

Мониторинг и классификация вулканической активности с помощью компьютерного зрения

Volcanoes monitoring and classification using computer vision
algorithms

Лялина Инна

ОГЛАВЛЕНИЕ

Оглавление	1
1 Введение	2
2 Данные.....	3
3 Машинное обучение	5
3.1 Бинарная классификация с помощью алгоритмов машинного обучения.....	5
3.2 Бинарная классификации с помощью нейронных сетей.....	7
3.3 Множественная классификации с помощью нейронной сети.....	9
3.4 Классификация и сегментация изображения.....	10
3.5 Функция потерь	12
3.6 Метрики.....	14
4 Оценка моделей	16
5 Практическое применение.....	17
5.1 Мониторинг состояния	18
5.2 Количественные характеристики	18
5.3 Сохранение истории наблюдений	19
6 Заключение	20
7 Ссылки	21
8 Приложение 1.....	22
9 Приложение 2.....	25
10 Приложение 3.....	26

А Б С Т Р А К Т

Данное исследование посвящено применению алгоритмов машинного обучения в мониторинге и классификации вулканической активности по данным веб камер, установленных на вулканологических станциях. Далее рассматривается сравнение результатов использования методов компьютерного зрения на основе алгоритмов классического машинного обучения (логистическая регрессия, дерево решений) и сверточных нейронных сетей (ResNet, VGG, U-net).

1 ВВЕДЕНИЕ

Мониторинг вулканов требует обобщения и анализа большого объема разнородных данных. Требование к непрерывности и своевременности проведения исследований необработанных данных, полученных с помощью различных специализированных устройств, приводит к необходимости автоматизации процесса наблюдения.

Большинство вулканологических станций оснащено веб камерами, изображение с которых транслируется в сеть Интернет для непрерывного мониторинга вулканической активности и своевременного реагирования на изменение состояния. Ручная проверка видеоряда для отслеживания и фиксации состояния вулкана в рассматриваемом временном периоде может быть заменена на автоматическую классификацию и сегментацию изображения с использованием алгоритмов компьютерного зрения.

Помимо автоматического определения состояния вулкана в реальном времени, применение компьютерного зрения может помочь в фиксации изменений состояний и автоматическом формировании базы данных с параметрами активности вулкана, такие как направление и площадь выбросов, сокращая требования к системам хранения данных при ведении истории наблюдений.



Рисунок 1 Схема исследования

Данные для исследования представляют из себя массив фотографий нескольких вулканов, полученных из сети Интернет, в том числе изображения с веб камер Института вулканологии и сейсмологии ДВО РАН, публикуемые на геопортале geoportal.kscnet.ru.

Так как все изображения получены из разных источников, для обучения и прогнозирования определения состояния, изображение приводится к единому размеру 256x256 пикселей. Пропорции изображения при этом могут изменяться.

Для моделей классификации на основе классических алгоритмов машинного обучения изображения обрабатывались (применяется фильтр Собеля, приведение к черно белому виду) и вычислялись дополнительные признаки и дескрипторы изображения. В моделях классификации на основе сверточных нейронных сетей изображения не изменялись дополнительно. В модели сегментации дополнительно проводилась аугментация изображения.

Для обучения и проверки моделей массив данных был разделена на обучающую и валидационную выборку в пропорции 75% и 25% соответственно. Тестирование проводилось на отдельном наборе данных.

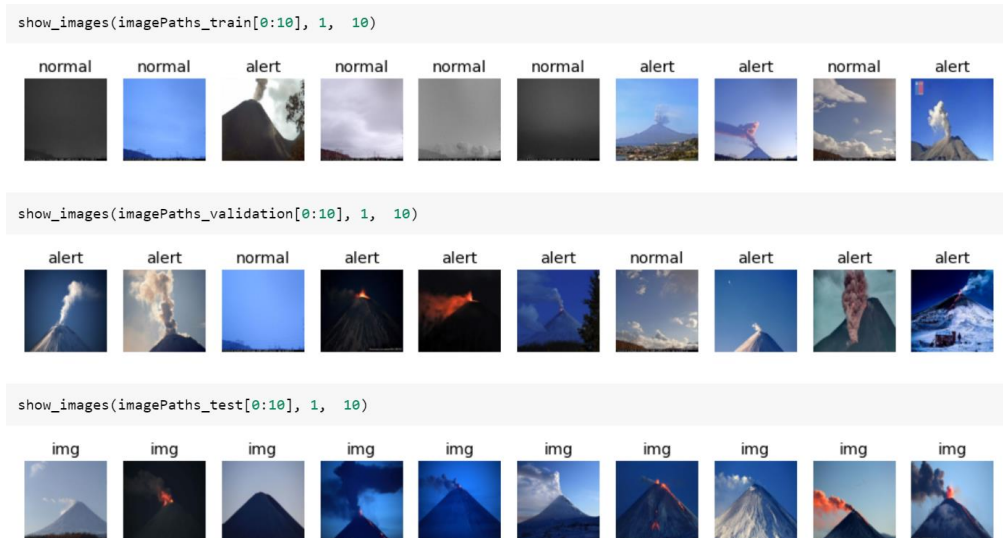


Рисунок 2 Входные данные (обучающая, валидационная, тестовая выборка)

Классификация данных проводилась по размеченным вручную классам:

- **класс состояние:**
 - 0 – внимание, зафиксирована активность вулкана,
 - 1 – нормальное состояние, либо видимость не позволяет проводить наблюдение
- **класс видимость:**
 - clear – ясно
 - cloud – облачно
 - mist – туман, сумерки, плохая видимость
- **класс выбросов - парогазовый столб (шлейф):**
 - 0 / 1 – не определено / определено наличие парогазовых выбросов
- **класс выбросов - лава:**
 - 0 / 1 - не определено / определено наличие лавовых потоков.

Для сегментации области изображения соответствующей классам выбросов на обучающей выборке подготовлен набор масок изображения. Маска изображения создавалась в два этапа:

- 1 этап на исходном изображении вручную задавалась область, принадлежащая заданному классу
- 2 этап автоматически проводилось изменение размера изображения, бинаризация, уточнение маски указанного сегмента с помощью алгоритма [GrabCut](#).

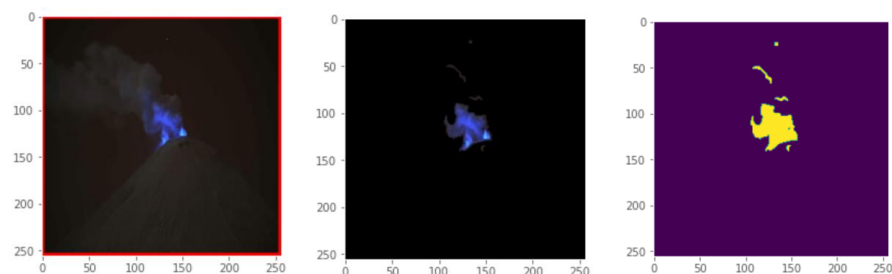


Рисунок 3 Создание маски изображения

В ходе анализа входных данных были применены разные способы обработки изображения, фильтрации, преобразования цветов, сглаживания, выделения контуров и областей, поиска якорных точек. На рисунке приводится пример выделения областей с помощью алгоритма [суперпиксель](#).

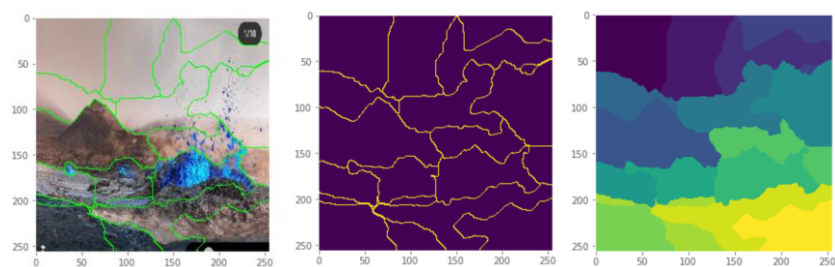


Рисунок 4 Пример выделения областей на изображении

Рассматривался так же метод выделения характерных точек, для создания масок и сегментации.

