



Использование нейронных сетей и методов машинного обучения в исследованиях и разработках Дальневосточного центра «НИЦ «Планета»

Докладчик: к.т.н., Кучма Михаил Олегович

XXI Международная научная Школа-конференция молодых ученых по
фундаментальным проблемам дистанционного зондирования Земли из космоса

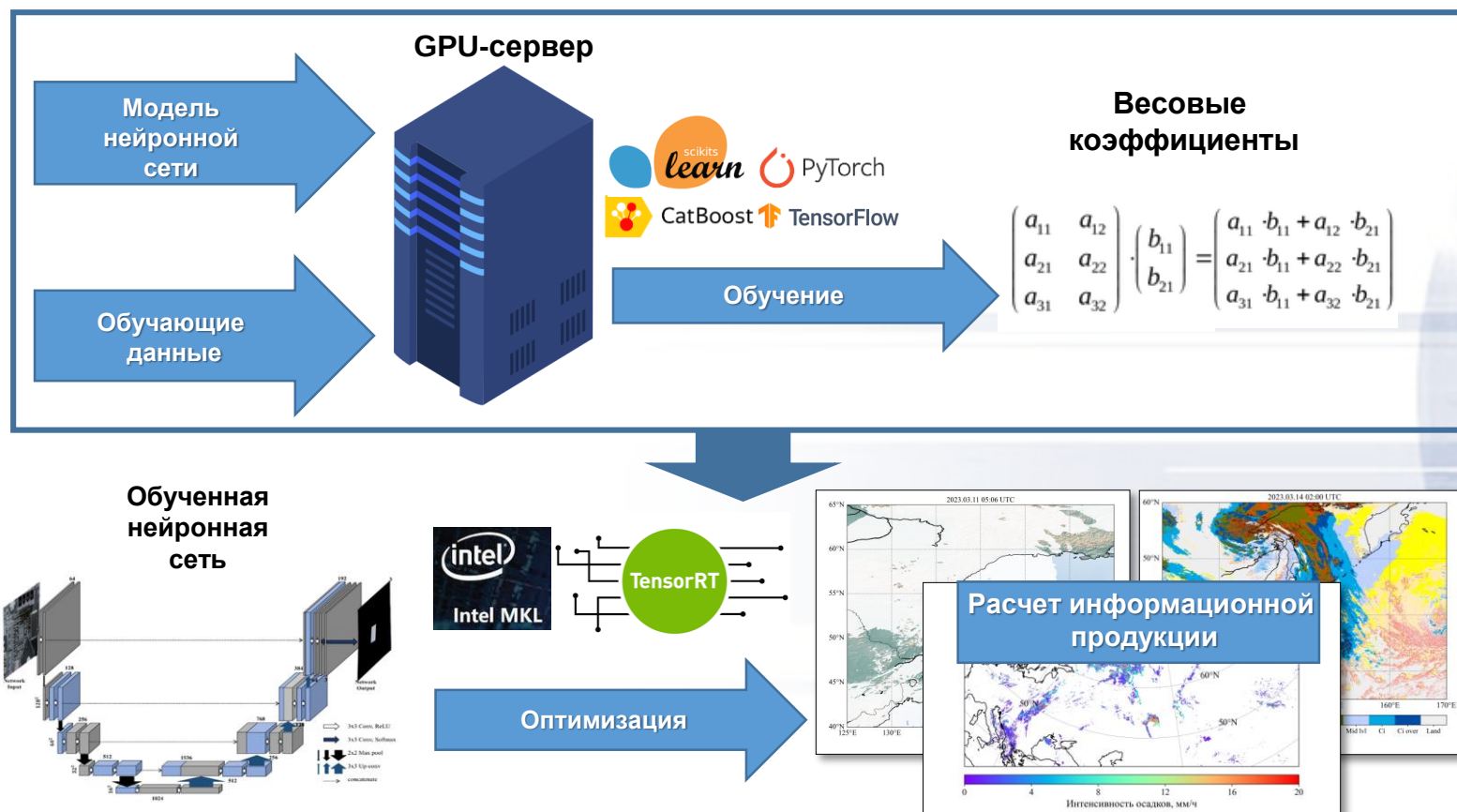
Актуальность нейронных сетей

- ✓ Сопоставимая или превосходящая точность при сопоставлении с физическими и эмпирическими алгоритмами
- ✓ Выявление неявных связей и их использование для создания информационного продукта
- ✓ Быстрое время создания алгоритма и запуск его в оперативную работу



Использование методов искусственного интеллекта

В Дальневосточном центре НИЦ «Планета» для получения спутниковой информационной продукции широко применяются методы искусственного интеллекта (ИИ) на основе алгоритмов машинного обучения, в том числе нейронные сети: полносвязные прямого распространения, сверточные и рекуррентные, трансформеры, генеративно-сопоставительные и другие.



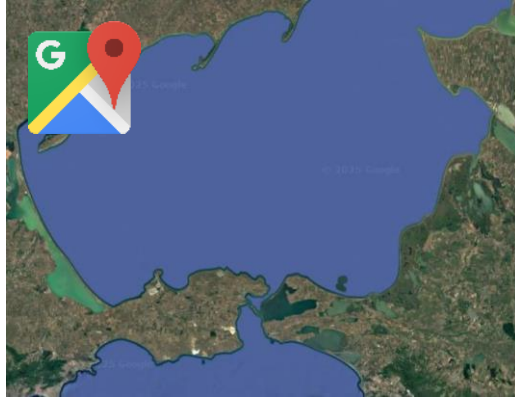
Для получения алгоритмов на основе ИИ необходима обучающая выборка сверх большого объема, включающая все возможные условия наблюдений. Обучение ИИ-моделей – многократно повторяющийся итеративный процесс, требующий больших вычислительных мощностей. Для обучения моделей применяются специализированные GPU-серверы, которые позволяют на порядок повысить скорость обработки в сравнении с обработкой на центральном процессоре и в несколько раз – в сравнении с обработкой на рабочей станции с одним графическим ускорителем.

Создание обучающей выборки

В каком виде использовать спутниковые данные?



Снимки Google Maps

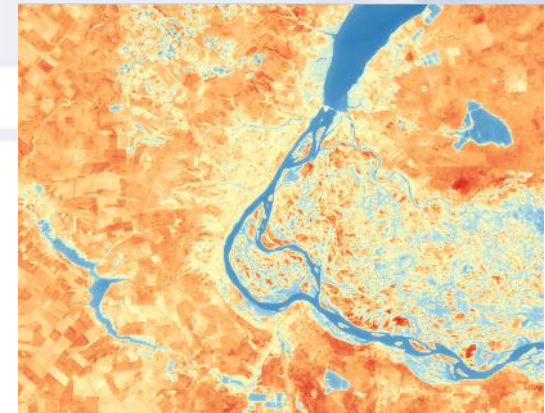


Снимки в формате JPEG, PNG



Физические величины:

- ✓ Radiance – СПЭЯ (спектральная плотность энергетической яркости), $\text{Вт/м}^2 \cdot \text{ср}$
- ✓ Reflectance – КСЯ (коэффициент спектральной яркости), %
- ✓ Brightness temperature – яркостная температура, К



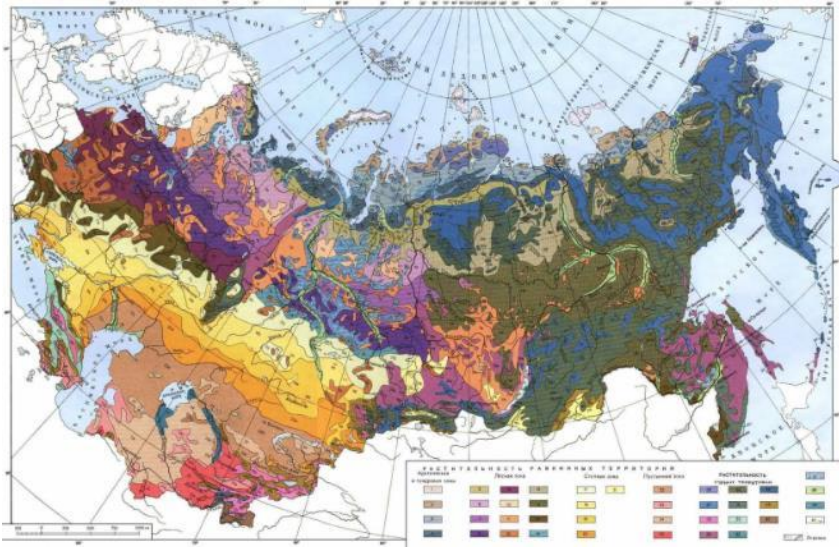
Создание обучающей выборки

Что важно учитывать?

