxes), которые могут содержать объекты интереса. Region Proposal Network (RPN) работает на карте признаков, полученной от Feature Pyramid Network (FPN). Четвертым компонент модели архитектуры Detectron2 является RoI Head. Данный компонент является модулем, ответственным за предсказание классов объектов и границ классов объектов (bounding boxes) на основе предложенных областей от Region Proposal Network (RPN). Пятый компонент модели архитектуры Detectron2 является Mask Head, которая служит для выполнения задач сегментации. Данный компонент является дополнительным модулем, который генерирует маски сегментации для обнаруженных объектов. Шестой компонент модели архитектуры Detectron2 является функции потерь. Различные функции потерь, используемые модели архитектуры Detectron2 для оптимизации задач обнаружения объектов на изображениях, классификации объектов на изображениях и сегментации объектов на изображениях. Седьмой компонент модели архитектуры Detectron2 является механизмы увеличения данных. Данный компонент является библиотекой, которая предоставляет способы увеличения данных, такие как масштабирование, вращение, случайные преобразования изображений, отражение и так далее, данный процесс улучшить обобщающую способность моделей. помогает Восьмым компонентом модели архитектуры Detectron2 является обучение и вывод. Данный компонент позволяет модели архитектуры Detectron2 поддерживать распределенное обучение на нескольких GPU, а также вывод на CPU и GPU с оптимизацией для задач различного типа.



Рисунок 12- структура архитектуры модели Detectron2

Отличительными чертами модели архитектуры Detectron2 являются следующие характеристики: модульность, гибкость и масштабируемость. Архитектура модели Detection2 позволяет без относительно больших затрат проводить эксперименты с различными компонентами и настройками для достижения оптимальной производительности для решения задач сегментации объектов на изображениях и обнаружение объектов на изображениях.

DETIC - это модель детекции и сегментации объектов, основанная на трансформерах. Она использует деформируемые ядра внимания для более точного моделирования геометрии объектов. Существуют несколько особенностей архитектуры DETIC. Во первых, Backbone - использует сверточные нейронные сети (CNN) для получения признаков изображения. Второй компонент, это деформируемый трансформер - применяет механизм деформируемого внимания для моделирования геометрии объектов. Третьим

компонентом является головка детекции - предсказывает ограничивающие рамки, классы объектов и маски. Четвертый компонент, это обучение, которое проводится в два этапа: предварительное обучение на задаче классификации, затем тонкая настройка на задаче детекции и сегментации.



Рисунок 13- структура архитектуры модели DETIC

YOLACT - это архитектура для быстрой сегментации объектов в 2019 реального времени. Она была разработана В режиме году исследователями из Университета Калифорнии в Санта-Крузе. Выделяют 5 основных компонентов архитектуры YOLACT. Первый компонент это Backbone, в качестве которого используется сверточная нейронная сеть (например, ResNet-101 или ResNet-50). Васкbone отвечает за извлечение признаков из входного изображения. Второй компонент - генератор масок (Mask Head), который состоит из 4 сверточных слоев, которые сокращают размерность признаков backbone. Mask Head генерирует набор протоматриц -

базовых масок, которые затем комбинируются для получения маски объекта. Третий компонент, это предиктор коэффициентов (Coefficient Predictor), который состоит из 3 сверточных слоев и предсказывает коэффициенты для линейного комбинирования протоматриц. Данные коэффициенты определяют, как протоматрицы должны быть скомбинированы для получения итоговой маски объекта. Четвертым компонентом идет предиктор боксов (Box Predictor), который состоит из нескольких сверточных слоев и предсказывает координаты ограничивающих боксов объектов. Последний компонент архитектуры модели, это процесс предсказания, который для каждого объекта предиктор боксов выдает координаты ограничивающего бокса. Генератор масок выдает набор протоматриц. Предиктор коэффициентов предсказывает коэффициенты для линейной комбинации протоматриц. Итоговая маска объекта получается как линейная комбинация протоматриц с предсказанными коэффициентами.



Рисунок 14- структура архитектуры модели YOLACT

Основное преимущество YOLACT - возможность быстрого предсказания масок объектов в режиме реального времени за счет использования линейной комбинации протоматриц вместо построения сложных архитектур сегментации.

Для дальнейшего анализа вопроса сегментации изображений осколков горных пород в открытых источниках так же помимо уже рассмотренных моделей предлагается изучить современные архитектуры сегментации изображений на базе трансформеров: Swin-Transformer, VIT (Vision Transformer Model).