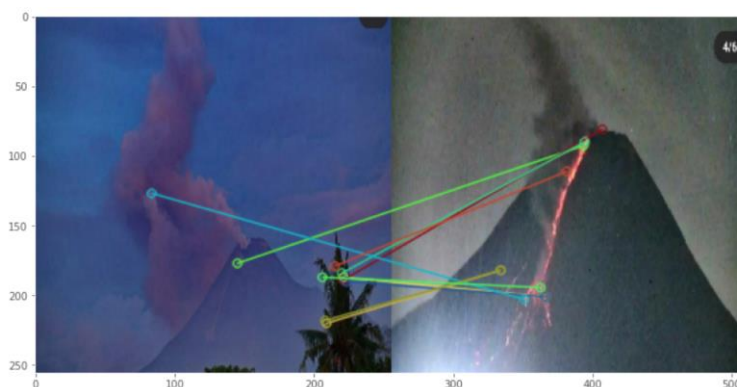


Рисунок 5 Пример поиска характерных точек

Но не использовался далее в моделях, так как целью было построение универсальной модели для любого вулкана с любой точки съемки.

Полезным оказалось применение фильтра Гауса, Лапласа для сглаживания шумов, определения градиента с помощью оператора Собеля с данными преобразованными в в схему HSV (1) и черно белую (2).

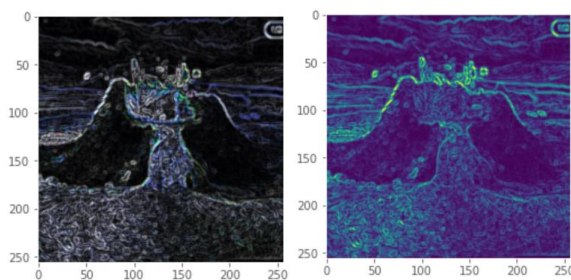


Рисунок 6 Преобразование входного изображения

Для построения дескрипторов изображения, которые будут использоваться в качестве признаков моделях классификации использовался алгоритм вычисления градиента:

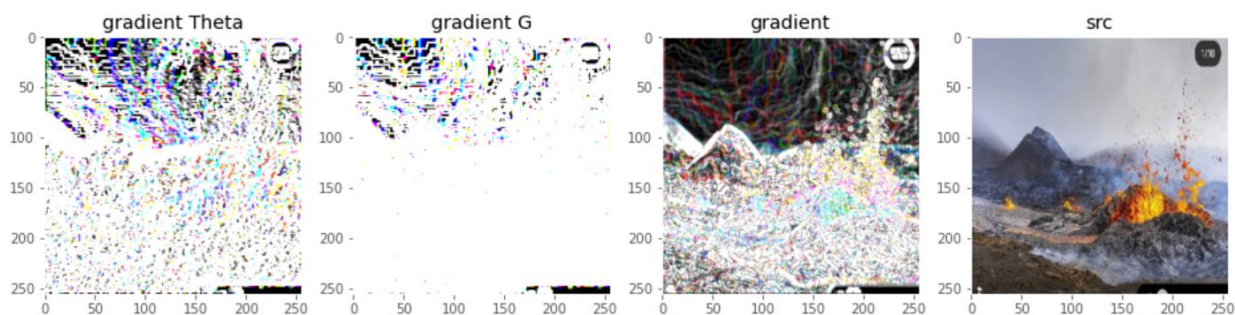


Рисунок 7 Пример дескриптора изображения

3 МАШИННОЕ ОБУЧЕНИЕ

Целью классификации изображения является предсказание вероятности принадлежности изображения к одному из заданных классов. В данной работе реализовано несколько алгоритмов классификации для определения возможности использования компьютерного зрения с обучением на небольшом наборе тестовых данных в промышленной эксплуатации.

- Бинарная классификации с помощью алгоритмов машинного обучения с предобработкой изображений
- Бинарная классификация с помощью нейронных сетей без дополнительной обработки изображений
- Множественная классификация с помощью нейронной сети для определения полного спектра состояний изображения

Рассмотрим подробнее примененные алгоритмы.

3.1 БИНАРНАЯ КЛАССИФИКАЦИЯ С ПОМОЩЬЮ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Рассматриваемые в данном разделе, модели машинного обучения в качестве входных признаков получали не изображение с веб камеры, а набор сгенерированных по данному изображению признаков. Генерация признаков выполнялась в несколько этапов:

1. Наложение фильтров на изображение: фильтр на основе распределения Гаусса для сглаживания и подавления шума, применение оператора Собеля для вычисления градиента и определения контуров на изображении.
2. Вычисление гистограммы изображения по цвету.
Цветовые гистограммы предоставляют подробную информацию о распределении цветов. С помощью разработанного экстрактора создается многомерная гистограмма изображения с вычисленными характеристиками для каждого канала.
3. Вычисление HOG дескриптора (альтернативный метод вычисления градиентов). Гистограмма направленных градиентов основана на подсчете количества направлений градиента в локальных областях изображения.

Для обучения модели использовались все три дескриптора. В итоговом датсете каждое изображение размера 256x256 представляется в виде набора из 1 017 124 признаков.

В качестве классификатора использовались следующие модели:

Логистическая регрессия (LR) – модель для прогнозирования вероятности принадлежности к одному из двух классов, является расширением линейной регрессии для моделирования дихотомическую (бинарной) переменной, который сообщает о вероятности наступления или не наступления события. Следовательно, чтобы использовать LR в качестве двоичного классификатора, необходимо назначить порог, чтобы различать два класса. Например, значение вероятности выше, чем 0,50 для входного экземпляра классифицирует его как «класс А» в противном случае - «класс В».

Алгоритм опорных векторов (SVM) можно классифицировать как линейные, так и нелинейные данные. Сначала он отображает каждый элемент данных в n -мерное пространство признаков, где n - количество функций. Затем он определяет гиперплоскость, которая разделяет элементы данных на два класса, максимально увеличивая расстояние для обоих классов и минимизируя ошибки классификации.

Дерево решений (DT) – данный метод моделирует логику принятия решений, исходя из значений переменных, полученных на вход. Каждый лист представляет собой значение целевой переменной, изменённой в ходе движения от корня по листу. Каждый внутренний узел соответствует одной из входных переменных. Все внутренние узлы (т.е. узлы, имеющие хотя бы один дочерний элемент) представляют из себя тесты для входных переменных. В зависимости от результата теста алгоритм классификации разветвляется к соответствующему дочернему узлу, где процесс проверки и ветвления повторяется, пока не достигнет конечного узла. Листовые узлы соответствуют результату решению.

Случайный лес (RF) является ансамблем решающих деревьев, каждое из которых само по себе даёт очень невысокое качество классификации, но за счёт их большого количества повышает результат классификации, другим словами выбирает классификацию с наибольшим количеством голосов (для бинарной классификации или среднее значение всех деревьев в лесу (для множественной классификации)).

Классификатор Bagging – это ансамбль классификаторов, который подбирает базовые классификаторы для каждого из случайных подмножеств исходного набора данных, а затем путем взвешивания или усреднения формирует окончательный прогноз. Комбинирование классификаторов позволяет дисперсию оценщика (например, дерева решений) за счет введения рандомизации в процедуру построения и последующего создания из него ансамбля.

Gradient Boosting - строит аддитивную модель с большим количеством моделей с учетом различных весов и позволяет оптимизировать произвольные функции потерь с поэтапным исправлением ошибки предыдущих этапов.

Ниже приведены результаты классификации некоторых из перечисленных моделей.

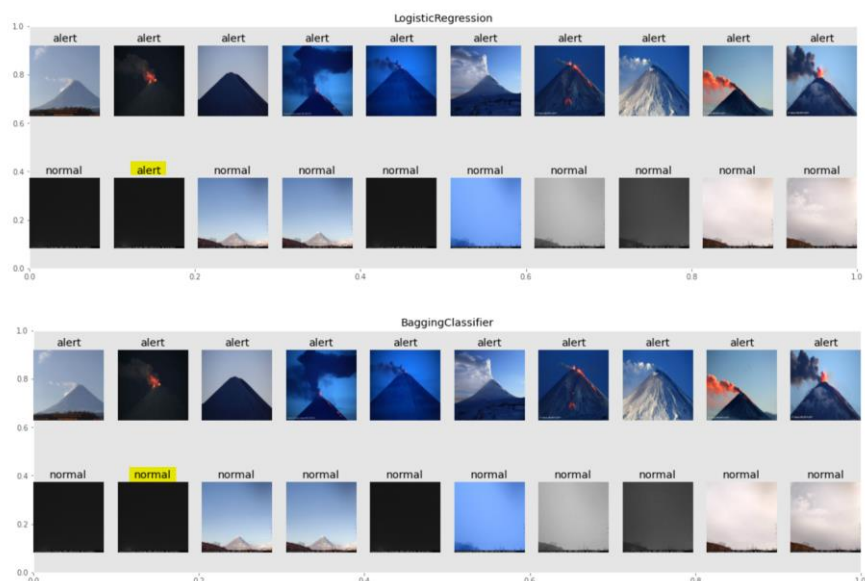


Рисунок 8 Результат работы LogisticRegression и BaggingClassifier

В качестве промежуточного заключения можно отметить, что данные модели дают прогнозы приемлемого качества с минимальными требованиями к аппаратной инфраструктуре и минимальной скорости обучения.