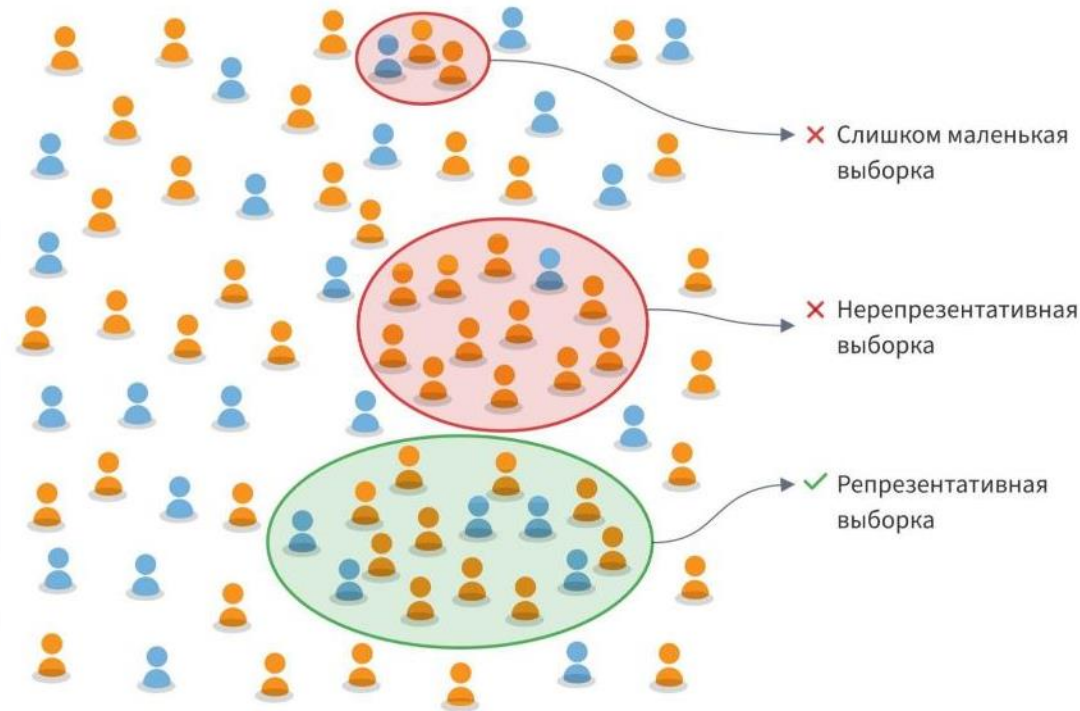
Сезонность



Тип подстилающей поверхности



Количество данных и баланс классов



Создание обучающей выборки

Какие еще данные можно использовать?

ЦМР (цифровая
модель рельефа)



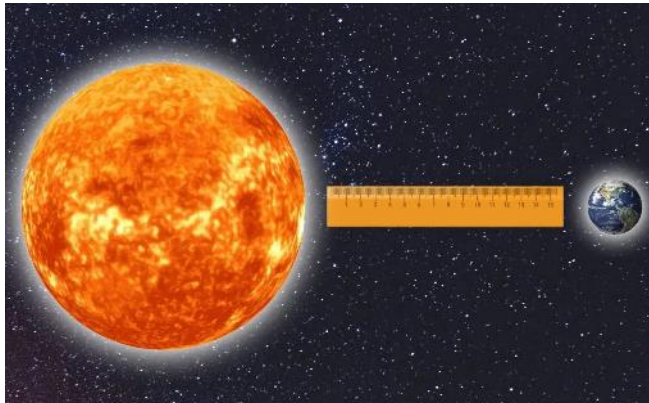
Данные
радиозондирования



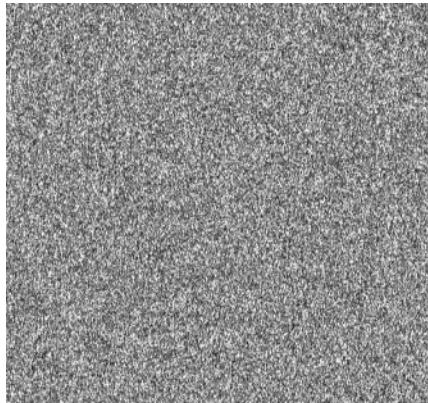
ЧПП (численный
прогноз погоды)



Расстояние до солнца в а.е.
(как учет сезонности)



Белый шум



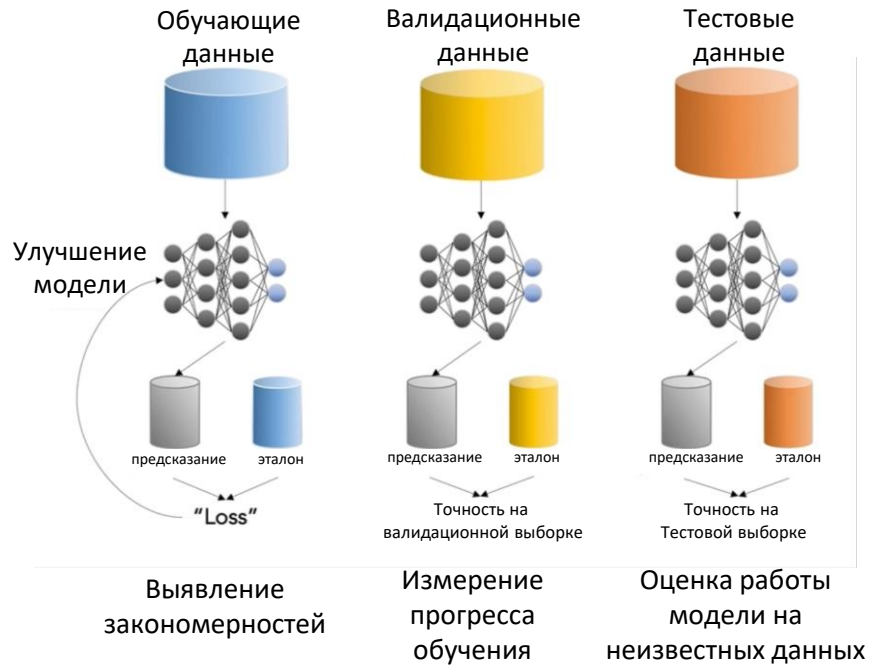
Эталонные измерения



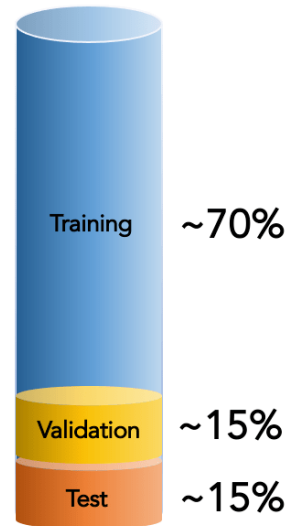
Создание обучающей выборки

Разделение

Для чего это нужно?

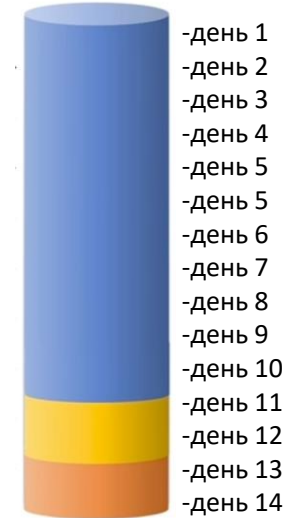


Какие пропорции?



Как делить?

