



ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ВЫДЕЛЕНИЯ ПЫЛЕВОГО ЗАГРЯЗНЕНИЯ ОБЛАСТИ ОТКРЫТОЙ ДОБЫЧИ УГЛЯ

Цыгулёв К.С., Окладников В.Е., Смагин С.И.
ВЦ ДВО РАН



Введение



Влияние пыли, возникающей при открытой добыче угля, на прилегающую территорию является важной экологической проблемой и может проявляться в различных аспектах. Угольная пыль может оседать на поверхности растений и почвы. Она может содержать тяжелые металлы и другие токсичные вещества, что загрязняют почву и подземные воды. Это приводит к нарушению фотосинтеза и ухудшению роста растительности. В связи с этим, необходимо проводить своевременную оценку воздействия пылевого загрязнения на прилегающую территорию для принятия мер по минимизации его негативного влияния.



Цель работы



Целью данной работы является исследование возможности применения методов машинного обучения для выделения пылевого загрязнения области открытой добычи угля по данным ДЗЗ.



Исходные данные



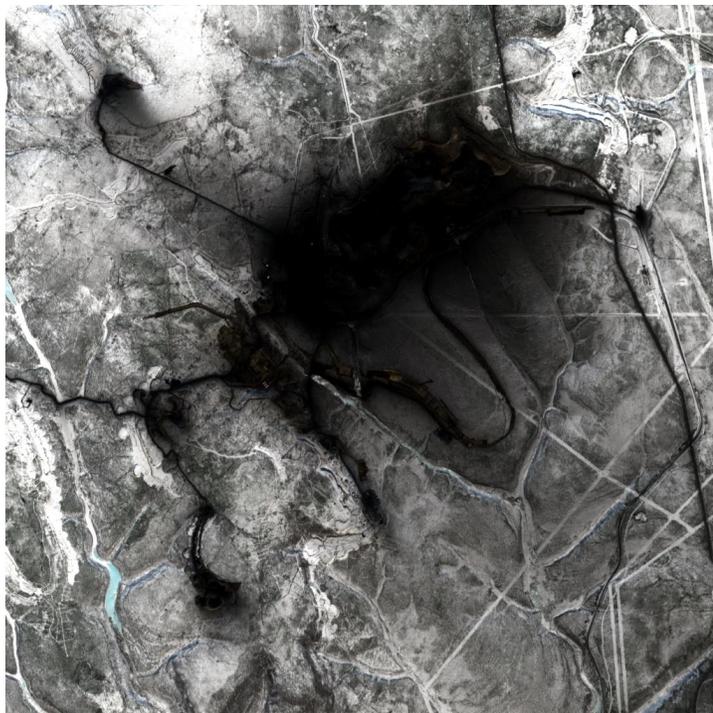
В качестве исходных данных использовалась информация об угольных разрезах в различных субъектах РФ, полученная со спутников Sentinel-2 за периоды начала таяния снега в 2018-2024 гг. В эти периоды пылевое загрязнение видно наиболее отчётливо.

Для составления обучающей выборки отбирались разрезы со следующих областей:

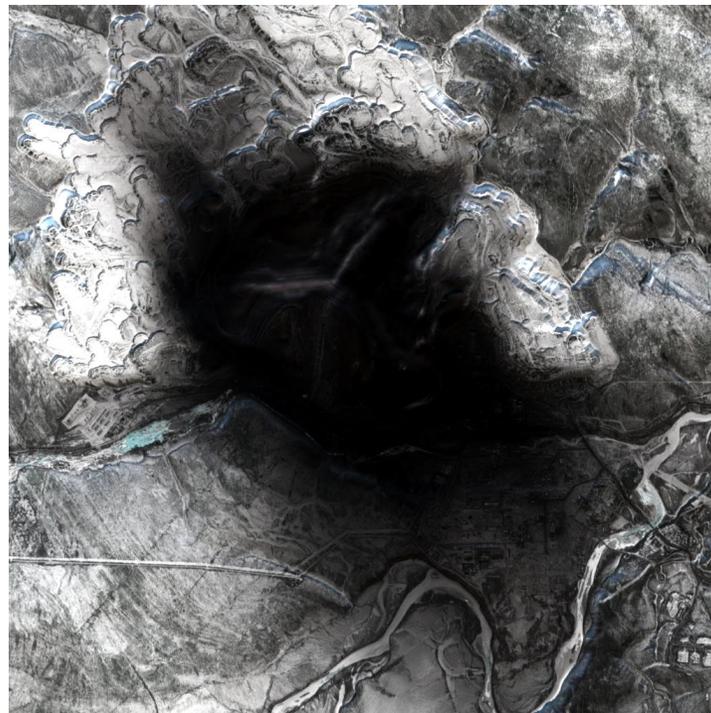
- Республика Саха (Якутия);
- Республика Бурятия;
- Республика Хакасия;
- Кемеровская область.

Для апробации моделей машинного обучения использовались разрезы с Хабаровского края.

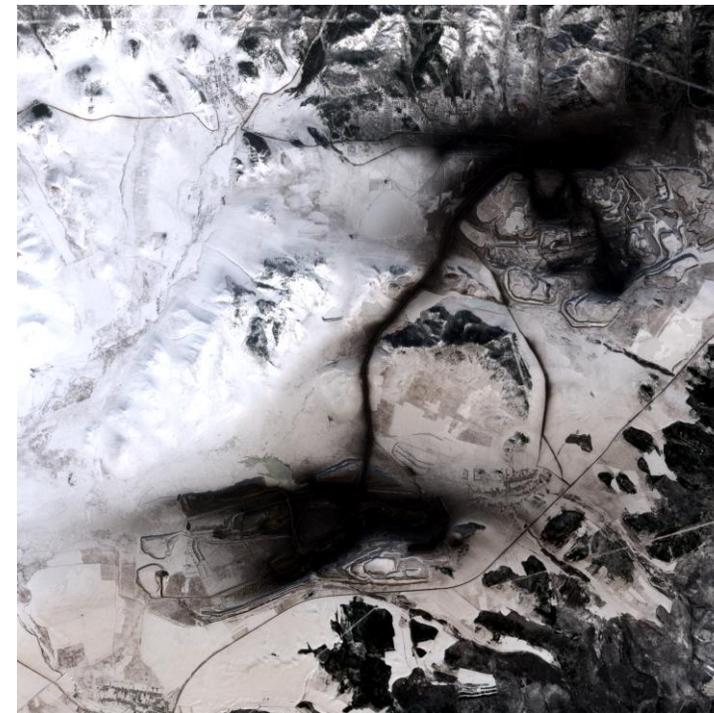
Примеры разрезов



Разрез вблизи поселка Чульман
за 2019 г.

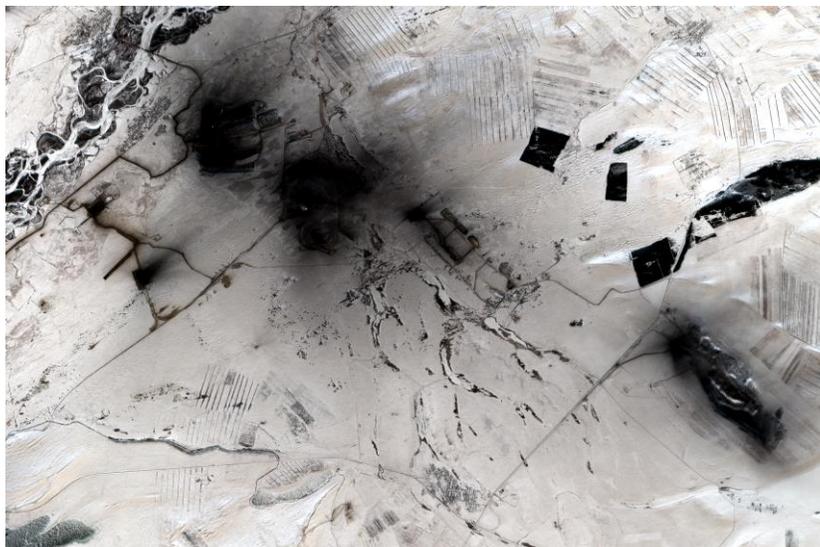


Нерюнгринский разрез за 2020 г.