3.2 БИНАРНАЯ КЛАССИФИКАЦИИ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Бинарная классификация с помощью нейронных сетей применена для определения вероятности наличия на изображении признаков активности вулкана (класс состояние) и наличия выбросов (лава или парогазовый столб). В данном разделе проводилось сравнение качества предсказания модели сверточной нейронной сети построенной со случайной инициализацией и двух моделей сверточных нейронных сетей на основе предобученных сетей ResNet50 и VGG16 и весами [Imagenet](#).

Тип нейронной сети, называемый сверточными нейронными сетями (CNN), обычно используется для задачи компьютерного зрения. Эти нейронные сети используют тот факт, существует некоторая взаимосвязь между пикселями в различных изображениях одинакового класса. CNN могут быть разных типов архитектуры, которые определяются наличием разного вида и назначения слоев и входных параметров. Сверточный слой имеет нейроны расположены в трехмерном виде (ширина, высота и глубина), которые называются сверточными фильтрами. Глубина в данном контексте означает количество фильтров, а не количество слоев в нейронная сети.

Результат этой операции с помощью функции активации дает реакцию фильтра в каждой пространственной позиции. За выходом сверточного слоя обычно следует активационный слой или нелинейная функция (например, ReLu, сигмовидная или tanh-функция). Интуитивно сеть изучает части изображения (такие как края, цвета или форм) через фильтры, которые активируются определенными функциями. Качество такого изучения оценивается с помощью функции потерь. Для снижения размерности добавляются пулинговые слои, для сокращения переобучения – слои регуляризации пулинговый слой сокращает размер изображения за счет выбора максимального значения пикселя из рассматриваемого окна. Слой регуляризации в случайном порядке «выключает» некоторые параметры изображения. Последним слоем является полносвязный слой признаков, который выводит оценку для каждого из классов, т.е. последний слой CNN содержит N оценок входного изображения, где N – желаемое количество разных классов. Оценка представляет собой вероятность того, что входные данные принадлежат каждому классу.

Простая модель CNN представляла из себя сверточную сеть, на вход которой подается изображение 256x256. Далее применяются несколько слоев свертки, пулинга, сглаживания шума, регуляризации, нормализации и в итоге полносвязный слой с сигмоидальной функцией активации и общими параметрами сети в количестве 1, бмлн.

Так же для сравнения было построено несколько моделей сверточной сети с применением transfer learning на основе 50 слойной сети ResNet и 19 слойной сети VGG и весами Imagenet.

conv5_block3_2_conv (Conv2D)	(None, 8, 8, 512)	2359888	conv5_block3_1_relu[0][0]
conv5_block3_2_bn (BatchNormali	(None, 8, 8, 512)	2048	conv5_block3_2_conv[0][0]
conv5_block3_2_relu (Activation	(None, 8, 8, 512)	0	conv5_block3_2_bn[0][0]
conv2d_8 (Conv2D)	(None, 8, 8, 128)	589952	conv5_block3_2_relu[0][0]
conv2d_9 (Conv2D)	(None, 8, 8, 128)	147584	conv2d_8[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, 4, 4, 128)	0	conv2d_9[0][0]
dropout_2 (Dropout)	(None, 4, 4, 128)	0	max_pooling2d_1[0][0]
conv2d_10 (Conv2D)	(None, 4, 4, 64)	73792	dropout_2[0][0]
conv2d_11 (Conv2D)	(None, 4, 4, 64)	36928	conv2d_10[0][0]
max_pooling2d_2 (MaxPooling2D)	(None, 2, 2, 64)	0	conv2d_11[0][0]
dense_1 (Dense)	(None, 2, 2, 32)	2080	max_pooling2d_2[0][0]
dense_2 (Dense)	(None, 2, 2, 32)	1056	dense_1[0][0]
flatten_1 (Flatten)	(None, 128)	0	dense_2[0][0]
batch_normalization_3 (BatchNor	(None, 128)	512	flatten_1[0][0]
dense_3 (Dense)	(None, 1)	129	batch_normalization_3[0][0]

Total params: 23,380,929			
Trainable params: 851,777			
Non-trainable params: 22,529,152			

Рисунок 9 Архитектура сети на базе ResNet50

В одном случае для указанных моделей transfer Learning были добавлены только последний полносвязный слой для бинарной классификации. Во других для сетей ResNet и VGG применялся механизм замораживания слоев в базовых нейронных сетях, без обновления весов и добавления новых слоев для fine-tuning.

block4_conv4 (Conv2D)	(None, 32, 32, 512)	2359888
block4_pool (MaxPooling2D)	(None, 16, 16, 512)	0
block5_conv1 (Conv2D)	(None, 16, 16, 512)	2359888
conv2d_40 (Conv2D)	(None, 16, 16, 128)	589952
conv2d_41 (Conv2D)	(None, 16, 16, 128)	147584
max_pooling2d_17 (MaxPooling	(None, 8, 8, 128)	0
dropout_10 (Dropout)	(None, 8, 8, 128)	0
conv2d_42 (Conv2D)	(None, 8, 8, 64)	73792
conv2d_43 (Conv2D)	(None, 8, 8, 64)	36928
max_pooling2d_18 (MaxPooling	(None, 4, 4, 64)	0
dense_27 (Dense)	(None, 4, 4, 32)	2080
dense_28 (Dense)	(None, 4, 4, 32)	1056
flatten_14 (Flatten)	(None, 512)	0
batch_normalization_16 (Bate	(None, 512)	2048
flatten_15 (Flatten)	(None, 512)	0
batch_normalization_17 (Bate	(None, 512)	2048
dense_29 (Dense)	(None, 5)	2565

Total params: 13,803,013		
Trainable params: 856,005		
Non-trainable params: 12,947,008		

Рисунок 10 Архитектура сети на базе VGG19

Качество классификации для классов выбросов можно сравнить на графиках:

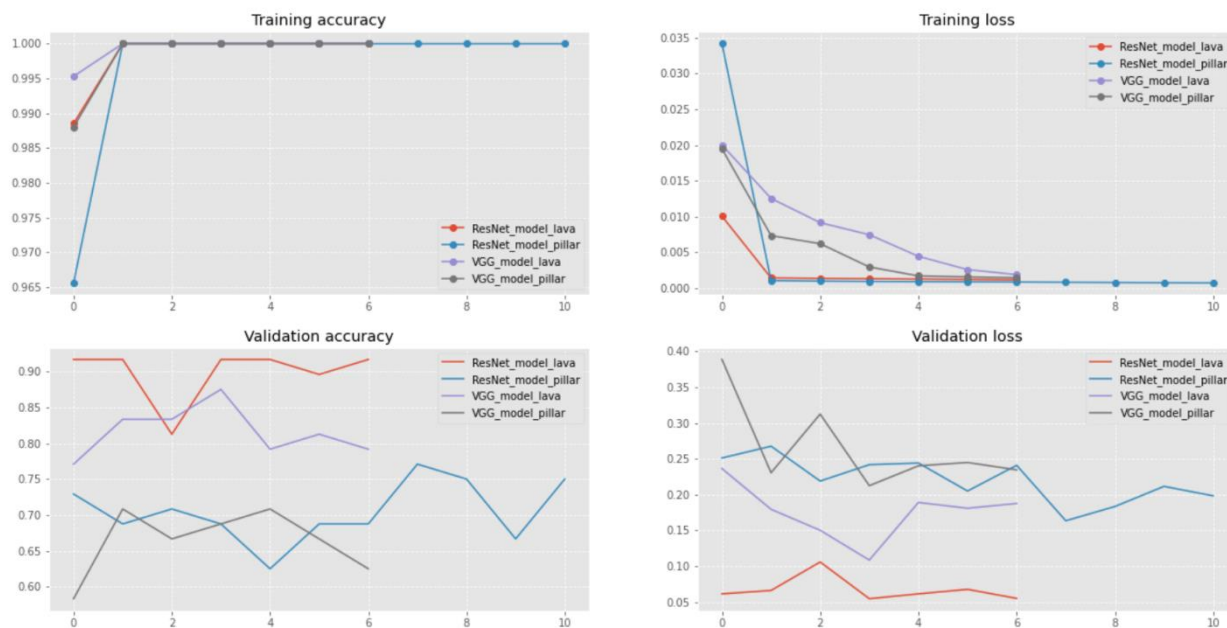


Рисунок 11 Оценка качества обучения моделей

Результат бинарной классификации:



Рисунок 12 Вероятность принадлежности к классу

3.3 Множественная классификации с помощью нейронной сети.