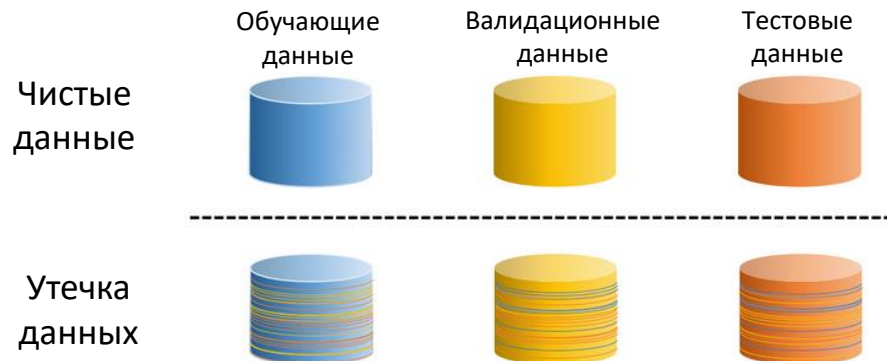
Отдельные сроки



Случайно

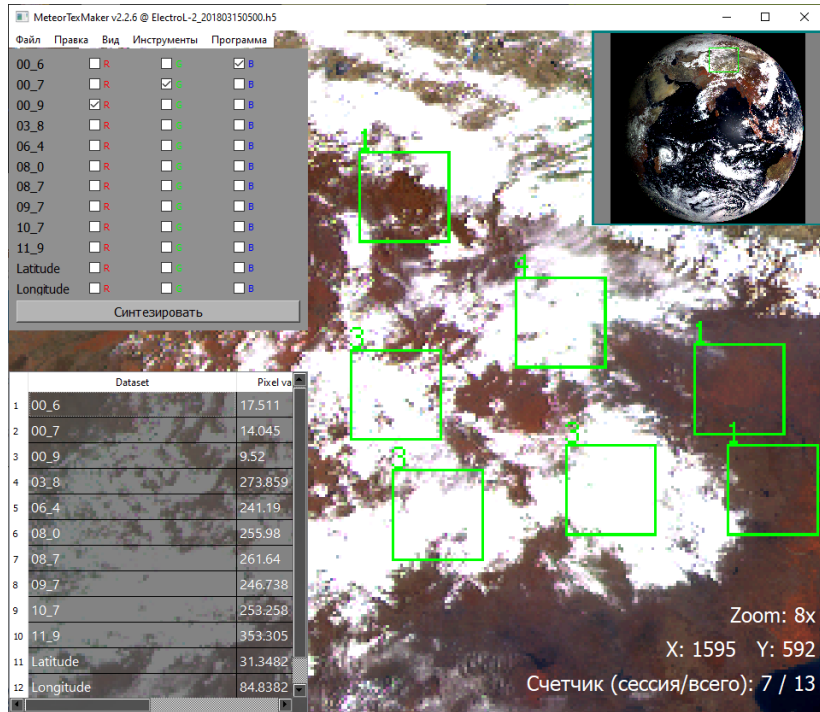


Что будет, если делить данные неправильно?



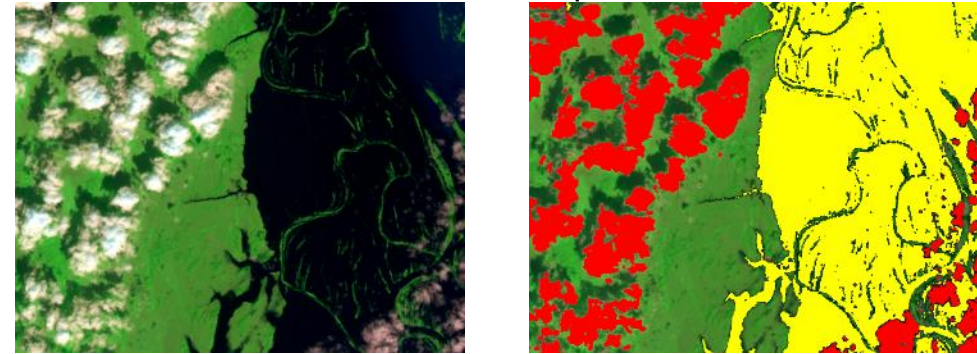
Создание обучающей выборки

Классификация



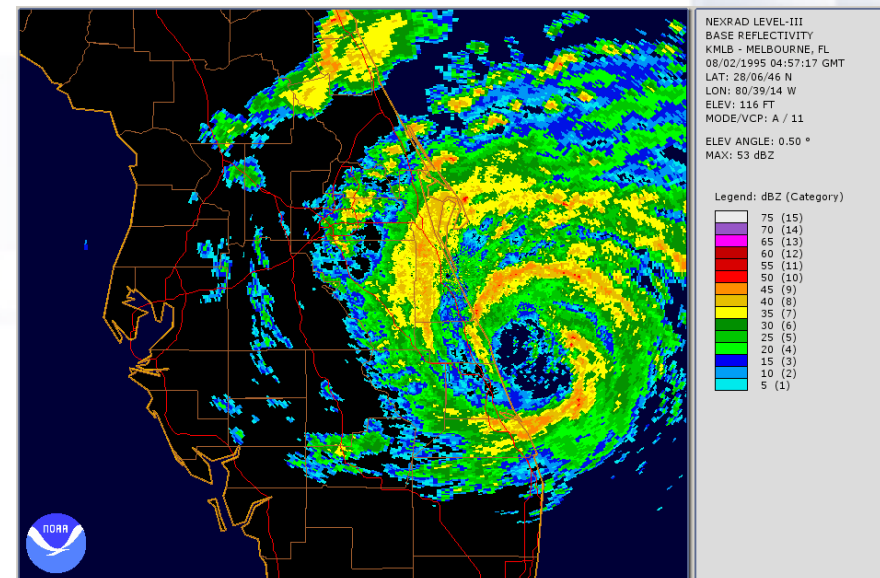
Специализированное ПО для
разметки данных

Сегментация



Ручное сегментирование данных

Прогноз



Нарезка сегментов разновременных данных

Хранение обучающей выборки

Хранение в одном или нескольких файлах формата HDF5

Внутренняя структура файла HDF5 (визуализация в Panoply)

Имя	Тип	Размер
EL4_202509021730_202509022300_140_1550.h5	Файл "H5"	5 143 КБ
EL4_202509021730_202509022300_140_1450.h5	Файл "H5"	5 321 КБ
EL4_202509021730_202509022300_140_1350.h5	Файл "H5"	5 304 КБ
EL4_202509021730_202509022300_140_1250.h5	Файл "H5"	5 035 КБ
EL4_202509021730_202509022300_140_1150.h5	Файл "H5"	4 735 КБ
EL4_202509021730_202509022300_140_1050.h5	Файл "H5"	4 660 КБ
EL4_202509021730_202509022300_140_950.h5	Файл "H5"	4 727 КБ
EL4_202509021730_202509022300_140_850.h5	Файл "H5"	4 859 КБ
EL4_202509021730_202509022300_140_750.h5	Файл "H5"	5 048 КБ
EL4_202509021730_202509022300_140_650.h5	Файл "H5"	5 223 КБ
EL4_202509021730_202509022300_140_550.h5	Файл "H5"	5 030 КБ
EL4_202509020500_202509021030_140_1697.h5	Файл "H5"	4 429 КБ
EL4_202509020500_202509021030_140_1650.h5	Файл "H5"	4 568 КБ
EL4_202509020500_202509021030_140_1550.h5	Файл "H5"	4 904 КБ
EL4_202509020500_202509021030_140_1450.h5	Файл "H5"	5 145 КБ
EL4_202509020500_202509021030_140_1350.h5	Файл "H5"	5 174 КБ
EL4_202509020500_202509021030_140_1250.h5	Файл "H5"	5 080 КБ
EL4_202509020500_202509021030_140_1150.h5	Файл "H5"	4 759 КБ
EL4_202509020500_202509021030_140_1050.h5	Файл "H5"	4 612 КБ
EL4_202509020500_202509021030_140_950.h5	Файл "H5"	4 553 КБ

Name	Long Name	Type
EL4_202509021730_202...	EL4_20250902...	Local File
bt_11.2um	bt 11.2um	Geo2D
cloud_mask	cloud mask	Geo2D
latitude	latitude	Geo2D
longitude	longitude	Geo2D
rain_rate	rain rate	Geo2D

File "EL4_202509021730_202509022300_140_550.h5"