Swin Transformer - это архитектура, основанная на трансформерах, которая была предложена исследователями из Университета Тунцзи в 2021 году. Она стала важным развитием в области применения трансформеров для задач компьютерного зрения, а также показала высокую эффективность на множестве задач компьютерного зрения.

Есть пять основных компонентов архитектуры Swin Transformer. Первый компонент, ЭТО иерархическая структура, Swin Transformer использует иерархическую структуру, аналогичную сверточным нейронным сетям. Входное изображение последовательно проходит через несколько стадий, каждая из которых уменьшает пространственное разрешение, но увеличивает размерность признаков. Второй компонент, это оконное внимание (Window Attention), который является механизмом внимания в Swin Transformer применяется не ко всему изображению, а к локальным оконным областям. Данный компонент позволяет эффективно моделировать пространственные зависимости И снижает вычислительную сложность. Третьим компонентом является сдвиг окон (Window Shift), между стадиями которого происходит сдвиг окон внимания, что позволяет глобальные зависимости. моделировать Сдвиг окон (Window Shift) обеспечивает эффективный захват как локальных, так и глобальных

особенностей. Четвертый компонент, это многомасштабность Swin Transformer, которая использует многомасштабные признаки, полученные на разных стадиях. Многомасштабность Swin Transformer позволяет эффективно совмещать низкоуровневые и высокоуровневые особенности. Последний пятый компонент, это модульность. Архитектура Swin Transformer построена из модульных блоков, что обеспечивает гибкость и масштабируемость. Модули могут быть легко адаптированы для различных задач компьютерного зрения.

Среди основных преимуществ Swin Transformer выделяют высокую производительность на множестве задач компьютерного зрения, эффективное моделирование пространственных зависимостей благодаря оконному глобальных вниманию, возможность захвата как локальных, так И особенностей, модульность и масштабируемость архитектуры.



Рисунок 15- структура архитектуры модели Swin Transformer

VIT (Vision Transformer Model) - это модель, основанная на трансформерах, которая была предложена исследователями в 2020 году. ViT стала прорывом в области компьютерного зрения и продемонстрировала высокую эффективность на задачах компьютерного зрения, таких как классификация изображений.

Всего насчитывается пять основных компонентов модели архитектуры ViT (Vision Transformer Model). Во первых, изображение разбивается на патчи. Входное изображение разбивается на фиксированные по размеру патчи (например, 16x16 пикселей). Каждый патч рассматривается как отдельный "токен" входной последовательности.Второй компонент, это линейная проекция патчей, где каждый патч проецируется в вектор фиксированной размерности с помощью полносвязного слоя. Это позволяет представить изображение в виде последовательности токенов. Затем производится добавление позиционной информации. Для сохранения пространственной информации к каждому токену добавляются позиционные эмбеддинги. Позиционные эмбеддинги могут быть либо фиксированными, либо обучаемыми. Четвертый компонент, это трансформерный кодировщик, собой который представляет последовательность токенов, включая позиционные эмбеддинги, подается на вход трансформерного кодировщика. Трансформерный кодировщик состоит из множества идентичных блоков, каждый из включает слой мультиголовочного которых внимания и полносвязный слой. Последний компонент архитектуры модели ViT (Vision Transformer Model). ЭТО классификационный токен. В начало последовательности токенов добавляется специальный классификационный токен *<*CLS*>*, который используется для финальной классификации. Выход трансформерного кодировщика для классификационный токена <CLS> подается на полносвязный слой для предсказания класса изображения.



Рисунок 16- структура архитектуры модели ViT (Vision Transformer Model)

Среди преимуществ ViT (Vision Transformer Model) основных выделяют высокую производительность на задачах классификации изображений, возможность захвата глобальных и локальных зависимостей трансформеров, масштабируемость и благодаря механизму внимания возможность обучения на больших объемах данных, гибкость архитектуры, позволяющая применять ее к различным задачам компьютерного зрения.

Из недостатков можно указать, что модель Vision Transformer (ViT), показала, что трансформеры могут быть эффективны для обработки изображений, однако ViT имеет ограничения в моделировании

пространственных зависимостей из-за использования глобального механизма внимания.

