



РОСАТОМ

Команда УИИ

Кейс

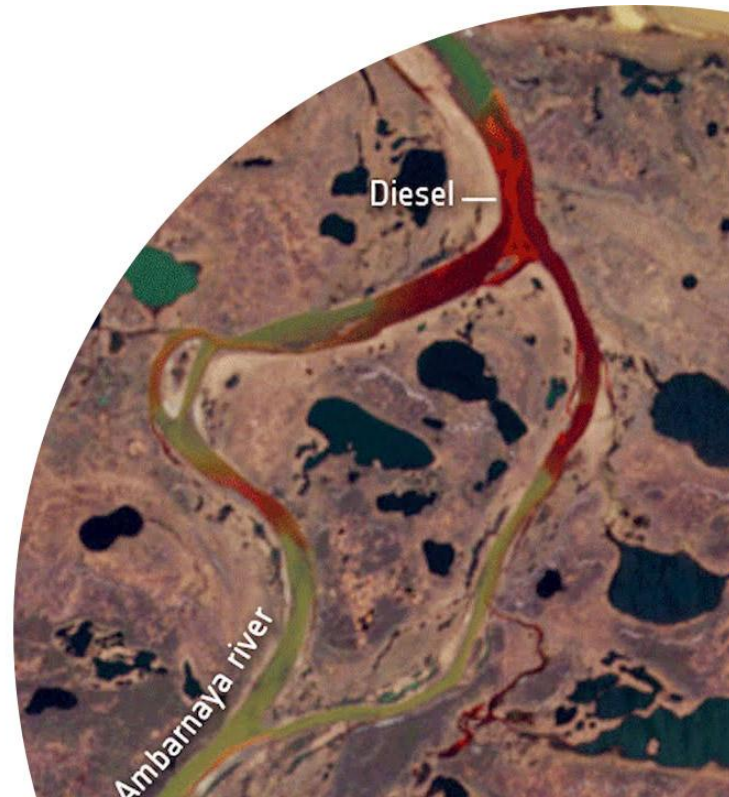
Разработка системы для обнаружения, оценки и мониторинга нефтеразливов (экологического ущерба) с использованием технологий дистанционного зондирования Земли.

Проблема

Региональным органам исполнительной власти из-за нехватки специалистов сложно оперативно обнаруживать аварии, оценивать нанесенный природе ущерб, а также собирать материалы по каждой из них для привлечения виновных лиц к ответственности.

Решение

Создан веб-сервис для анализа спутниковых снимков с целью выявления загрязнений окружающей среды. На основе программного кода производится очистка входных данных и последующее соответствие снимков с координатами, видом загрязнения, датой снимка. Особенность данного сервиса является простое и интуитивное управление не требующее знаний программирования.





POCATOM

СТЕК ТЕХНОЛОГИЙ



библиотека eo-learn



РОСАТОМ

Работа с данными

- Чистка данных —————→
- Создание датафреймов в необходимом формате для работы модели
- Совмещение данных со снимков с таблицей загрязнений в одном датафрейме.
- Аугментация данных (поворот изображения)

Чистка данных

Произведена очистка таблицы от нулевых столбцов и записей до 2017 года.

12096 rows × 40 columns



4719 rows × 23 columns

Путем анализа всех значений были выделены такие столбцы:

Категория загрязнения | широта | долгота | дата | категория земель

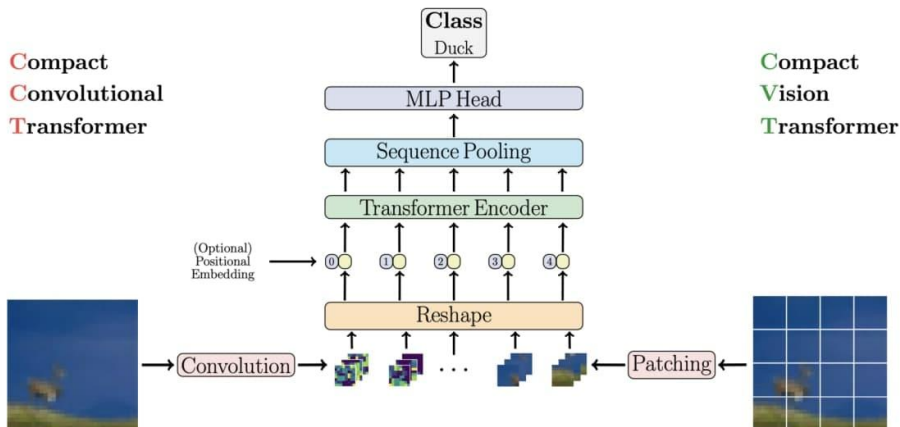
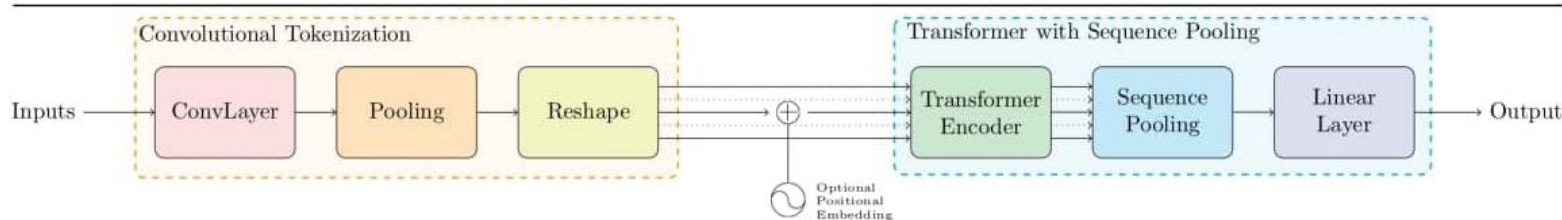


POCATOM

РЕШЕНИЕ

Модель визуального трансформера

Compact Convolutional Transformer (CCT)



Трансформер делит картинку на патчи - меньшие части, анализирует их и дает им токены значения содержимого, и токены значения позиции, а дальше работает как в нлп - анализирует, как коррелируют значения содержимого каждого кусочка картинки с тем, на каком месте эта часть картинки находится.

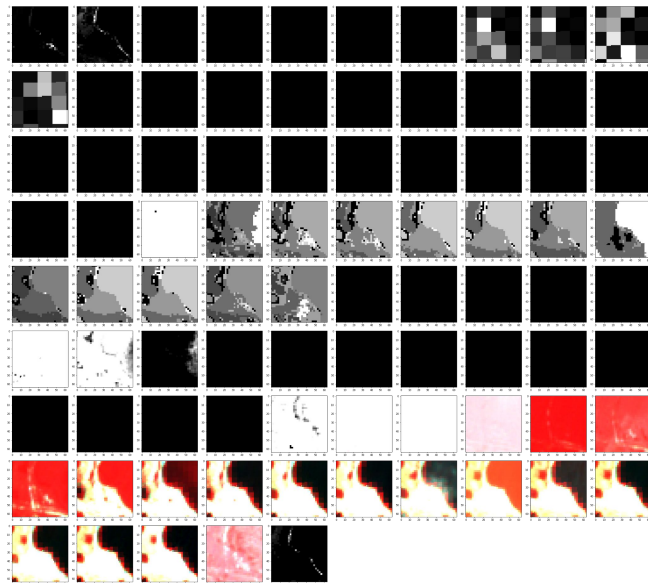


РОСАТОМ

РЕШЕНИЕ

Снимки

Тестирования базы снимков, изучение прямыми методами. Для работы со снимками использовалась **библиотека eo-learn**.





РОСАТОМ

РЕШЕНИЕ

Получение Vbox

Раздел, где из архивов снимков извлекается Vbox с координатами.

**Имея все координаты снимков и координаты загрязнений была создана
таблица соответствий файлов и загрязнений.**

Номер	Директория	Название	Дата и время	Год
0	2021-04	РН-ЮНГ-2013-988	2020-01-21 06:42:12+00:00	2020
1	2021-04	РН-ЮНГ-2013-988	2020-01-28 06:32:13+00:00	2020
2	2021-04	РН-ЮНГ-2013-988	2020-02-22 06:32:15+00:00	2020



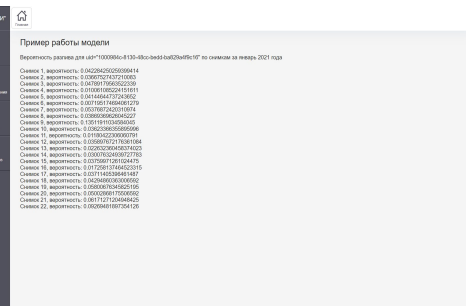
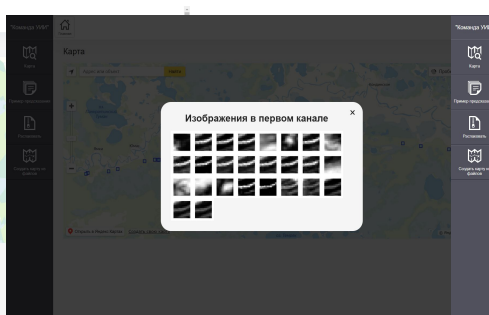
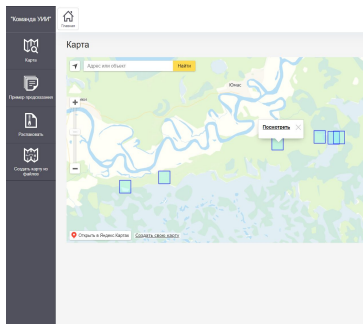
РОСАТОМ

ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС



Новаторские идеи и фишки

1. Распаковка датасета
2. Обработка загруженного изображения нейросетью с выдачей и генерацией интерактивной карты
3. Просмотр снимков на карте
4. Вероятность разлива
5. Экспорт в PDF



[Видеодемонстрация](#)



РОСАТОМ

СТОИМОСТЬ И СРОКИ

Внедрение

Для запуска пилотной версии с базовым функционалом

– **3 месяца**

Для полномасштабной надежной работы системы

– **6 месяцев**

Стоимость внедрения

960 000 руб.

Команда готова к внедрению решения



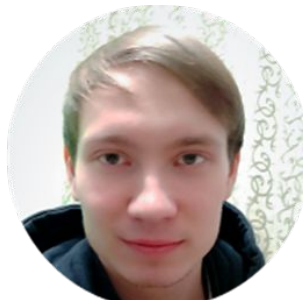
РОСАТОМ

КОМАНДА УИИ



**Качалкин
Артём**

Data Scientist



**Соснин
Дмитрий**

Data Scientist



**Курочкин
Алесе́й**

Data Scientist



**Хуторной
Борис**

Data Scientist



**Домненко
Алексей**

Data Scientist
Full-stack



РОСАТОМ

ПОЛЕЗНЫЕ ССЫЛКИ

Пользовательский интерфейс

<http://144.76.152.13:9999/system/>

логин: admin

пароль: 12345678



ЦП, обработка базы

<https://colab.research.google.com/drive/18uFLm5dA6lz3H54EKRnoRvehDq9-gKyf#scrollTo=intCeDazPfxX>

Модель обучения

<https://colab.research.google.com/drive/1X3ZpL2pr2xMYZgp3JLEDt0NTJz896G10?usp=sharing>

GitHub

<https://github.com/borisstrong/rosatom>



РОСАТОМ

ПРИЛОЖЕНИЕ

Очистка данных

С помощью библиотеки для анализа данных “pandas” в ноутбук была загружена таблица загрязнений.

Произвелась очистка от нулевых столбцов и строк.

Путем анализа всех значений были выделены такие столбцы:

Категория загрязнения | широта | долгота | дата | категория земель

Широта и долгота были приведены к единому десятичному формату.

Так же удалены все записи ранее 2017 года, так как по ним нет спутниковых снимков. И чуть позже записи по которым не было снимков.

	cat_pollution	lat	lon	date	cat_ground	square
0	нефть	73.506833	62.592583	2017-02-15 00:00:00	земли лесного фонда	0.1703
1	нефть	73.508194	62.543500	2017-02-15 00:00:00	земли лесного фонда	0.2595
2	нефть	73.347778	62.616667	2019-03-04 00:00:00	земли лесного фонда	0.1559
3	нефть	73.547472	62.557611	2021-02-18 00:00:00	земли лесного фонда	0.0100
4	нефть	73.983397	62.129353	2021-02-18 00:00:00	земли лесного фонда	0.0050
...
4714	нефть	77.792639	61.173361	2020-03-11 00:00:00	земли лесного фонда	0.0200
4715	нефть	77.738583	61.227111	2020-03-11 00:00:00	земли лесного фонда	0.4500
4716	подтоварная вода	77.742867	61.166608	2021-02-18 00:00:00	земли лесного фонда, земли промышленности	0.0983
4717	нефтепродукты	77.539306	61.316889	2020-02-18 00:00:00	земли лесного фонда	0.0200
4718	нефть	75.120758	61.429492	2020-03-05 00:00:00	земли лесного фонда/земли иных категорий	0.7663

4719 rows x 6 columns



РОСАТОМ

ПРИЛОЖЕНИЕ

Получение координат и работа с данными

Следующий этап было получение координат по снимкам и временные метки по каждому снимку. Приводилось все в табличный вид.