File type: Hierarchical Data Format, version 5

```
netcdf file:///d:/niokr/Mikhail/Out/DsForNowcastTrain
variables:
    float bt_11.2um(12, 512, 512);
        :min = 0.21f; // float
        :max = 298.02f; // float
        :mean = 264.23947f; // float
        :std = 43.8561f; // float
        :_ChunkSizes = 2U, 64U, 64U; // uint

    float cloud_mask(12, 512, 512);
        :min = 0.0f; // float
        :max = 1.0f; // float
        :mean = 0.4604826f; // float
        :std = 0.4984359f; // float
        :_ChunkSizes = 2U, 64U, 64U; // uint

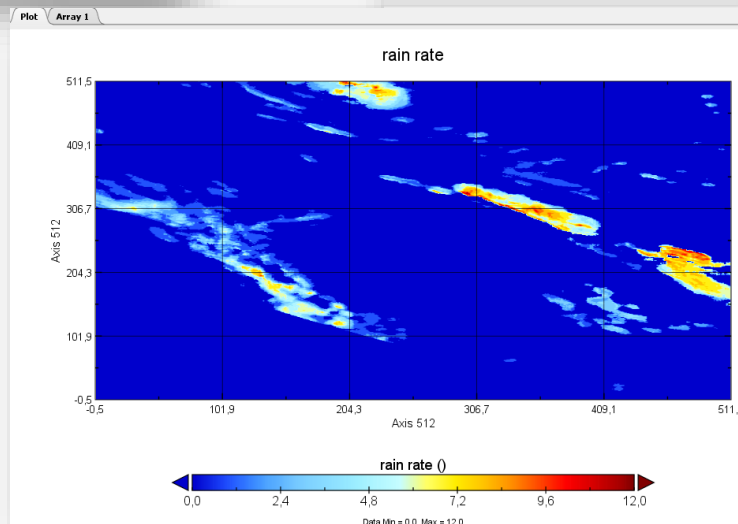
    float latitude(512, 512);
        :min = 28.783941f; // float
        :max = 65.71573f; // float
        :mean = 42.026592f; // float
        :std = 8.580802f; // float
        :_ChunkSizes = 64U, 64U; // uint

    float longitude(512, 512);
        :min = 93.65393f; // float
        :max = 151.72133f; // float
        :mean = 133.0541f; // float
        :std = 11.167198f; // float
        :_ChunkSizes = 64U, 64U; // uint

    float rain_rate(12, 512, 512);
        :min = 0.0f; // float
        :max = 22.0f; // float
        :mean = 0.33979163f; // float
        :std = 1.2634016f; // float
        :_ChunkSizes = 2U, 64U, 64U; // uint

// global attributes:
:start_i = 140; // int
:start_j = 550; // int
:step = 512; // int
```

Визуализация одного примера



Валидация результатов

- Сравнение необходимо проводить с:
 - *in-situ* данные
 - аналогичные продукты (зарубежные или отечественные)
 - другие доступные эталонные данные
- Необходимо использовать корректные метрики для оценки точности



Виды используемых нейронных сетей

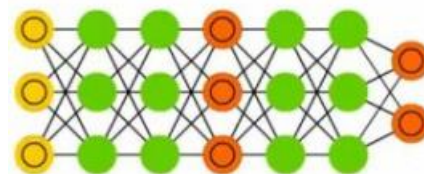
**Многослойная нейронная сеть
прямого распространения**
Deep Feed Forward (DFF)



Рекуррентная нейронная сеть
Recurrent Neural Network (RNN)



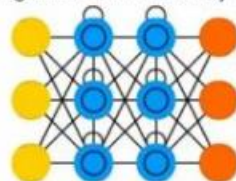
Генеративная состязательная сеть
Generative Adversarial Network (GAN)



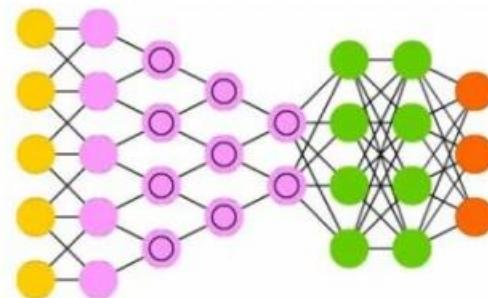
Автоэнкодер
Auto Encoder (AE)



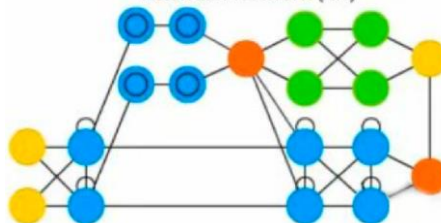
Нейронная сеть долгой краткосрочной памяти
Long Short-Term Memory (LSTM)



Глубокая сверточная нейронная сеть
Deep Convolutional Network (DAE)

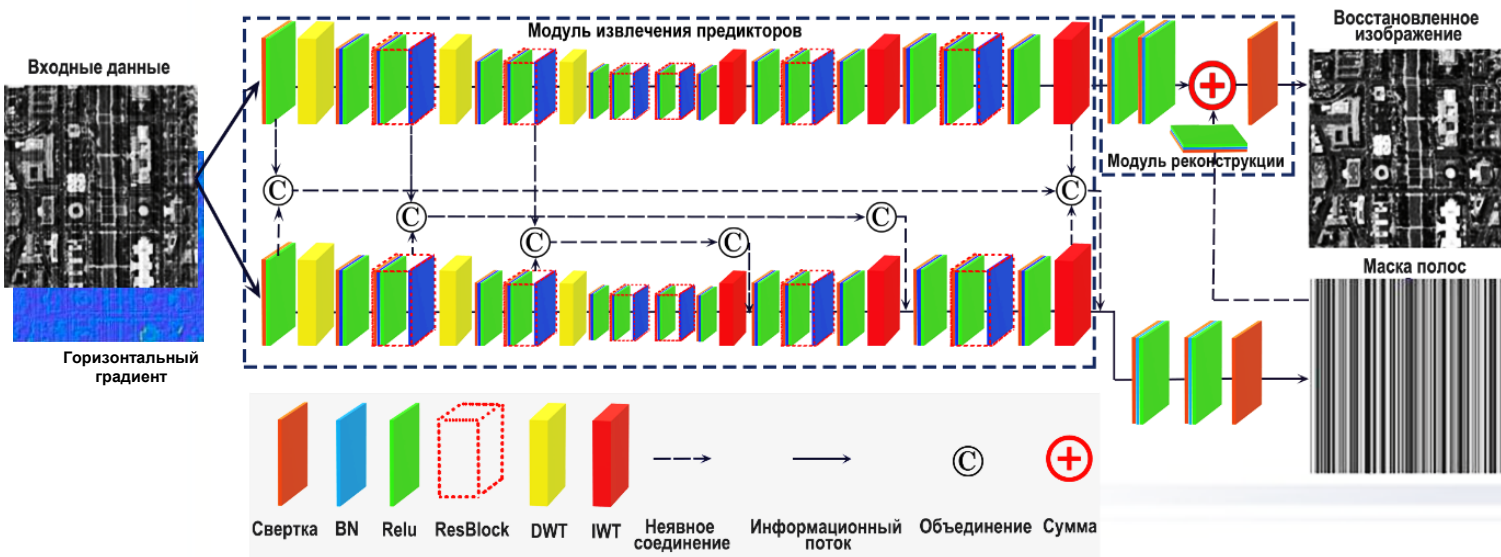


Нейронная сеть внимания
Attention Network (AN)

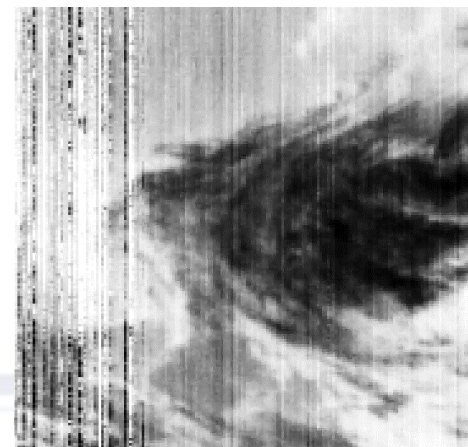


- Входы нейронной сети
- Нейроны с обратными связями
- △ Зашумленные входы нейронной сети
- Скрытые нейроны
- Вероятностные скрытые нейроны
- △ Импульсные нейроны
- Капсульные нейроны
- Выходы нейронной сети
- Нейроны сравнения входов и выходов
- Нейроны с рекуррентными или неполными связями
- Нейроны памяти
- △ Нейроны памяти с фильтром
- Ядро
- Свертка

Минимизация помех в ИК каналах МСУ-ГС КА Арктика-М № 1 с помощью сверточной нейронной сети