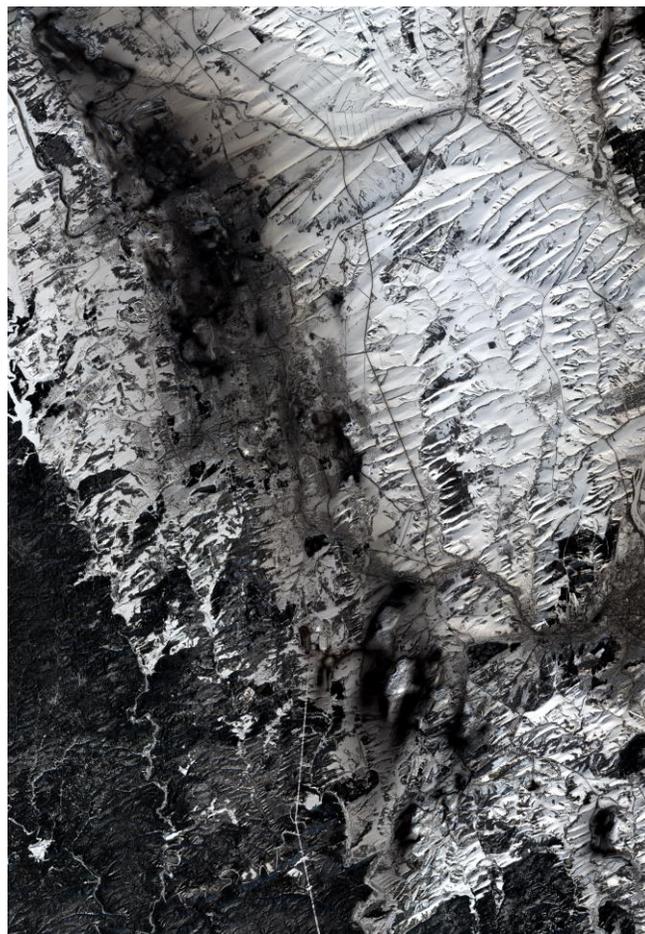


Никольский и Тугнуйский
разрезы за 2023 г.

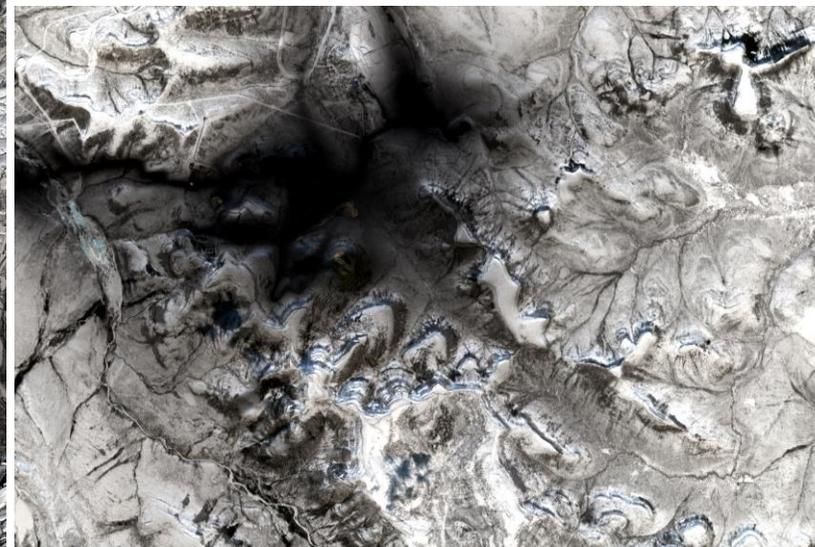
Примеры разрезов



Разрезы вблизи села Аршаново за
2020 г.