Множественная классификация рассматривается в данной работе для оценки возможности определения полного спектра возможных состояний вулкана по изображению и может применяться для того, чтобы получить общее описание изображения, преобразовав таким образом входные данные в дискретные/структурированные.

По изображению определяется классы выбросов (лава, шлейф) и класс видимость.

- clear – ясно
- cloud – облачно
- mist – туман, сумерки, плохая видимость

Так как классы не сбалансированы и объединены в одну модель классификации были добавлены веса. Значимость классов выбросов больше чем класса видимости.

В моделях множественной классификации также применялся методы transfer Learning и freezing. В качестве инициализации использовались веса Imagenet. Использование только многоклассовой кросс-энтропии давала очень большую погрешность, поэтому было добавлено MSE. При отключении функции call back значение метрики val-loss уменьшалось, но прогноз вероятности классов распределялся равномерно для всех изображений. Ситуация характерна как для модели на основе ResNet так и VGG.

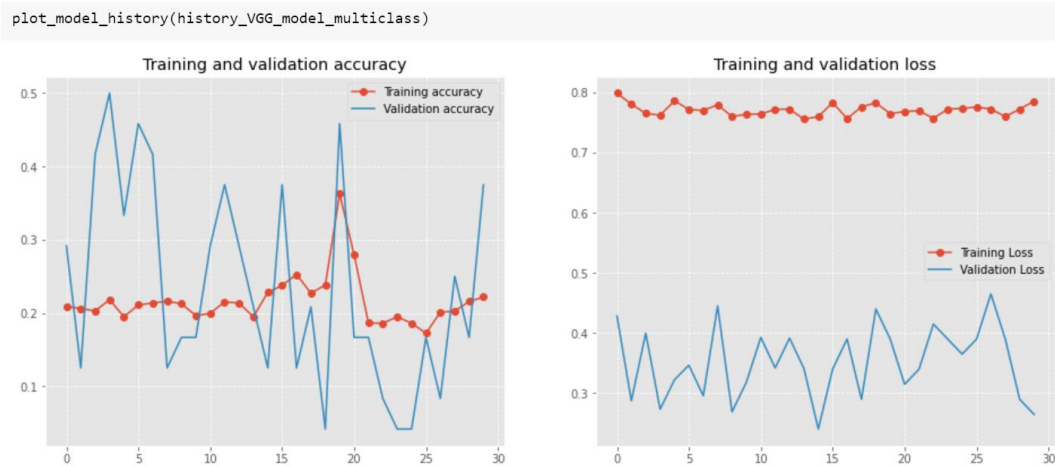


Рисунок 13 Обучение модели множественной классификации

Пример переобучения модели:



Рисунок 14 Пример переобучения модели

3.4 КЛАССИФИКАЦИЯ И СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЯ

Для выделения сегментов на изображении относящихся к одному из классов выбросов была применена сверточная нейронная сеть, за основу которой взята архитектура U-net [1]. U-Net была изначально построена для работы с обучающими выборками малого объема. U-Net является полносверточной нейронной сетью, т. е. принимает на вход изображение целиком и возвращает его полную карту сегментации. Сеть состоит из сверточной(слева) и разверточной части (справа). Рис.

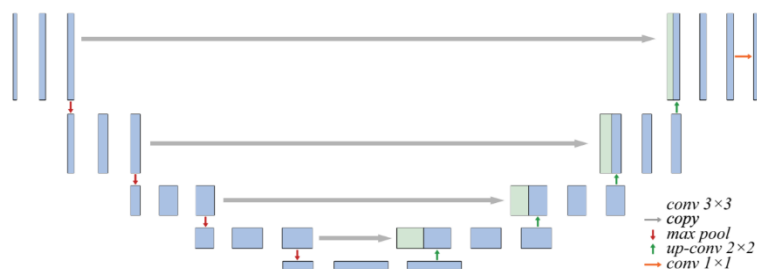


Рисунок 15 Архитектура базовой нейронной сети U-net

Сверточная часть U-net похожа на обычную свёрточную сеть, она содержит два подряд свёрточных слоя 3×3 , после которых идет слой ReLU и пулинг с функцией максимума 2×2 с шагом 2. Каждый шаг разверточной части содержит слой, обратный пулингу, который расширяет карту признаков, после которого следует свертка 2×2 , которая уменьшает количество каналов признаков. После идет конкатенация с соответствующим образом обрезанной картой признаков из сжимающей части и две свертки 3×3 . Обрезка нужна из-за того, что мы теряем пограничные пиксели в каждой свёртке. На последнем слое свертка 1×1 используется для приведения каждого 64-компонентного вектора признаков до требуемого количества классов.

Сначала входное изображение пропускается через свёрточный слой с фильтрами размера 3×3 пикселя. Далее применяется функция активации ReLU ($f(x) = \max(0, x)$). Изображение пропускается через указанную последовательность слоёв еще несколько раз, пока разрешение не достигнет минимального. На расширяющем пути применяется та же последовательность свёрточных фильтров, перемежающаяся операциями Upsampling, повышающими разрешение изображения вдвое по каждой стороне за счет размножения каждого из пикселей в четыре пикселя (в два по каждой стороне).