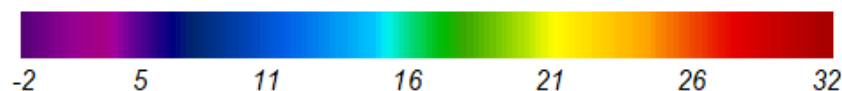
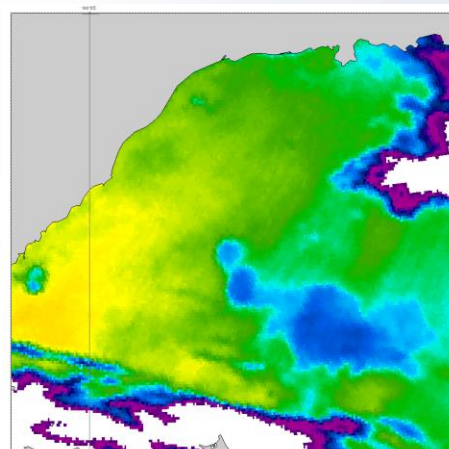
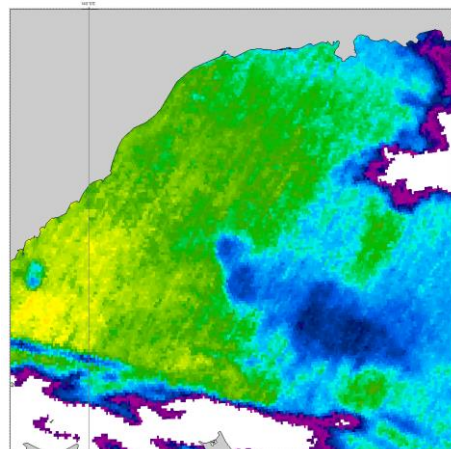


Исходные данные

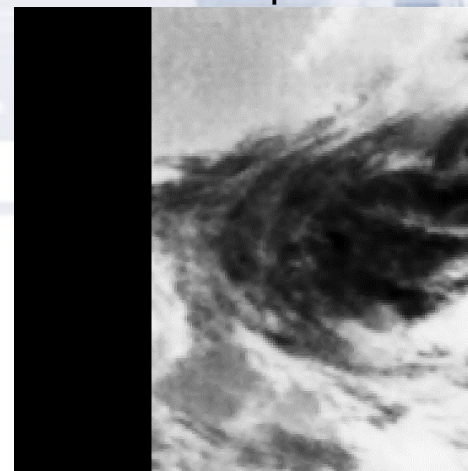


До применения ИИ

После применение ИИ



После нейросети



Карта
температуры
поверхности
океана, °C

Маска облачности по данным Арктика-М

по данным прибора МСУ-ГС КА серии Арктика-М

Выборка
текстурных данных



Обучающая выборка:

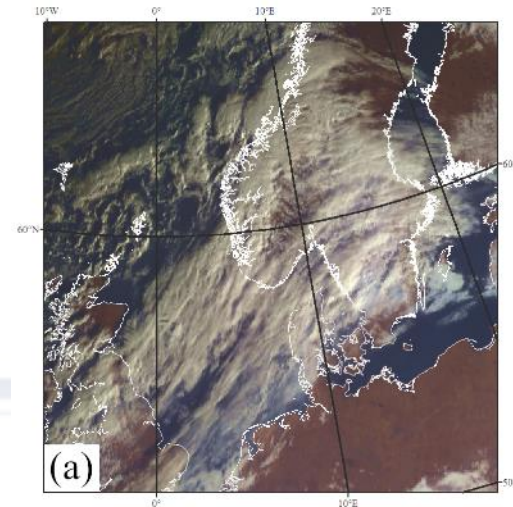
- Сегменты размером 256x256 пикселей
- Маски классифицировались вручную опытными специалистами
- Выборка охватывает период с июня 2022 по март 2023
- Общий объем выборки: 17 тыс. сегментов

**Сходство масок по данным
МСУ-ГС и SEVIRI $\approx 93\%$**

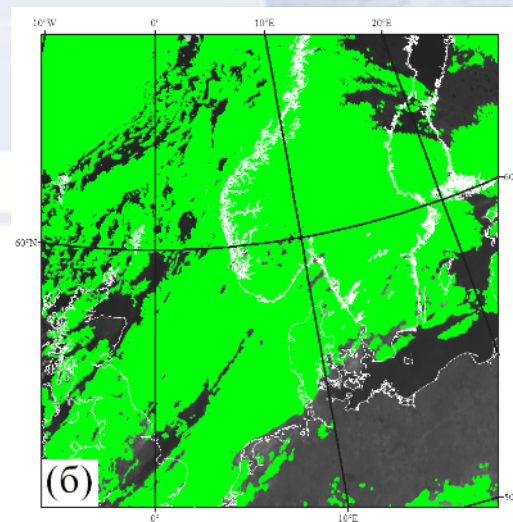
**Точность масок по данным
МСУ-ГС $\approx 92\%$**

**Точность масок по
данным SEVIRI $\approx 94\%$**

Мультиспектральное изображение

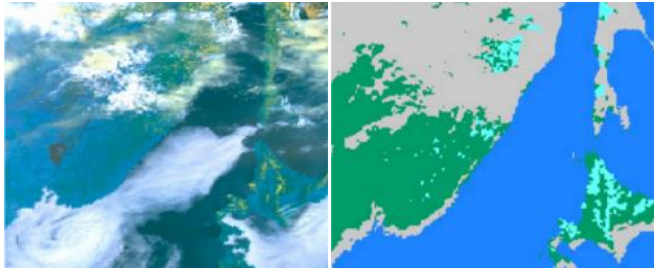


Маска облачности



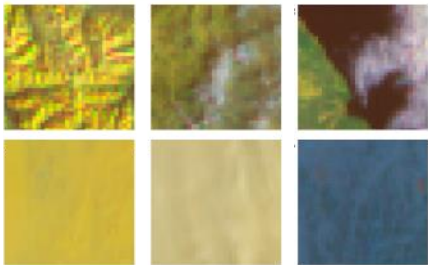
Использование сверточных нейронных сетей CNN

Карта распределения снежного покрова по данным прибора MSU-MP KA Метеор-М №2



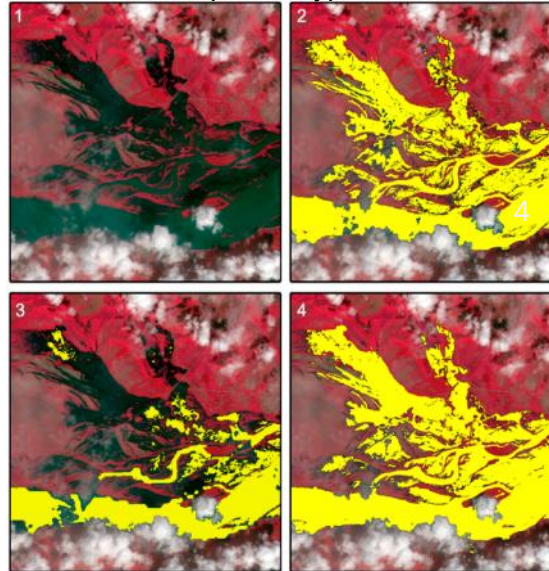
Используется полносверточная нейросетевая архитектура для классификации типов подстилающей поверхности. Входной информацией являются текстурные изображения размером 32x32 пикселя

Примеры цветосинтезированных текстурных изображений снега и облачности



Детектирование разлива реки Амур по данным приборов MSI и OLI KA Sentinel-2 и Landsat-8

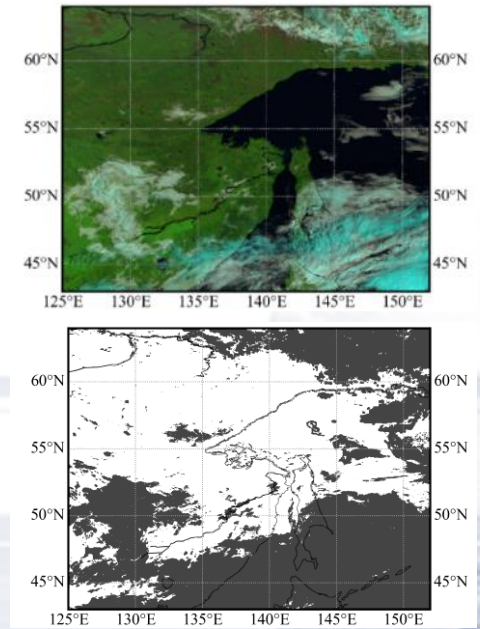
Модель классификации: сверточная сеть на базе архитектуры Unet