Разрезы вблизи поселка
Березово за 2022 г.



Эльгинский разрез за 2019 г.



Random Forest

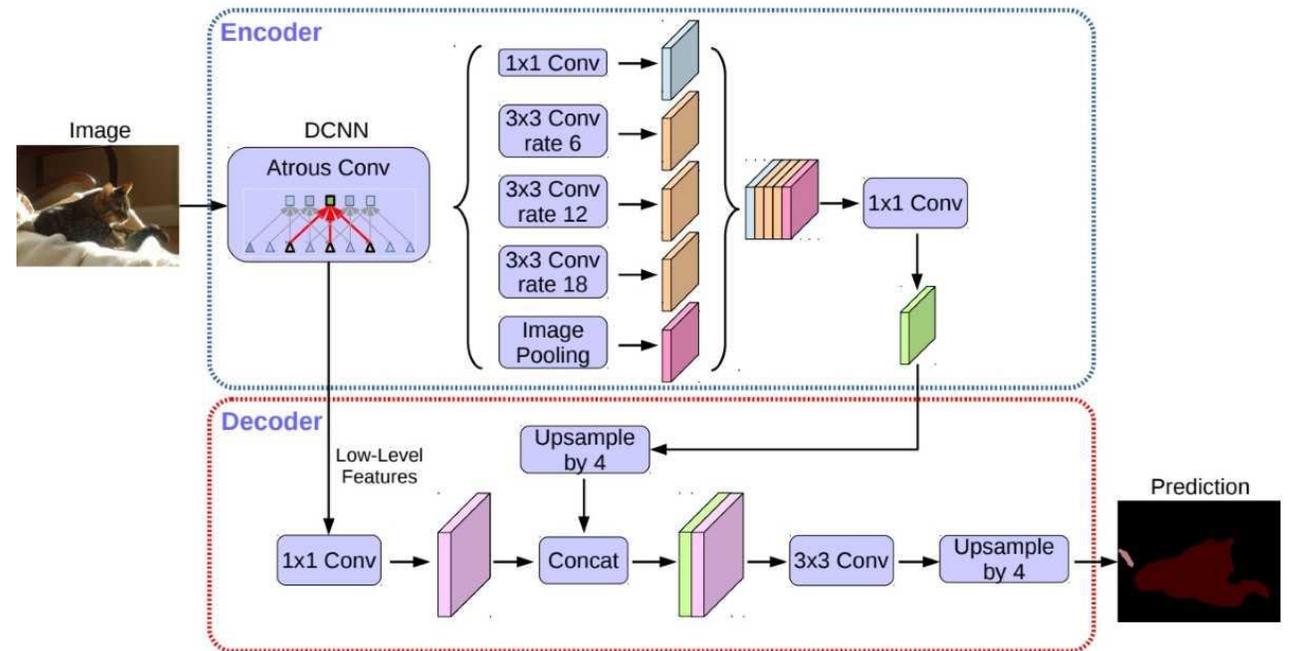


Random Forest – это алгоритм машинного обучения, который использует ансамбли деревьев решений для повышения точности и устойчивости модели. Он строит множество деревьев на случайных подвыборках данных, применяя случайные подмножества признаков для каждого дерева, что снижает риск переобучения. Алгоритм объединяет прогнозы деревьев, выбирая большинство голосов для классификации или среднее значение для регрессии.

DeepLabV3

DeepLabV3 – это модель глубокого обучения для семантической сегментации изображений, разработанная Google. Её основными характеристиками являются:

- Расширенные свёрточные сети (atrous convolutions)
- Многоуровневая обработка с помощью коэффициентов расширения (atrous rates).