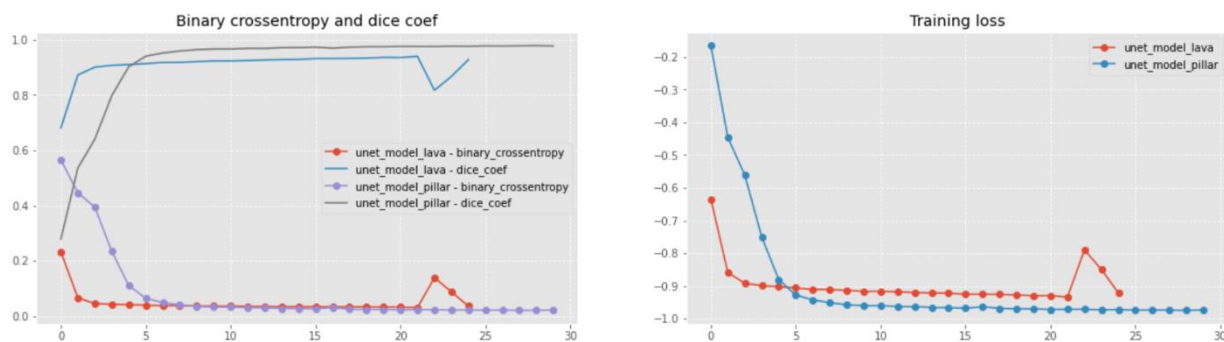


Рисунок 16 Оценка качества обучения модели для сегментации объектов

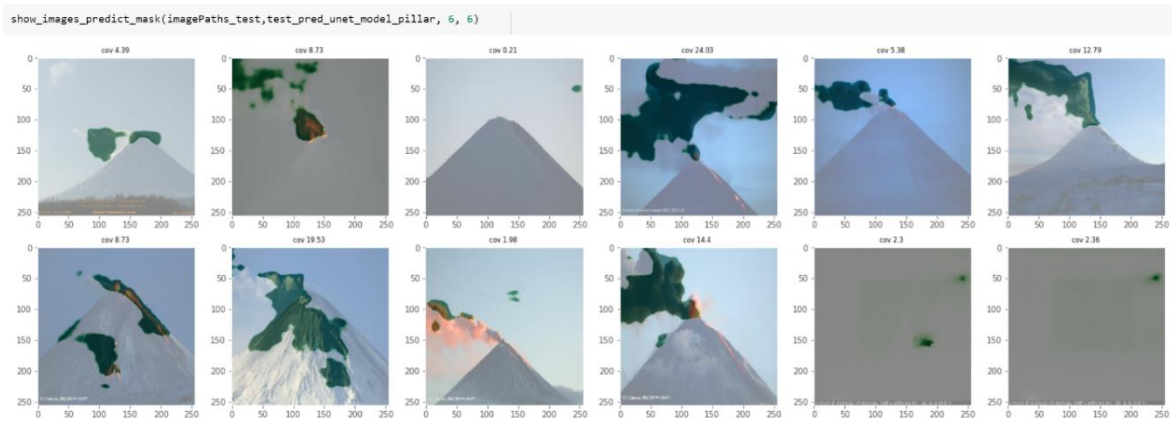


Рисунок 18 Результат сегментации участков шлейфа с помощью нейронной сети

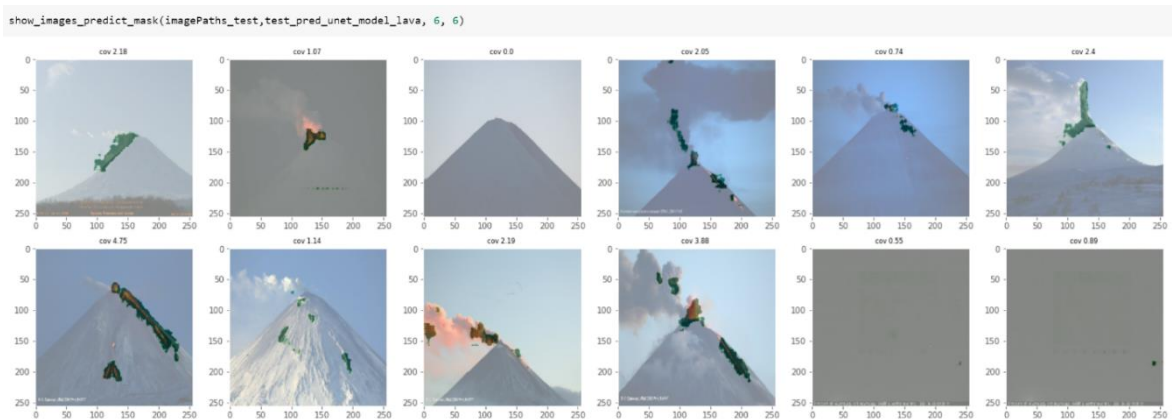


Рисунок 17 Результат сегментации участков лавы с помощью нейронной сети

Предварительные результаты показывают, что более эффективно использовать сегментацию участков с соответствующими классами выбросы для изображений, которые уже прошли бинарную классификацию и отмечены как принадлежащие данному классу. В моделях на классификацию точность выше, поэтому если на картинке выделены сегментированные области, надо проверять определен ли этот класс в предыдущих моделях.

3.5 Функция потерь

Цель обучения любой модели машинного обучения минимизация расхождения между предсказанными значениями и истинными значениями. Выбор функция потерь является важным решением алгоритма и должна зависеть от проблемы и данных. В данной работе выбраны часто используемые функции потерь - это среднеквадратичная ошибка и бинарная кросс-энтропия для сравнения результатов обучения с базовыми параметрами. В моделях по сегментации для решения проблемы несбалансированности данных применялась сочетание категориальной кросс-энтропии и обобщённой функции потерь Дайса.

Среднеквадратичная ошибка (MSE) измеряет среднюю сумму квадратной разности между фактическим значением и прогнозируемым значением для всех точек данных:

$$MSE = \frac{1}{n} \sum_{i=0}^n (y_i - x_i)^2$$

MSE имеет некоторые преимущества:

- Она подчеркивает большие ошибки над меньших ошибках.
- Является дифференцируемым, что позволяет более эффективно использовать для поиска минимальных или максимальных значений с помощью математических методов.

Кросс-энтропийная (CCE) функция потерь часто используется в задачах семантической сегментации изображений [9]. Выходом функции является значение вероятности в диапазоне от 0 до 1 для каждой категории (класса). Значение кросс-энтропийной функции потерь увеличивается, когда прогнозируемая вероятность отклоняется от целевой метки:

$$CCE = - \sum_{i=0}^n (x_i * \log(y_i))$$

Здесь $y_i = 1$, когда объект принадлежит к классу i , и $y_i = 0$ иначе, а x_i – предсказанная вероятность принадлежности объекта к классу i .

В **бинарной классификации** (BCE), где количество классов равно двум, кросс-энтропия может быть посчитана так:

$$BCE = -(y \ln p + (1 - y) \ln(1 - p))$$

где $y = 0$ для объекта первого класса и $y = 1$ для второго, p – вероятность того, что объект принадлежит ко второму классу

Обобщённая **функция потерь Дайса** (Dice loss, DL) построена на основе меры сходства между множествами, известной как коэффициент Дайса–Сёренсена:

$$DL = 1 - 2 \frac{\sum y_i p}{\sum (p_i + y_i)}$$

где $\sum y_i$ – это сумма y_i по всему изображению, то есть количество пикселей класса i . Благодаря такому взвешиванию сеть лучше обучается на объектах редко встречающихся классов, так как вес ошибок на них в функции потерь увеличивается.

Функция потерь Дайса может быть рассмотрена как мощность пересечения множеств:

$$Dice(X, Y) = 1 + \frac{2 |X \cap Y|}{|X| + |Y|}$$

X – предсказание, Y истинной значение (подготовленная маска для изображении) $|X|$ - мощность множества X (количество элементов в этом множестве) и P пересечение между X и Y . Т.е. X построенная алгоритмом карта сегментации (возможно, содержащая вещественные значения), Y — корректная карта сегментации (с бинарными значениями). В модели пересечение реализуется как умножение, и мощность в виде `sum()` по axis 1 (сумма из трех каналов) заключается в том, что предсказания и цель являются one-hot encoded векторами. Расширенная функция потерь для сегментации:

$$BCE\ Dice = 0.2 * BCE - Dice$$

К функции добавлен коэффициент 1 для обратного распространения. Такое расширение функции (BCE Dice) позволяет вычислять градиент функции потерь и делает возможным обучение нейронной сети. Основное преимущество функции Dice перед кросс-энтропийной функцией потерь, что функция Dice не зависит от соотношения между площадью объекта и фона.

Во всех моделях применялся **оптимизатор Adam** (Адаптивная оценка момента). Данный алгоритм оптимизации вычисляет экспоненциальное скользящее среднее градиента и квадратичный градиент, а параметры β_1 и β_2 управляют скоростью затухания этих скользящих средних. К его преимуществам относятся:

- Простая реализация.
- Вычислительная эффективность.
- Небольшие требования к памяти.
- Подходит для задач с очень шумными или разреженными градиентами.
- Гиперпараметры имеют наглядную интерпретацию и обычно не требуют долгой настройки.

Функции активации полносвязного слоя в моделях:

- в случае бинарной классификации использовалась функция активации **Sigmoid**:

$$f(x_i) = \frac{1}{1 + e^{x_i}}$$

- в случае множественной классификации использовалась функция активации **Softmax**:

$$f(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

3.6 МЕТРИКИ

Для оценки качества модели использовались следующие метрики. Для бинарной классификации - стандартные метрики Accuracy, Recall, Confusion Matrix. Для оценки качества сегментации дополнительно использовался коэффициент Dice, как мера сходства двух множеств.

Accuracy

Базовая метрика, которая измеряет количество верно классифицированных объектов относительно общего количества всех объектов.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Так как она не идеальна для несбалансированных классов, где может быть много экземпляров одного класса и мало другого. В некоторых случаях применялись дополнительные метрики.

Recall/Sensitivity

Количество объектов правильно классифицированных с позитивной меткой из всего множества позитивных.

$$recall = \frac{TP}{TP + FN}$$

Precision

Количество объектов, которые классифицируются как положительные, и действительно являются положительными, относительно общего количества полученных в модели позитивных меток.

$$precision = \frac{TP}{TP + FP}$$

F1 score

Сочетание precision и recall, дает некоторый компромисс между ними двумя: 1 –наилучшее, 0 – наихудшее качество модели

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

Confusion Matrix

Для некоторых моделей для описания точности классификатора использовалась матрица ошибок, которая представляется в виде таблицы:

		Prediction outcome		
		positive	negative	
Actual value	positive	TP	FN	TP + FN
	negative	FP	TN	FP + TN
		TP + FP	FN + TN	

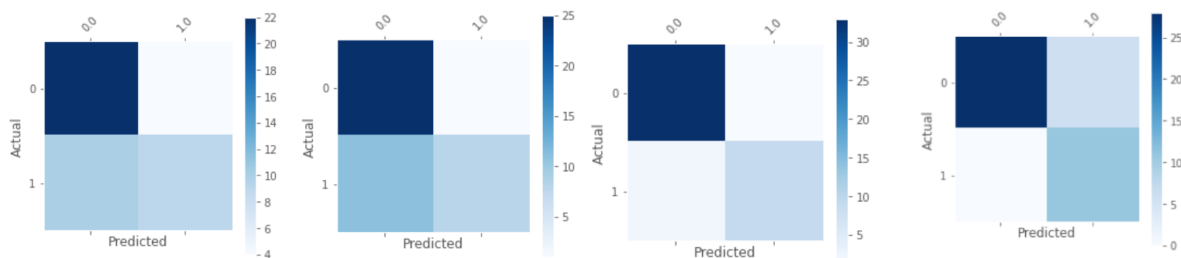


Рисунок 19 Confusion matrix для сверточных сетей с бинарной классификацией (слева направо) :
ResNet_model_pillar, VGG_model_pillar, ResNet_model_java, VGG_model_java

Для сравнения качества бинарной классификации на основе алгоритмов машинного обучения так же использовалась кривая ROC и показатель AUC (англ. area under ROC curve, площадь под ROC-кривой) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций.