Оценка точности на 6000 тестовых участках.
При сравнении с эталонными картами:
Precision – 94.51%, Recall – 89.95%
F1-measure – 92.17%

При сравнении с Sentinel Scene Classification:
Precision – 85.45%, Recall – 79.84%,
F1-measure – 82.55%

Обнаружение облачности по данным КА Himawari-8/9:

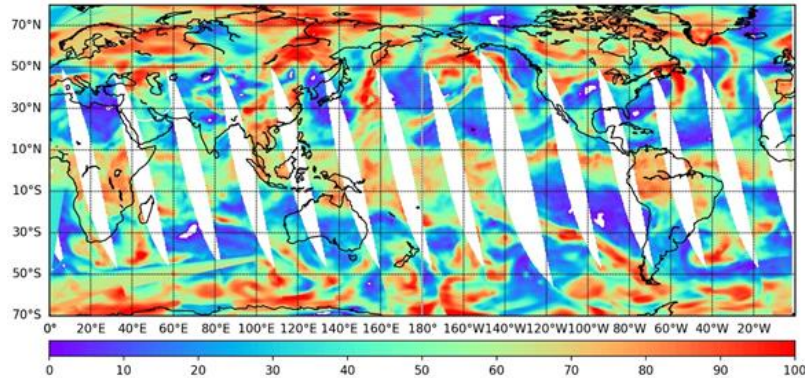


Ансамбль из 9 сверточных нейросетевых классификаторов на основе спектральной и текстурной информации

Значение F1-меры в сравнении с маской облачности Himawari по данным NOAA: 87,5% и 83,98% для светлого и темного времени суток

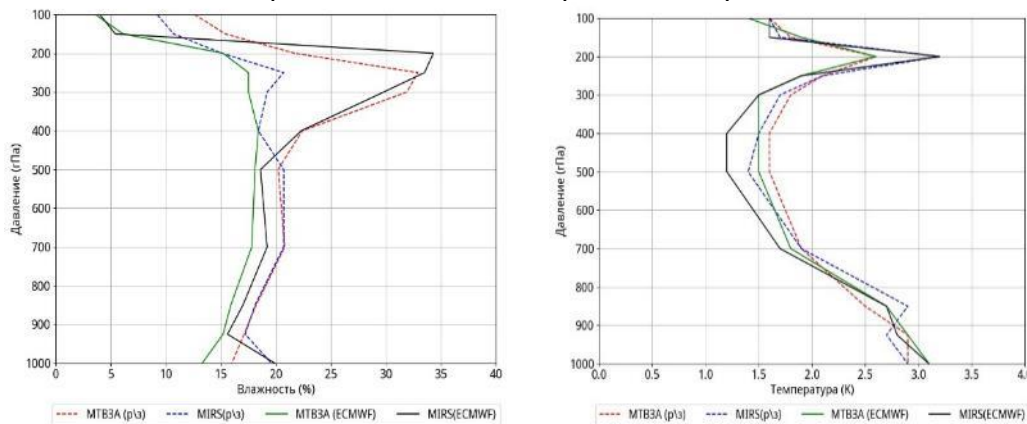
Использование сетей MLP

Карта суточного композита профиля влажности на уровне 700 гПа по состоянию на 2021.04.26 по данным МТВЗА-ГЯ КА Метеор-М №2-2



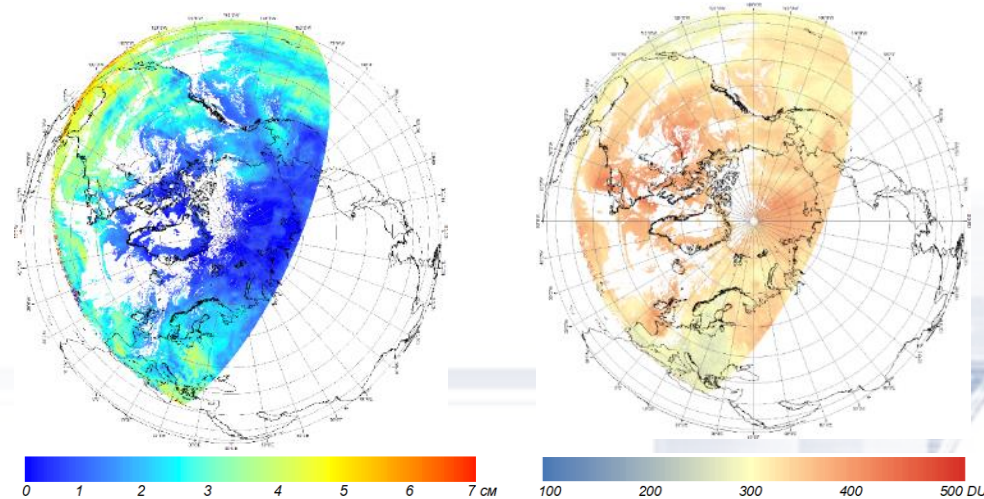
Обучающие данные: ECMWF-ERA5 (около 30 млн. примеров);
Входные данные: измерения в каналах прибора МТВЗА-ГЯ, широта, высота над уровнем моря;
Модель: многослойный персептрон с 3 скрытыми слоями;
Выходная информация: температура и влажность на 37 изоб. уровнях.

Точность в сравнении с данными радиозондирования и ECMWF



Филей А. А., Андреев А. И., Успенский А. Б. Использование искусственных нейронных сетей для восстановления температурно-влажностного состояния атмосферы по данным спутникового микроволнового радиометра МТВЗА-ГЯ КА Метеор-М № 2-2 // Исследование Земли из космоса. – 2021. – № 6. – С. 83-95.

Карты общего содержания озона (слева) и водяного пара (справа) по состоянию на 2022.06.26 17:00 UTC по данным КА Арктика-М №1



Обучающие данные: приборы OMPS KA Suomi NPP (225 тыс. точек);
Входные данные: измерения в ИК каналах прибора МСУ-ГС, профиль температуры, приземное давление, зенитные углы наблюдения;
Модель: многослойный персептрон с 4 скрытыми слоями.

Ошибка RMSE определения озона: 0.41 см (эталон AERONET)
Ошибка RMSE определения водяного пара: 19 DU (эталон WOUDC)

Блощинский, В. Д. Определение содержания водяного пара в столбе атмосферы по данным КА "Электро-Л" №3 с использованием нейронных сетей / В. Д. Блощинский, А. А. Филей, Е. И. Холодов // Оптика атмосферы и океана. – 2021. – Т. 34. – № 10(393). – С. 808-811.

В. Д. Блощинский, М. О. Кучма, А. В. Кухарский. Определение общего содержания озона в столбе атмосферы по данным КА Электро-Л № 3 с использованием нейронных сетей // Исследование Земли из космоса. – 2022. – № 4. – С. 79-85.

Использование визуальных трансформеров ViT

Обучающие данные: измерения в рамках проекта GPM IMERG (около 500 тыс. текстурных изображений);

Входные данные: измерения КА Himawari-8/9 в ИК каналах, микро- и макрофизические параметры облачности, представленные в виде изображений 5x5 пикселей;

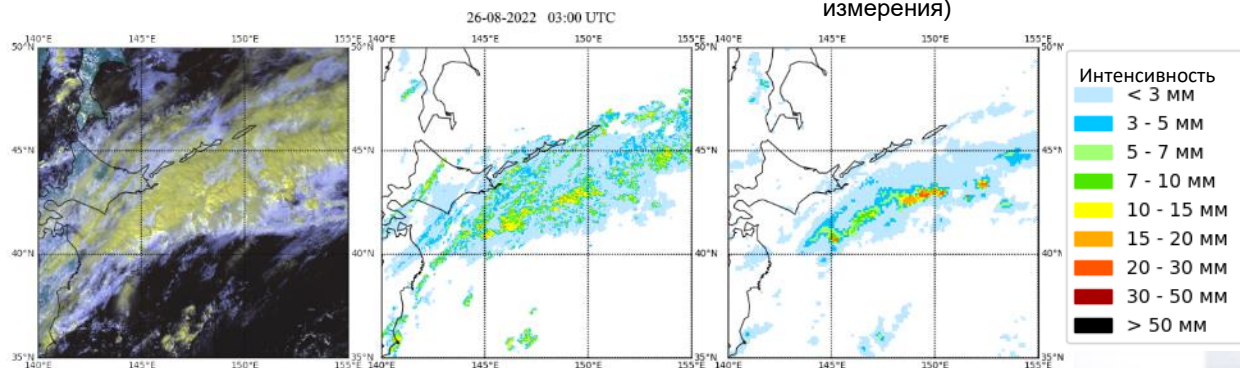
Модель: визуальный трансформер на основе SWIN-UNet для обнаружения осадков, сверточная сеть для оценки интенсивности осадков;

Выходная информация: поля осадков и значения мгновенной интенсивности в мм/ч.