ЗАКЛЮЧЕНИЕ

В данной работе мы рассмотрели несколько моделей сегментации изображений на примере задачи фрагментации горных пород в открытых карьерах после проведения взрывных работ. Были рассмотрены модель Mask-RCNN (Mask Region-Convolutional Neural Network) и две модели серии YOLO (You Only Look Once): модель YOLOv8l и модель YOLOv9c. Сравнение моделей происходил через параметр взвешенной средней точности (mAP). Лучший результат продемонстрировала модель YOLOv8l с параметром взвешенной средней точности (mAP) равным 0,751, второй по точности была модель YOLOv9c с параметром взвешенной средней точности (mAP) равным 0,743 и на третьем месте была модель архитектуры Mask-RCNN (Mask Region-Convolutional Neural Network) с параметром взвешенной средней точности (mAP) равным 0,743 и на третьем месте была модель архитектуры Mask-RCNN (Mask Region-Convolutional Neural Network) с параметром взвешенной средней точности (mAP) равным 0,48.

Также модель YOLOv8l показала самую быструю скорость равную 13.3 миллисекунд за одно изображение, а модель Mask-RCNN оказалась самой медленной моделью скорость которой равна 167,3 миллисекунд за одно изображение. Второй по скорости результат достается модели серии YOLO - модель YOLOv9c, скорость которой равна 19.5 миллисекунд за одно изображение.

Построение графиков кумулятивных функции распределения (Cumulative distribution function (CMD)) для каждой модели, которые были выбраны для исследования вопроса данной работы, показывает следующие результаты: во первых, максимальный размер модели архитектуры YOLOv8l имеет наименьшее отклонение от максимального размера разметки; во вторых, максимальный размер, полученный для модели архитектуры YOLOv9c визуально имеет отклонение от максимального размера разметки больше, чем модель архитектуры YOLOv8l; в третьих, самое больше

отклонение от максимального размера разметки из всех трех моделей имеет отклонение максимального размера полученное для модели архитектуры Mask-RCNN (Mask Region-Convolutional Neural Network).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Gonzalez, R.C., Woods, R.E., Eddins, S.L. Digital Image Processing Using MATLAB. [Текст]: учеб. пособие для студ. среднего проф. образования / Gonzalez, R.C., Woods, R.E., Eddins, S.L. Torino: Pearson Education. 2009. -609 с.

2. Szeliski, R. Computer Vision: Algorithms and Applications. [Текст]: учебник для студ. вузов / Szeliski, R. Germany. Springer. 2010. - 812 с.

3. Jain, A.K., Dubes, R.C. Algorithms for Clustering Data. [Текст]: учебник для студ. вузов / Jain, A.K., Dubes, R.C. Torino: Prentice-Hall. 1988. - 334

4. Russell, S.J., Norvig, P. Artificial Intelligence: A Modern Approach. [Текст]: учебник для студ. вузов / Russell, S.J., Norvig, P. Torino: Pearson Education. 2016. - 1151 с.

5. Shafer, S. Computer Vision: From Surfaces to 3D Objects. [Текст]: учебник для студ. вузов / Shafer, S. London: Chapman & Hall. 2014. - 214 с.

6. Haralick, R.M., Shapiro, L.G. Computer and Robot Vision. [Текст]: учебник для студ. вузов / Haralick, R.M., Shapiro, L.G. Boston: Addison-Wesley. 1992. - 617 с.

 Smith, J.R., Clarke, T., Farid, H. (2016). Classification of rocks using computer vision and convolutional neural networks. [Teκct] / Smith, J.R., Clarke, T., Farid, H.. // Journal of Applied Remote Sensing. 2016. №10. C. 56-72.

8. Arfaghani, A., Shivanian, E. Segmentation and characterization of rock micro-CT images: A review. [Teκct] / Arfaghani, A., Shivanian, E. // Computers & Geosciences. 2020. №138. C. 104-153.

9. Yu, L., Lu, J. A Review on Image Segmentation Techniques for Rock Image Analysis. [Teкст] / Yu, L., Lu, J. // ISPRS International Journal of Geo-Information. 2020 №9. C. 62-106.

10. Haralick, R.M., Shapiro, L.G. Computer and Robot Vision. [Текст]: учебник для студ. вузов / Haralick, R.M., Shapiro, L.G. Boston: Addison-Wesley. 1992. - 617 с.

11. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. Mask R-CNN. [Текст] / Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. // Computer Science. 2017. №67. С. 25-39.

12. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko. End-to-End Object Detection with Transformers. [Текст] / Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko. // Computer Science. 2020. №95. C. 61-93.