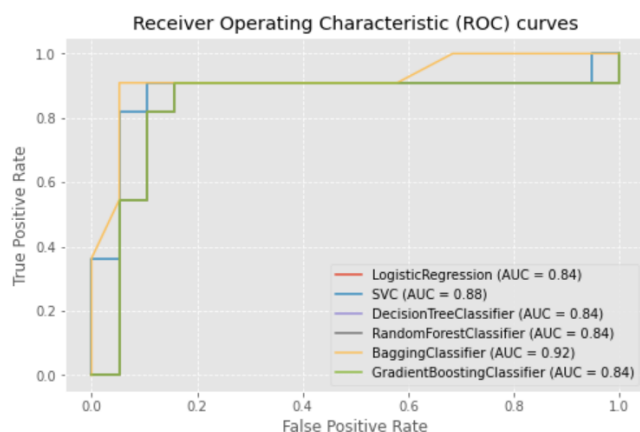


Рисунок 20 Сравнение классификаторов по ROC AUC

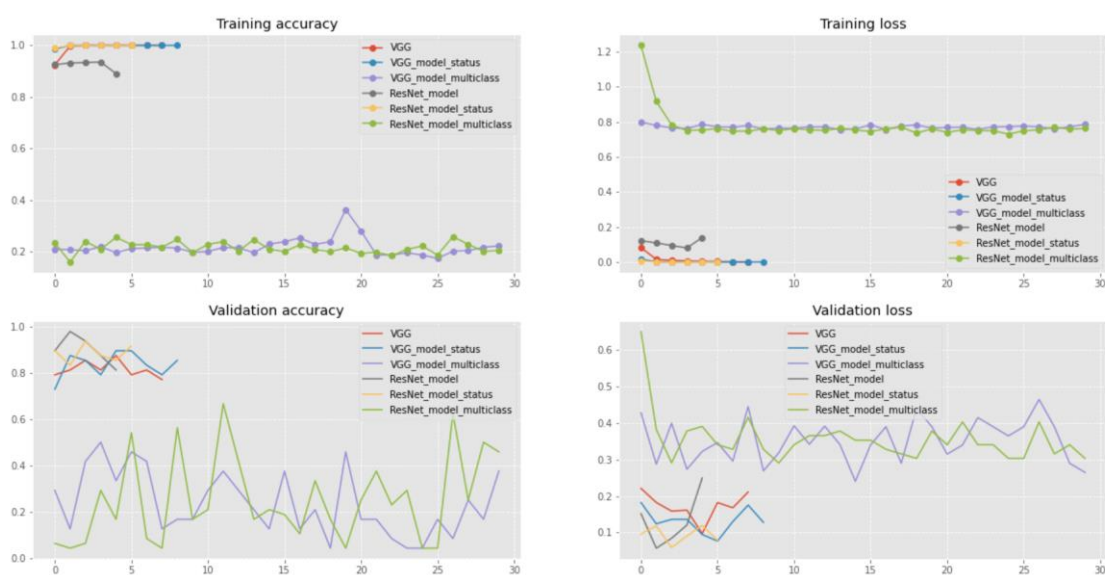


Рисунок 21 Сравнение качества моделей на основе нейронных сетей

4 ОЦЕНКА МОДЕЛЕЙ

Модели на основе алгоритмов машинного обучения по качеству сопоставимы с моделями на основе нейронных сетей. Принимая во внимание скорость обучения и размер файла модели, наиболее предпочтительными из них выглядят линейные классификаторы: логистическая регрессия и классификатор на основе метода опорных векторов. Из группы классификаторов на основе деревьев решений хорошую точность показал Bagging, но требует больше времени на обучение.

Среди моделей на основе сверточных нейронных сетей явных лидеров по точности не выявлено, точность сильно зависит от сбалансированности классов и обучающей выборки. Сеть без инициализации не выводится в итоговый результат так как показала низкую точность. Сети на основе предобученных сетей ResNet и VGG с весами Imagenet показали приемлемый результат бинарной классификации. Множественная классификация на основе сверточных сетей требует более глубоких настроек и может использоваться в базовом виде только для очень обобщенного описания состояния наблюдаемого объекта.

Сверточная сеть для сегментации изображений показала приемлемый результат согласно метрики пересечения нужного сегмента и предсказанной маски. Тем не менее в промышленной эксплуатации для более точного определения наличия класса изображения для его последующей сегментации правильнее будет использовать комбинацию unet сети с одним из бинарных классификаторов описанных выше. Обусловлено тем, что на фото часто бывают схожие текстуры, например, облако и парогазовый столб или участки снега яркого снега среди камней и участки ярких пятен лавы среди камней.

Итоговые оценки качества обученных моделей классификации и сегментации приведены в таблице:

Модель	Accuracy	Recall	Precision	F1 score	Binary crossentropy	Dice coef	Размер файла модели, Мб
Machine Learning Algorithm (on validation dataset)							
Logistic Regression	0.86	0.91	0.88	0.87			8
SVM	0.87	0.91	0.88	0.87			528
Decision Tree	0.8	0.82	0.81	0.8			0.002
Random Forest	0.87	0.81	0.87	0.87			0.067
Bagging	0.93	0.91	0.93	0.93			233
Gradient Boosting	0.8	0.82	0.81	0.8			0.12
Convolutional Networks (on test dataset)							
ResNet_model_status	0.89	0.84	0.9	0.89			96
VGG_model_status	0.91	0.84	0.93	0.91			59
ResNet_model_pillar	0.68	0.47	0.69	0.68			96
VGG_model_pillar	0.73	0.42	0.78	0.71			78
ResNet_model_lava	0.93	0.81	0.93	0.93			96
VGG_model_lava	0.82	0.91	0.87	0.83			78
ResNet_model_multiclass	0.67	0	0.48	0.55			96
VGG_model_multiclass	0.65	0.07	0.51	0.56			59
UNet_model_lava	0.78	0.09	0.83	0.7	0.04	0.93	89
UNet_model_pillar	0.73	0.58	0.73	0.73	0.02	0.98	89

Таблица 1 Результат оценки качества моделей

5 ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

В результате исследования было определено, что алгоритмы машинного обучения и компьютерного зрения могут использоваться для автоматической обработки изображения с веб камер для нескольких целей:

1. Мониторинг состояния и уведомление об изменении состояния вулкана
2. Вычисление предварительных количественных характеристик
3. Сохранение истории наблюдений в структурированном виде

5.1 МОНИТОРИНГ СОСТОЯНИЯ

Полученные оценки качества модели говорят о том, что модель может определять наличие извержения вулкана, наличия выбросов в виде лавы или парогазового столба по изображению с веб камеры с точностью 86%-90%. Такая точность позволяет использовать данные модели в мониторинге вулканической деятельности. В случае изменения состояния наблюдаемого объекта и определения соответствующего класса система может рассылать уведомление (электронная почта, Telegram) с изображением и описанием изображения сгенерированного автоматически. На рисунках ниже представлены примеры описания состояния объекта по изображению.