Для работы со снимками использовалась **библиотека eo-learn**.

Собраны все BBox в виде координат.

	dir	name	x1	y1	x2	y2
0	2021-04	РН-СНГ-2013-5116	76.989089	61.190537	77.001351	61.196104
1	2021-04	РН-СНГ-2013-4345	76.686472	61.227745	76.698695	61.233340
2	2021-04	РН-СНГ-2013-4361	76.728468	61.229940	76.740699	61.235530
3	2021-04	РН-СНГ-2013-5163	76.923933	61.152721	76.936168	61.158294
4	2021-04	РН-СНГ-2013-5380	76.890414	61.206888	76.902665	61.212464
...
11678	2021-07	РН-СНГ-2018-626	76.485370	61.446162	76.497643	61.451774
11679	2021-07	РН-СНГ-2017-240	76.764541	61.376538	76.776836	61.382125
11680	2021-07	РН-СНГ-2018-553	76.667407	61.182925	76.679610	61.188521
11681	2021-07	РН-СНГ-2018-434	76.471186	61.143595	76.483339	61.149209
11682	2021-07	РН-СНГ-2018-237	76.561286	61.219327	76.573483	61.224933

11683 rows x 6 columns



РОСАТОМ

ПРИЛОЖЕНИЕ

Создание таблицы соответствий

Затем, имея все координаты снимков и координаты загрязнений была создана таблица соответствий файлов и загрязнений. В расчет бралось то понимание, что координата разлива будет между координатами у снимка.

Так же понимая дату разлива было легко узнать год, и как следствие архив, где его нужно искать. Сами координаты нам не нужны, а нужна была дата, ну и для возможных улучшений связь с таблицей загрязнений

Эта таблица выглядит так:

	year	dir	name	date	id_pollution
0	2017	01	ЛУК-2016-6	2017-02-15 00:00:00	0
1	2017	01	ЛУК-2016-7	2017-02-15 00:00:00	1
2	2019	01	ЛУК-КГ-2018-1	2019-03-04 00:00:00	2
3	2021	01	ЛУК-КГ-2020-1	2021-02-18 00:00:00	3
4	2021	01	ЛУК-КГ-2020-2	2021-02-18 00:00:00	4
...
76649	2021	01	ПП-2013-129	2021-02-18 00:00:00	11807
76650	2021	01	ПП-2013-130	2021-02-18 00:00:00	11807
76651	2014	01	ПП-2013-207	2014-02-15 00:00:00	11809
76652	2020	16	РУФ-2020-1	2020-02-18 00:00:00	11810
76653	2020	16	ТС-2019	2020-03-05 00:00:00	11811

76654 rows x 5 columns

	dir	name	id	date
0	2020-11	РН-ЮНГ-2013-988	0	2020-01-21 06:42:12+00:00
1	2020-11	РН-ЮНГ-2013-988	1	2020-01-28 06:32:13+00:00
2	2020-11	РН-ЮНГ-2013-988	2	2020-02-22 06:32:15+00:00
3	2020-11	РН-ЮНГ-2013-988	3	2020-03-18 06:32:18+00:00
4	2020-11	РН-ЮНГ-2013-988	4	2020-04-10 06:42:15+00:00
...
179420	2017-07	РН-СНГ-2018-601	4	2017-08-31 06:30:12+00:00
179421	2017-07	РН-СНГ-2018-601	5	2017-10-07 06:17:15+00:00
179422	2017-07	РН-СНГ-2018-601	6	2017-10-12 06:18:04+00:00
179423	2017-07	РН-СНГ-2018-601	7	2017-10-25 06:29:33+00:00
179424	2017-07	РН-СНГ-2018-601	8	2017-11-14 06:31:20+00:00

179425 rows x 4 columns

Кроме этого нам нужны временные метки каждого участка, что мы оставили, они будут использоваться в модели при обучения, чтобы брать нужный батч, где есть разлив и сравнивать с другими, которые были до него.



POCATOM

ПРИЛОЖЕНИЕ

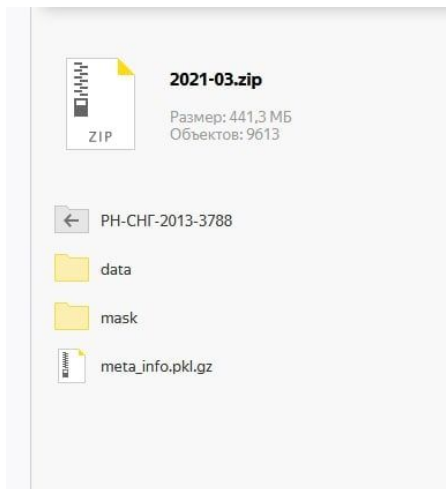
Создание облегченной базы для дальнейшего обучения

Имея таблицу соответствия файлов и загрязнений была создана облегченная база, где есть все необходимые снимки и данные для обучения нейронной сети

Ещё хочется отметить момент, что достаточно немало файлов не имели координат для снимков или их временные метки. Такие данные откидывались.