13. Xinlei Chen, Saining Xie, Kaiming He. An Empirical Study of Training Self-Supervised Vision Transformers. [Текст] / Xinlei Chen, Saining Xie, Kaiming He. // Computer Science. 2021. №102. С. 34- 51.

14. Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. [Текст] / Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, Ross Girshick. // Computer Science. 2021. №102. C. 106-115.

15. Xinlei Chen, Li-Jia Li, Li Fei-Fei, Abhinav Gupta. Deforming Autoencoders: Unsupervised Disentangling of Shape and Appearance. [Текст] / Xinlei Chen, Li-Jia Li, Li Fei-Fei, Abhinav Gupta. // Computer Science. 2020. №93. C. 75- 81.

16. Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarkiewicz, Cécile Cappelle, Theo Coomans, Michael Cordts. Self-supervised Semantic Segmentation for Autonomous Driving. [Текст] /

Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarkiewicz, Cécile Cappelle, Theo Coomans, Michael Cordts. // Computer Science. 2022. №110. C. 12- 31.

17. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. The pascal visual object classes (voc) challenge. [Текст] / Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. // International journal of computer vision. 2010 №88. C. 303-338.

18. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D & Zitnick, C. L. Microsoft coco: Common objects in context. In European conference on computer vision [Текст]: учебник для студ. вузов / Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D & Zitnick, C. L. Cham: Springer. 2014. 755 c.

19. Padilla, R., Neira, W. J., & Aitken, N. (2020). A simple implementation of the improved precision-recall metric for object detection evaluation. [Текст] / Padilla, R., Neira, W. J., & Aitken, N. // Computer Science. 2020. №192. C. 15-31.

20. Salton, G., & McGill, M. J. Introduction to modern information retrieval. [Текст]: учеб. пособие для студ. среднего проф. образования / Salton, G., & McGill, M. J. New York: Pearson Education. 1986. - 569с.

21. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. [Текст] / Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. // IEEE . 2019 №106. С. 658-666.

22. Hogg, R. V., McKean, J. W., & Craig, A. T. Introduction to mathematical statistics. [Текст]: учеб. пособие для студ. среднего проф. образования / Hogg, R. V., McKean, J. W., & Craig, A. T. Torino: Pearson Education. 2019. - 762 с.

23. Rice, J. A. Mathematical statistics and data analysis. [Текст]: учеб. пособие для студ. среднего проф. образования / Rice, J. A. Boston: Cengage Learning. 2006. - 685 с.

24. Casella, G., & Berger, R. L. Statistical inference. [Текст]: учеб. пособие для студ. среднего проф. образования / Casella, G., & Berger, R. L. Pacific Grove, CA: Duxbury. 2002. - 686 с.

25. Ross, S. M. Introduction to probability and statistics for engineers and scientists. [Текст]: учеб. пособие для студ. среднего проф. образования / Ross, S. M. New York: Academic Press. 2014. - 730 с.

26. Papoulis, A., & Pillai, S. U. (2002). Probability, random variables, and stochastic processes. [Текст]: учеб. пособие для студ. среднего проф. образования / Papoulis, A., & Pillai, S. U. New York: Tata McGraw-Hill Education. 2002. - 678 с.

27. Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). Mathematical statistics with applications. [Текст]: учеб. пособие для студ. среднего проф. образования / Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. Pacific Grove, CA: Brooks/Cole. 2008. - 939 с.

28. Mood, A. M., Graybill, F. A., & Boes, D. C. Introduction to the Theory of Statistics. [Текст]: учеб. пособие для студ. среднего проф. образования / Mood, A. M., Graybill, F. A., & Boes, D. C. New York: Tata McGraw-Hill Education. 1974. - 715 с.

29. Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. [Текст] / Ren, S., He, K., Girshick, R., & Sun, J. // Computer Science. 2015. №64. С. 45-53.

30. Redmon J, Divvala S, Girshick R, Farhadi F. You only look once: Unified, real-time object detection. [Teκcτ] / Redmon J, Divvala S, Girshick R, Farhadi F. // IEEE . 2016 №65. C. 779-788. 31. He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. [Текст] / He, K., Gkioxari, G., Dollár, P., & Girshick, R. // Computer Science. 2017. №76. C. 7-23.

32. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. [Teкct] / Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. // IEEE . 2017 №106. C. 2117-2125

33. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision. [Текст] / Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. & Zitnick, C. L. // Springer. 2014. №13. C. 740-755.