Цветосинтезированное изображение облачности

Разработанный алгоритм по данным Himawari-8/9

GPM-IMERG (спутниковые микроволновые и ИК измерения)



Валидация в сравнении с данными проекта GPM IMERG

	POD	FAR	CSI	F1	RMSE, мм/ч	Bias, мм/ч	Pearson
День	0.76	0.27	0.66	0.75	1.28	-0.43	0.45
Ночь	0.74	0.25	0.69	0.76	1.64	0.44	0.46
Сутки	0.73	0.28	0.64	0.73	1.33	-0.13	0.45

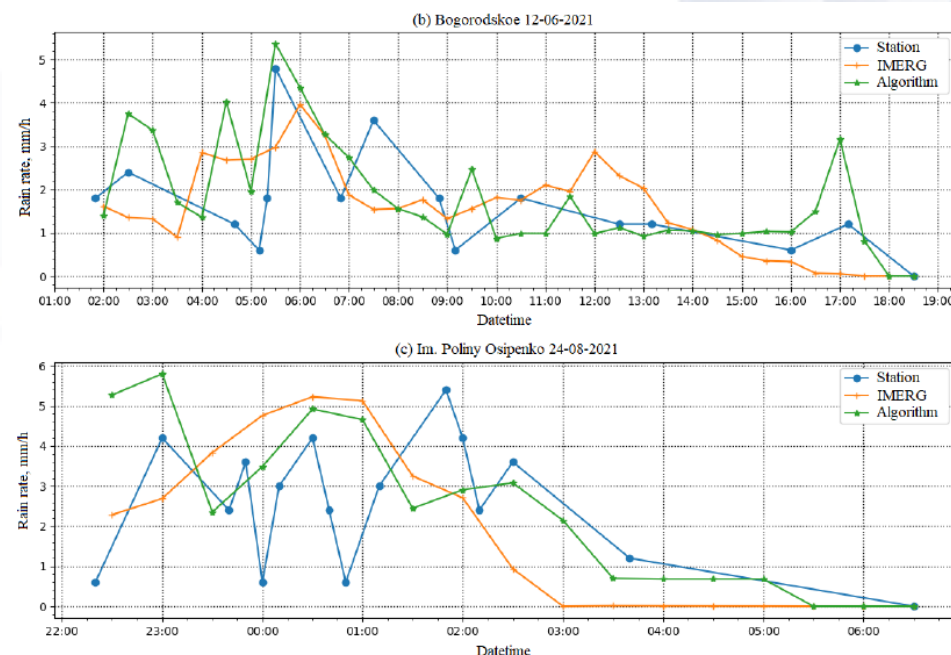
Сравнение накопленных сумм осадков с наземными измерениями на станциях за месяц (август 2023):