Набор данных для обучения



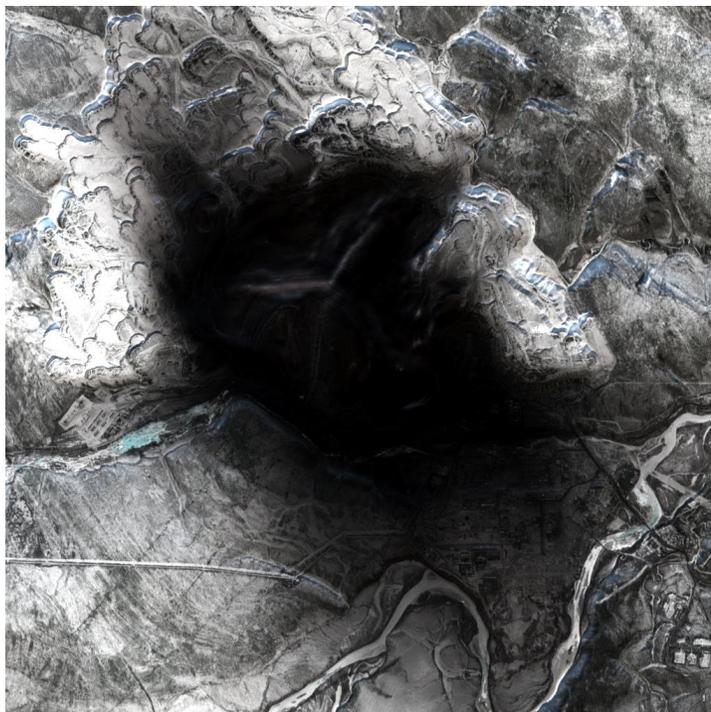
Для формирования обучающей выборки получено 44 спутниковых снимка Sentinel-2, из которых отобраны каналы видимого спектра (B₂, B₃, B₄), канал ближнего инфракрасного спектра (B₈) и два канала коротковолнового инфракрасного спектра (B₁₁, B₁₂).

Далее, полученные данные обрезались, а затем размечались двумя способами: вручную и с помощью индекса Enhanced Coal Dust Index (ECDI) путем его бинаризации по порогу больше нуля. Формула этого индекса выглядит следующим образом:

$$ECDI = (B_{11} - B_8 + B_{12}) / (B_{11} + B_8 - B_{12})$$

Для обучения DeepLabV3 данные дополнительно разрезались на квадраты размером 512x512 пикселей.