Довольно большая часть файлов имела нечитаемое расширение gz0000644, и их не удалось использовать.

Название ↑	Владелец	Последнее изменение	Размер файла
2017.zip	я	13:48 я	1,69 ГБ
2018.zip	я	13:47 я	4,41 ГБ
2019.zip	я	13:48 я	5,03 ГБ
2020.zip	я	13:52 я	2,41 ГБ
2021.zip	я	11:24 я	2,87 ГБ





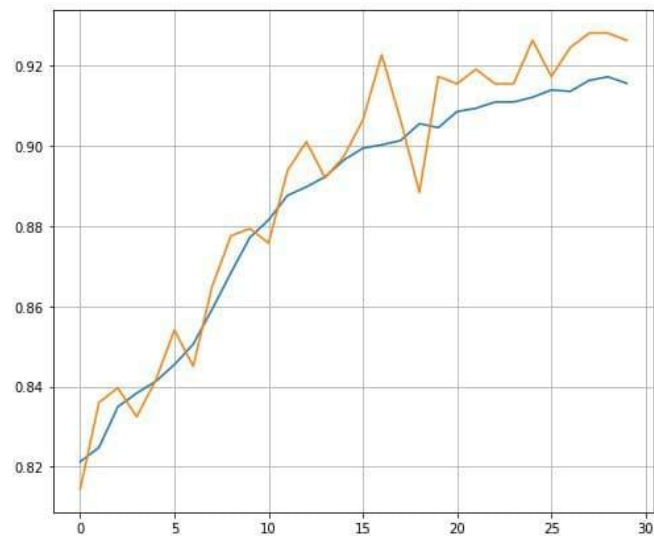
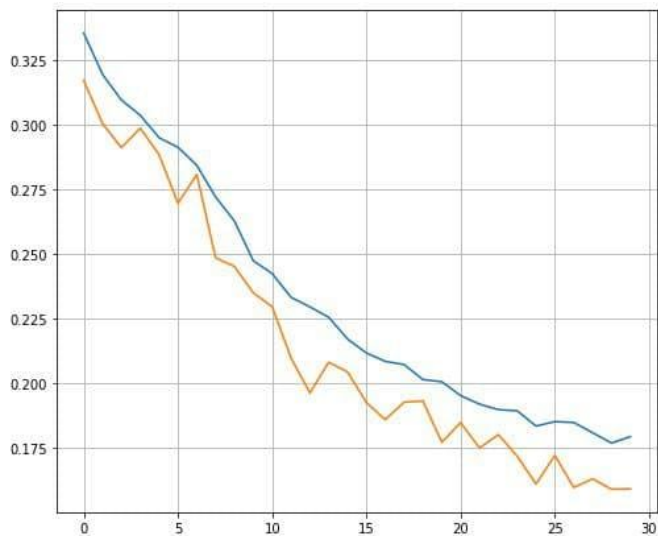
РОСАТОМ

Точность

Ошибки и точности нейросети во время первых 30 эпох обучения

синий график - функция во время обучения

оранжевый - во время валидации





РОСАТОМ

Повышение точности

Повышение точности сети возможно за счет аугментации

Аугментация - искусственное расширение датасета за счет дополнения имеющихся изображений их отраженными, транспонированными, повернутыми и зашумленными копиями, для улучшения обучения сети.

Для аугментации использовался пайплайн `tensorflow.data.Dataset`, настроенный с помощью написанных под текущую задачу функций. В пайплайн подавались списком директории, где находятся снимки местности, затем внутри пайплайна с помощью библиотеки `eo-learn` загружаются снимки, и с помощью методов `tensorflow` проводится аугментация. Кроме того, в последствии в модель был добавлен ряд дополнительных слоёв предварительной обработки.

В методы аугментации входят: перемешивание экземпляров фото по каждому квадрату, зашумление картинки, случайные скачки контрастности, вращение, случайная обрезка, отражение, транспонирование и смещение с отзеркаливанием.



POCATOM

Полная схема модели