Станции (141 насел. пункт): 10033 мм

Алгоритм: 10383 мм

GPM IMERG: 4924 мм

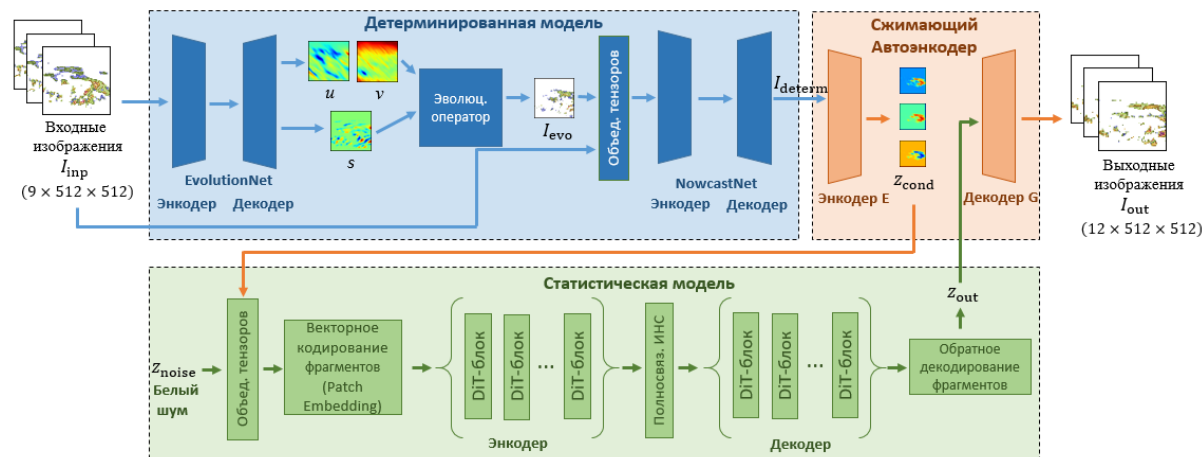
Численная модель CosmoRu-6: 8210 мм



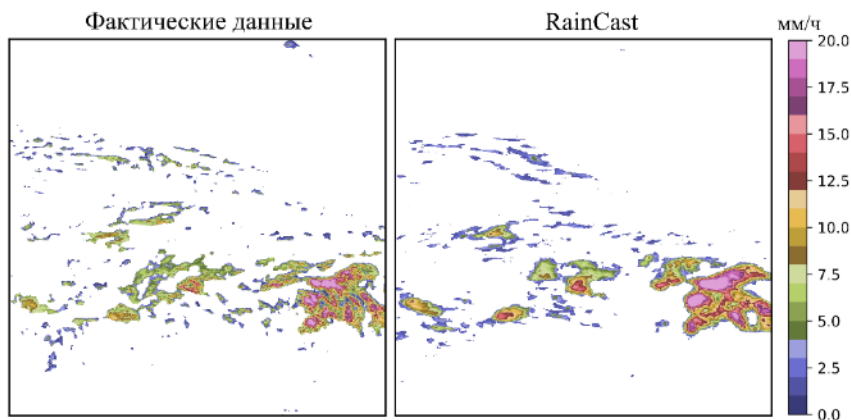
Наукастинг осадков по данным Himawari

Краткосрочный прогноз осадков (наукастинг) на срок до 2 часов

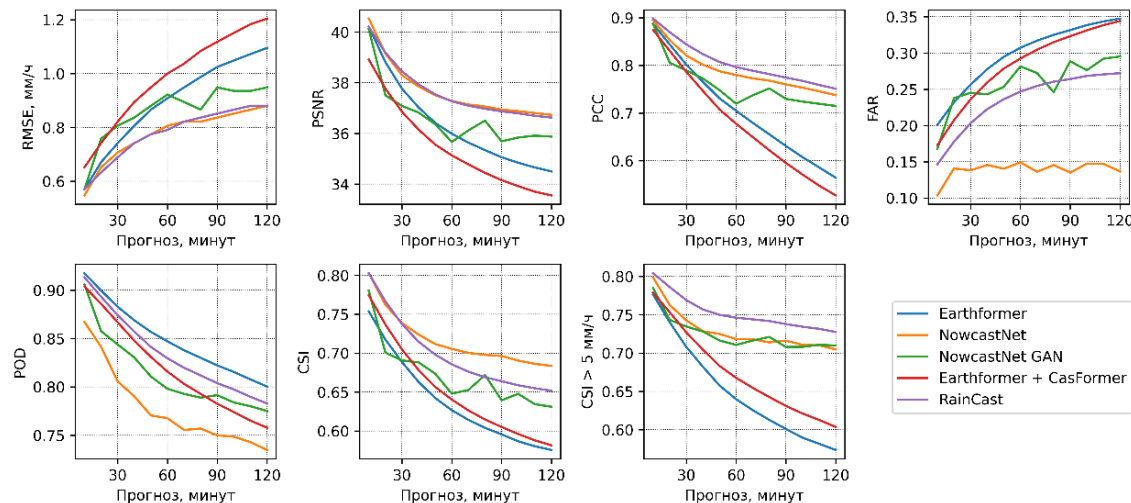
Алгоритм прогноза RainCast



Пример прогноза RainCast на 120 минут



Сравнение с другими актуальными алгоритмами



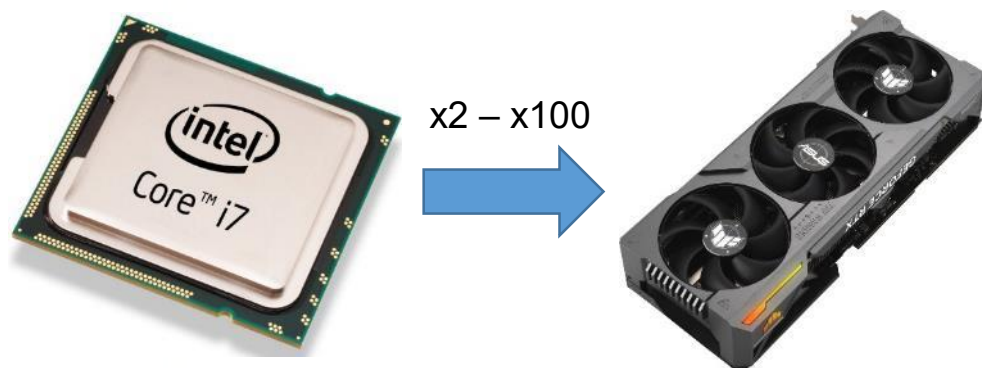
Андреев А. И. и др. RainCast: Гибридный алгоритм наукастинга интенсивности осадков по данным спутниковых наблюдений КА Himawari-8/9 // Информатика и автоматизация. – 2025. – Т. 24. – №. 4. – С. 1085-1113.

Оптимизация процесса обработки спутниковой информации

Проблемы обработки

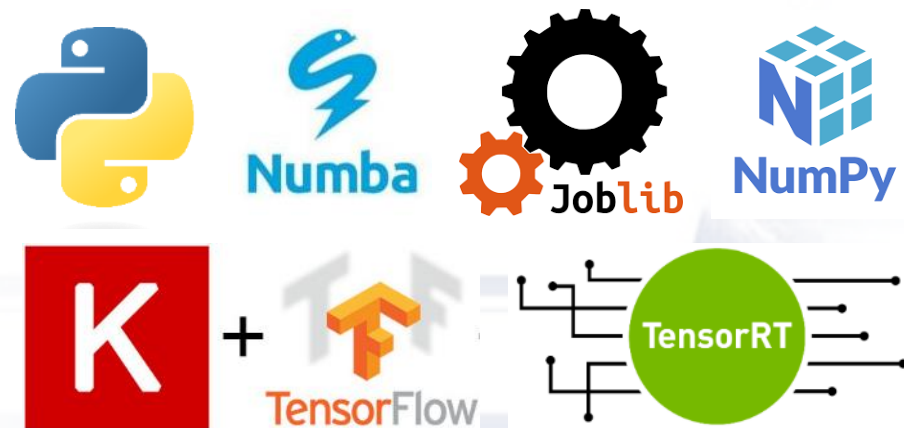
- Долгие вычисления на CPU
- Отсутствие распараллеливания расчётов
- Низкая точность

Аппаратное решение проблем (применение GPU)



- Увеличение скорости обработки однотипных операций для больших объемов данных
- Гибкие возможности для распараллеливания

Программное решение проблем (применение новых библиотек и технологий)



- Применение специализированных библиотек для распараллеливания задач
- Внедрение новых алгоритмов обработки данных

Пример успешной оптимизации

Получение маски облачности по данным МСУ-МР с помощью нейронной сети

Изначальный алгоритм

Время обработки
3 часа

Переход от CPU к GPU



Время обработки
20 минут

Оптимизация архитектуры нейронной сети



Время обработки
7 минут

Переход от Tensorflow к TensorRT



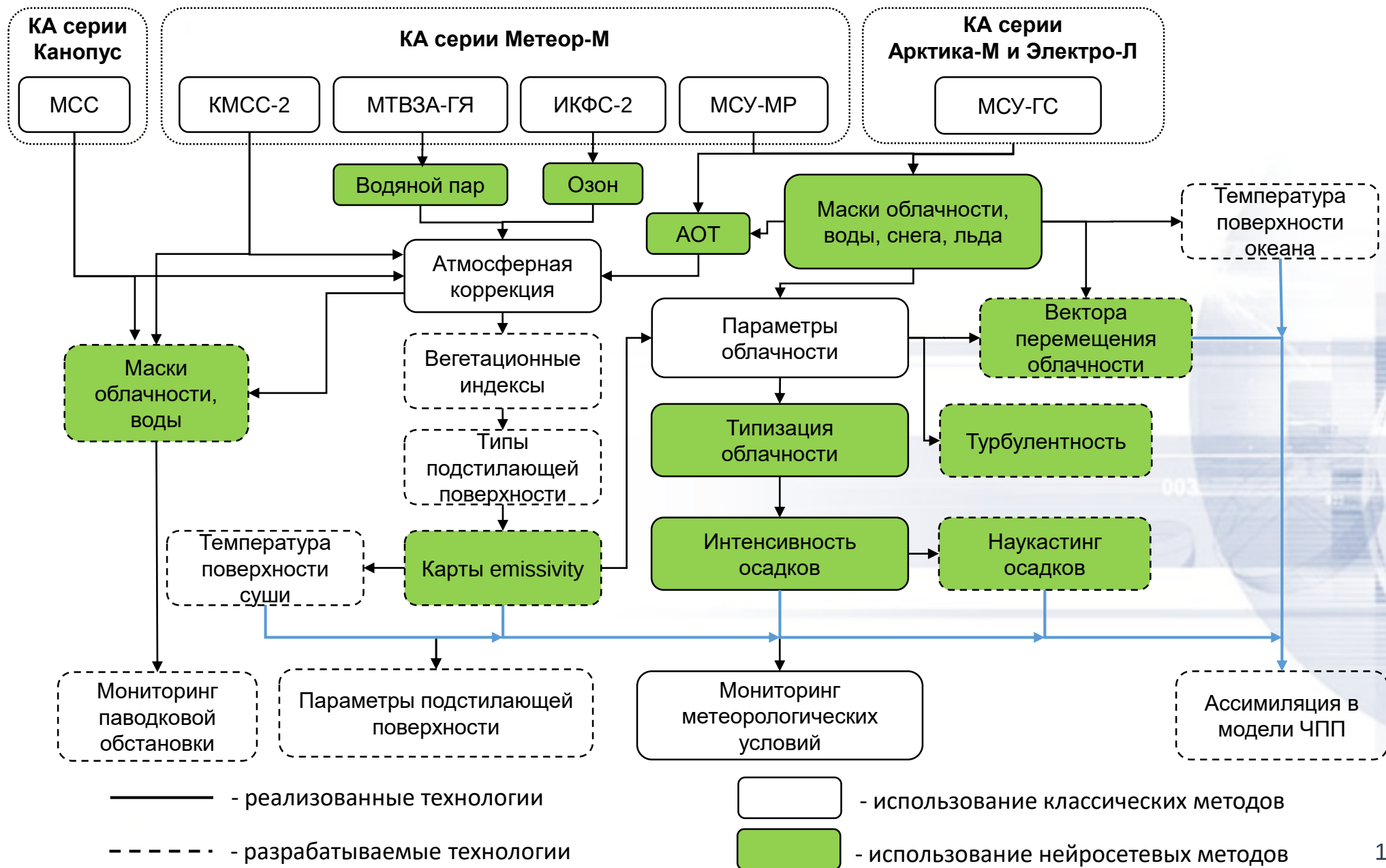
Распараллеливание и оптимизация (Numba, Numpy), изменение квантизации

Время обработки
3 минуты



Время обработки
40 секунд

Комплексное использование данных российских КА





Спасибо за внимание!

Россия, г. Хабаровск, ул. Ленина, д. 18
тел.: 8-(4212) 21-43-11
факс: 8-(4212) 21-40-07
e-mail: m.kuchma@dvrspod.ru