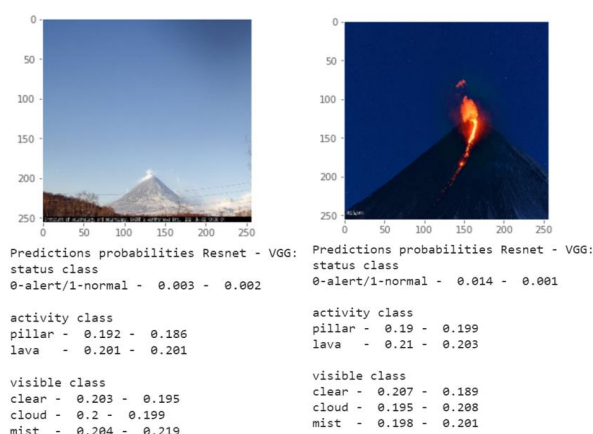


Рисунок 23 Пример уведомления о состоянии вулкана

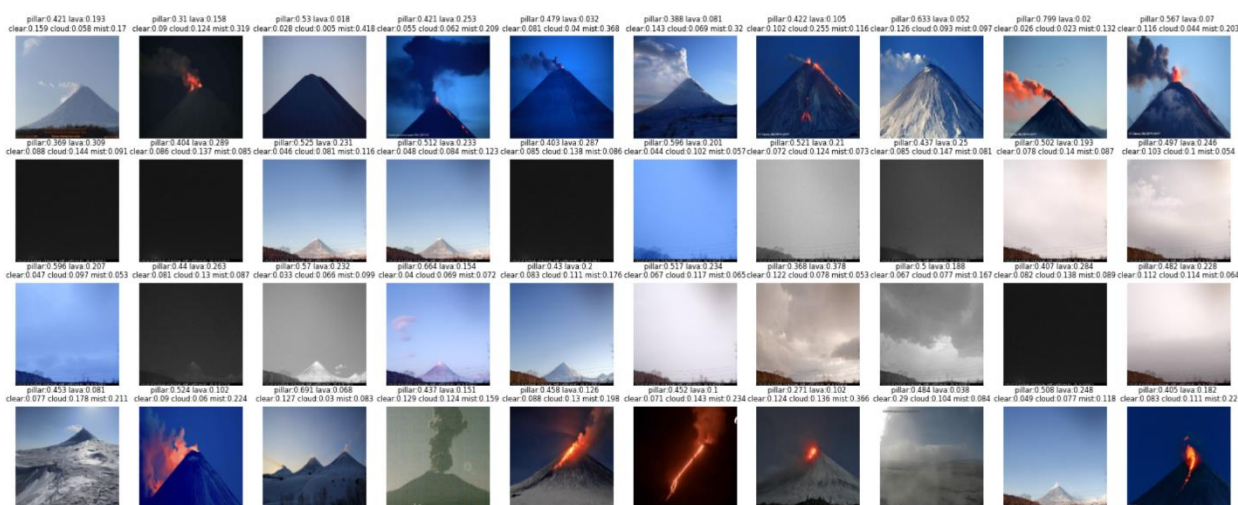


Рисунок 22 Пример описания состояния вулкана по фото (класс: вероятность наличия класса)

5.2 КОЛИЧЕСТВЕННЫЕ ХАРАКТЕРИСТИКИ

Возможность определения количественных характеристик (Quantification) классифицированных и сегментированных объектов по изображению одна из самых важных функций автоматического анализа данных. После классификации изображения и выделения участков, относящихся к рассматриваемому классу, для полученного сегмента (маски класса) может быть рассчитаны новые признаки и на основе этих признаков проведена категоризация. Например, для участка изображения, определенного как лавовый поток, проводится дополнительная обработка: выделение контуров маски, расчет количества сегментов в маске, расчет центра для

каждого сегмента, выделение очага по насыщенности и яркости пикселей. Затем вычисляется площадь полученных сегментов относительно площади всего изображения. Может быть рассчитана площадь очага относительно площади лавового потока. По полученной площади исходное изображение может быть отнесено к соответствующей категории извержения.

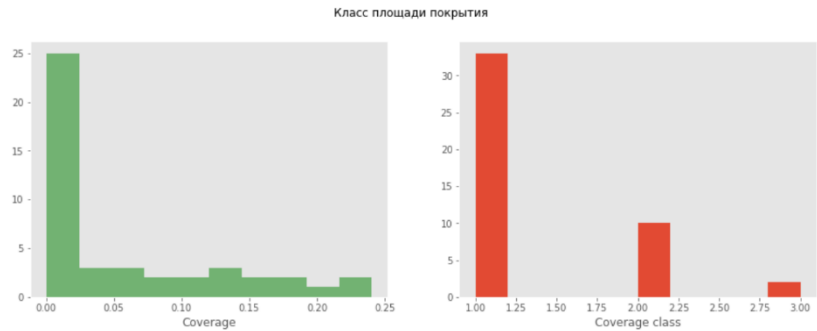


Рисунок 24 Категоризация по предсказанной маске изображения

Помимо площади сегмента (верхний левый угол) для изображения может быть рассчитана площадь очага (верхний правый угол):

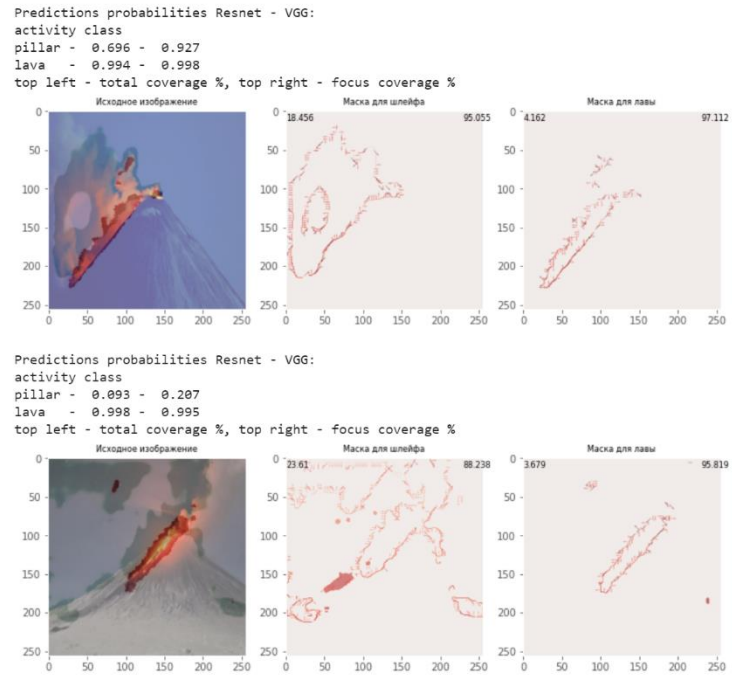


Рисунок 25 Пример уведомления о наличии участков лавы и шлейфа и рассчитанные площади

5.3 СОХРАНЕНИЕ ИСТОРИИ НАБЛЮДЕНИЙ

Еще одной целью выполнения автоматического анализа изображений в ходе мониторинга вулканической активности может быть сохранение истории наблюдений в структурированном виде в БД или текстовых файлах. Текстовый файлы занимают значительно меньше места на диске и могут хранить информацию о состоянии за длительный период для проведения дополнительного анализа в будущем. По определённой логике собираются данные об объекте наблюдения, при наличии какой-либо активности записывается исходное базовое

изображение, а дальше по ходу наблюдения в систему наблюдения записываются только рассчитанные количественные характеристики для отслеживания динамики изменений.

Одним из примеров практического использования может быть отслеживание направления движения потоков лавы или шлейфа. С помощью обработки маски класса, полученной на этапе сегментации, вычисляются градиенты и визуализируются в виде гистограмм. Рассчитанные характеристики хранятся в текстовых файлах и могут быть визуализированы или обработаны в дальнейшем в новой логике.

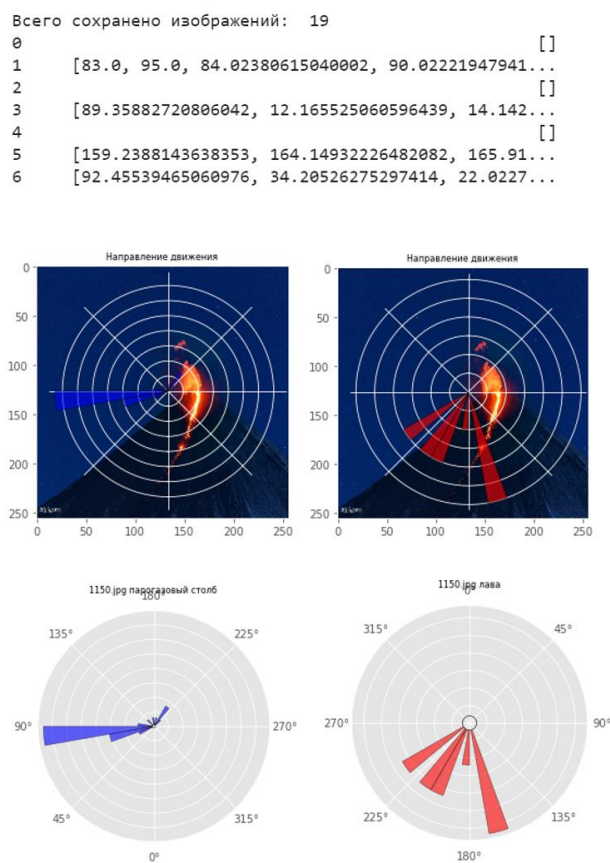


Рисунок 26 Пример сохранения результатов предсказания модели

6 ЗАКЛЮЧЕНИЕ

В последнее время установка веб камер для наблюдения за различными объектами, в том числе вулканами получила широкое распространение. Тем не менее, зачастую мониторинг с помощью установленных веб-камер предполагает постоянное наблюдение человека. Кроме того, не всегда данные полученные подобным методом автоматически обрабатываются и используются как инструмент количественной оценки.