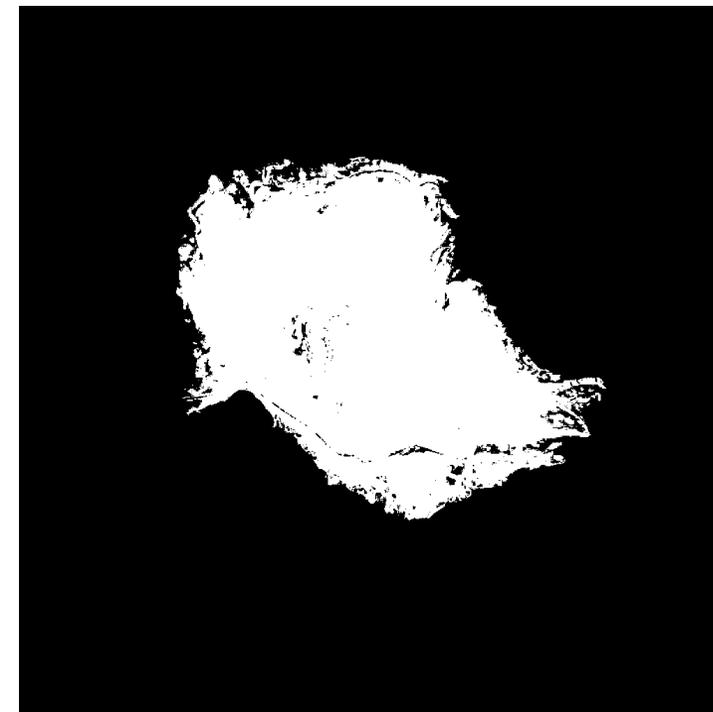
Примеры разметки данных



Исходное изображение



Область, размеченная вручную



Область, размеченная с
помощью ECDI



Результаты обучения



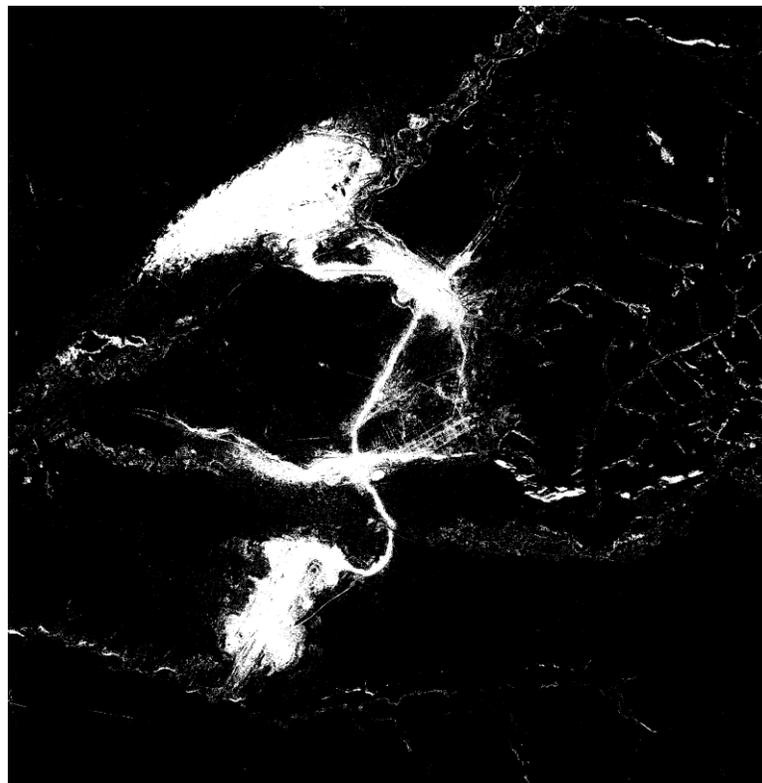
Точность модели DeepLabV3, обученной на вручную размеченной выборке, составила ~61%. Точность этой же модели, обученной выборке, на размеченной с помощью ECDI, составила ~75%. Оценка проводилась с помощью метрики Intersection over Union.

Точность модели случайного леса составила менее 60% независимо от выборки.

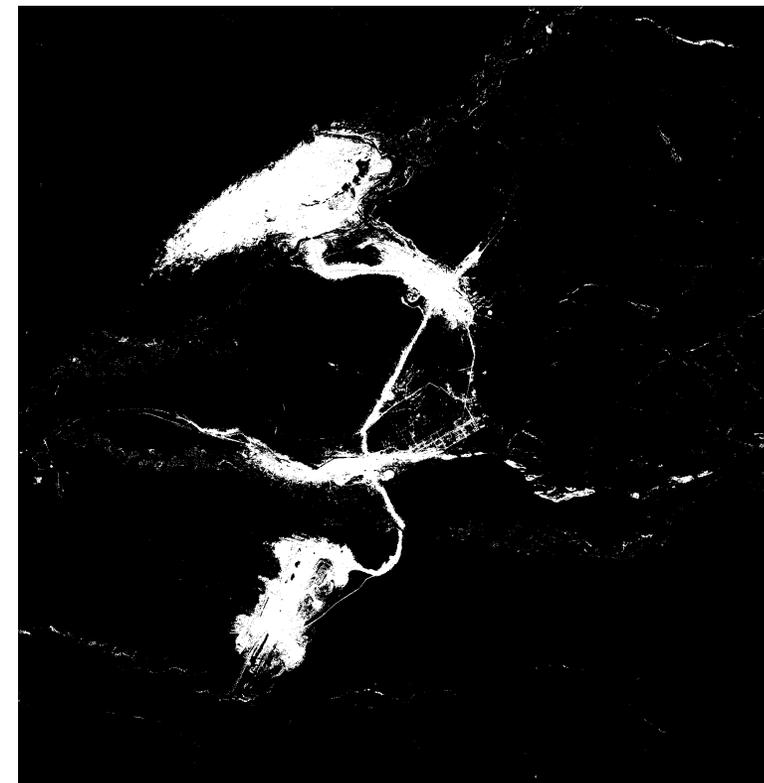
Апробация Random Forest



Исходное изображение



Результат обучения на данных с
ручной разметкой



Результат обучения на данных с
разметкой по ECDI

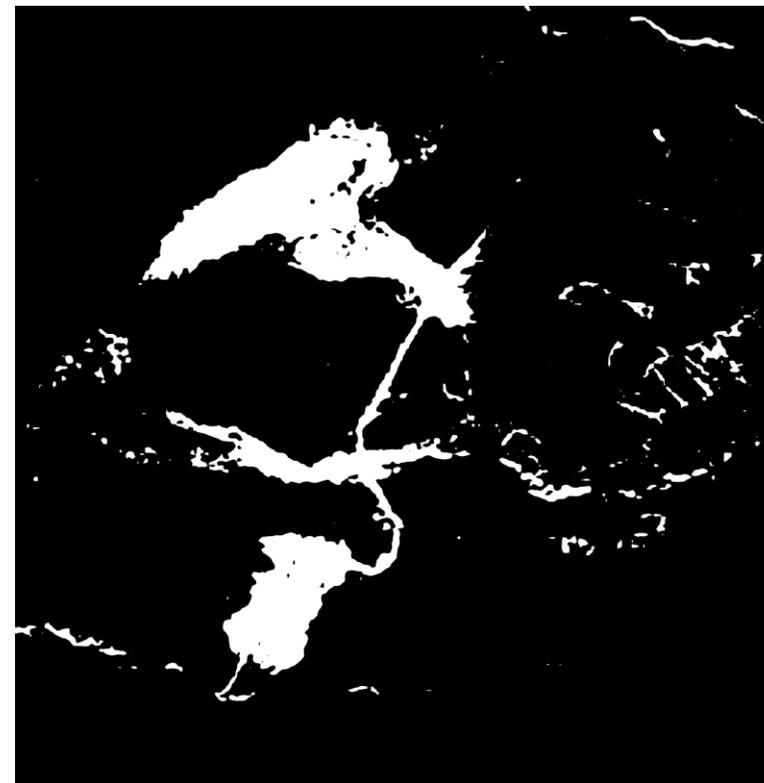
Апробация DeepLabV3



Исходное изображение



Результат обучения на данных с
ручной разметкой



Результат обучения на данных с
разметкой по ECDI



Заключение



По результатам проведенного исследования можно сделать вывод, что методы машинного обучения являются перспективным инструментом для выделения области пылевого загрязнения при открытой добыче угля.

При детектировании пылевого загрязнения с помощью методов машинного обучения:

- снизились временные затраты;
- повысилась точность определения границ области;
- появилась возможность проведения мониторинга области за большой период времени.

В будущем методика войдет как этап в алгоритм оценки состояния растительности, находящейся под влиянием пылевого загрязнения.



Для выполнения расчетов были использованы вычислительные ресурсы
ЦКП «Центр данных ДВО РАН».

Исследования выполнены при поддержке Российского научного фонда
(проект 24-11-20029)

СПАСИБО ЗА ВНИМАНИЕ