Рассмотренные в данной работе алгоритмы машинного обучения и методы компьютерного зрения могут быть применены в качестве дополнительных способов видеонаблюдения на вулканах. Модели, обученные на небольшом наборе исходных данных, показали точность предсказания от 86% до 93%. Реализованные в данной работе модели не привязаны к определённой камере и объекту наблюдения и могут быть применены к другим вулканам и локациям. Нейронные сети обучались по данным с нескольких камер и вулканов, что может говорить об универсальности модели, однако, для большей точности расчета количественных характеристик, целесообразно настраивать модель под параметры конкретной веб камеры. Выбросы выявленные в ходе фото-видео -фиксации могут быть оцифрованы и сохранены для изучения динамики в дальнейшем.

1. U-Net: Convolutional Networks for Biomedical Image Segmentation
Olaf Ronneberger, Philipp Fischer, Thomas Brox <https://arxiv.org/abs/1505.04597>
2. Comparing image features and machine learning algorithms for real-time parking space classification
M. Tschentscher¹, M. Neuhausen¹, C. Koch², M. König², J. Salmen¹, M. Schlipsing¹
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.708.1741&rep=rep1&type=pdf>
3. A Comparison of Deep Learning Algorithms for Medical Image Classification and Image Enhancement
Carina Pereira
https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/43294/Pereira_washington_0250O_19533.pdf
4. Machine learning for blob detection in high-resolution 3D microscopy images
Martin Ter Haak <https://kth.diva-portal.org/smash/get/diva2:1232431/FULLTEXT01.pdf>
5. The Hessian Blob Algorithm: Precise Particle Detection in Atomic Force Microscopy Imagery
Brendan P. Marsh, Nagaraju Chada, Raghavendar Reddy Sanganna Gari, Krishna P. Sigdel & Gavin M. King
<https://www.nature.com/articles/s41598-018-19379-x>
6. Image Classification Algorithm Based on Deep Learning-Kernel Function
Jun-e Liu and Feng-Ping An <https://downloads.hindawi.com/journals/sp/2020/7607612.pdf>
7. A Comparative Study of Image Classification Algorithms for Foraminifera Identification
Boxuan Zhong, Qian Ge, Bhargav Kanakiya, Ritayan Mitra, Thomas Marchitto and Edgar Lobaton
http://instaar.colorado.edu/~marchitt/SSCI_2017_Forams_Classification_final.pdf
8. Семантическая сегментация спутниковых снимков аэропортов с помощью свёрточных нейронных сетей
В.А. Горбачёв¹, И.А. Криворотов^{1,2}, А.О. Маркелов^{1,2}, Е.В. Котлярова²
<http://computeroptics.ru/KO/PDF/KO44-4/440418.pdf>

Predictions probabilities Resnet - VGG:

activity class

pillar - 0.901 - 0.893

lava - 0.017 - 0.522

top left - total coverage %, top right - focus coverage %



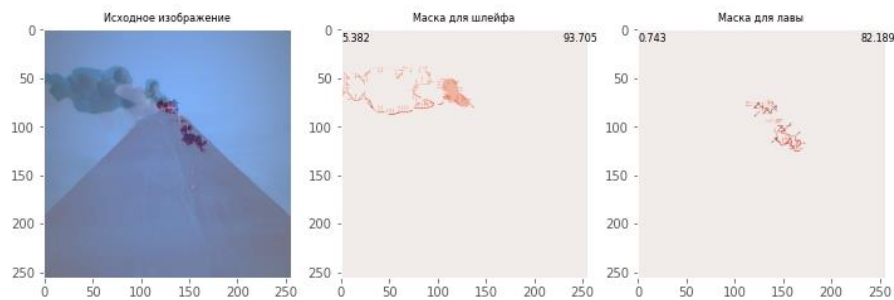
Predictions probabilities Resnet - VGG:

activity class

pillar - 0.173 - 0.162

lava - 0.999 - 0.915

top left - total coverage %, top right - focus coverage %



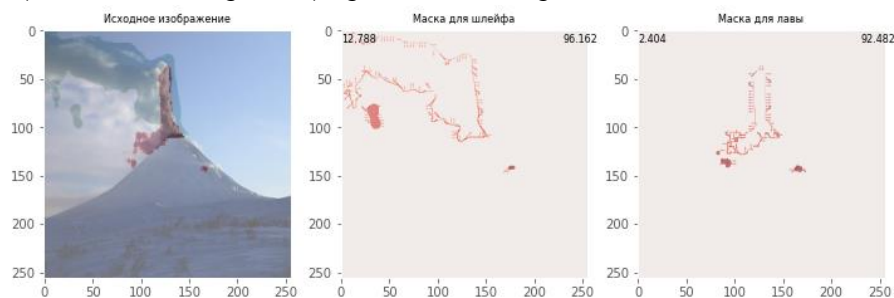
Predictions probabilities Resnet - VGG:

activity class

pillar - 0.977 - 0.817

lava - 0.243 - 0.974

top left - total coverage %, top right - focus coverage %



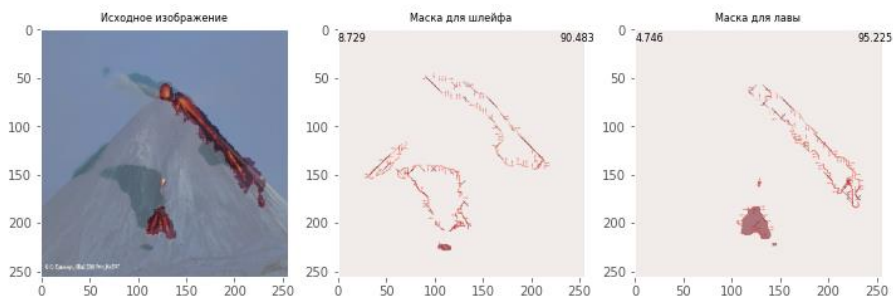
Predictions probabilities Resnet - VGG:

activity class

pillar - 0.337 - 0.081

lava - 0.981 - 1.0

top left - total coverage %, top right - focus coverage %



Predictions probabilities Resnet - VGG:

activity class

pillar - 0.964 - 0.97

lava - 0.145 - 0.51

top left - total coverage %, top right - focus coverage %



Predictions probabilities Resnet - VGG:

activity class

pillar - 0.875 - 0.468

lava - 0.945 - 0.973

top left - total coverage %, top right - focus coverage %



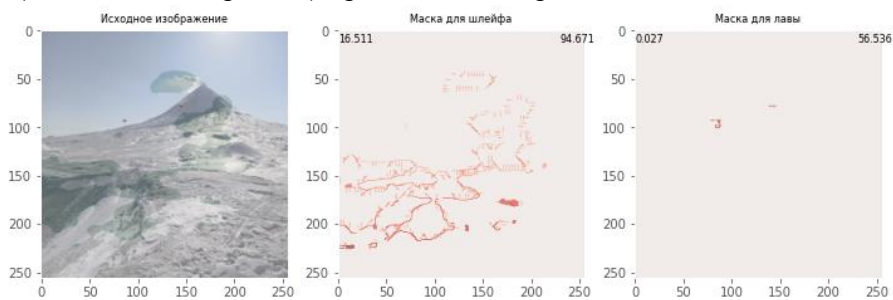
Predictions probabilities Resnet - VGG:

activity class

pillar - 0.27 - 0.03

lava - 0.606 - 0.969

top left - total coverage %, top right - focus coverage %



Predictions probabilities Resnet - VGG:

activity class

pillar - 0.696 - 0.927

lava - 0.994 - 0.998

top left - total coverage %, top right - focus coverage %



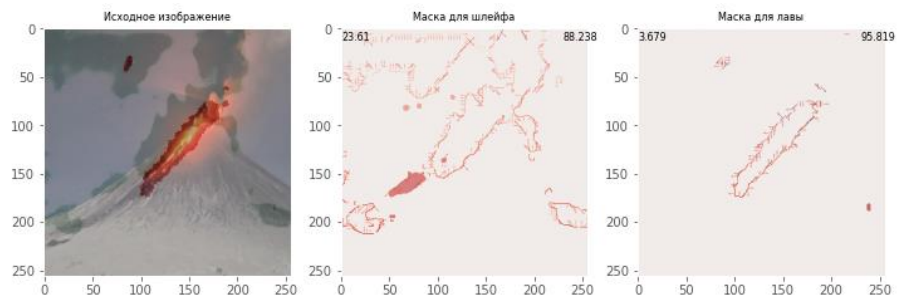
Predictions probabilities Resnet - VGG:

activity class

pillar - 0.093 - 0.207

lava - 0.998 - 0.995

top left - total coverage %, top right - focus coverage %



Predictions probabilities Resnet - VGG:

activity class

pillar - 0.734 - 0.514

lava - 0.994 - 0.995

top left - total coverage %, top right - focus coverage %

