

# **Применение методов машинного обучения для оценки количества биомассы по данным спутниковых снимков Sentinel-1 и Sentinel-2**

Автор: Денис Петренко

Научный руководитель: Алексей Озерин

## **Описание проведенных экспериментов**

### **S1-baseline**

Ранее для решения задачи предлагался baseline подход на основе данных спутниковых снимков Sentinel-1. К сожалению, такой подход нежизнеспособен из-за особенностей природы таких снимков.

В SAR изображениях присутствует спекл-эффект (т.е. случайная интерференционная картина), эффект от которого варьируется от пикселя к пикселю. Наличие интерференционной картины в амплитуде принимаемого сигнала приводит к тому, что значения признаков Sentinel-1 порождены гамма распределением. Таким образом не представляется возможным использовать данные этого спутника напрямую.

### **S2-baseline**

При исследовании причин неудовлетворительных результатов baseline модели было обнаружено, что ввиду природы данных снимков спутника Sentinel-1, их нельзя использовать напрямую без предобработки. Для проверки работоспособности текущей модели были использованы данные снимков Sentinel-2. Для отсутствующих снимков была симитирована экстремальная облачность: канал CLP был заполнен максимально допустимым значением 255, остальные каналы заполнены нулями. Данное восстановление пропусков позволило использовать модель S1-baseline без существенных изменений.

## Анализ результатов

Переход на снимки Sentinel-2 принес значительное увеличение качества модели: loss уменьшился до 27.185242, метрик ка до 42.7747. Детальный анализ снимков, сгенерированных моделью и сравнение с таргетом выявил системное занижение предсказаний. На рисунке ниже приведен пример целевого изображения и изображения полученного от модели. Можно заметить, что модель систематически занижает прогнозы для зон с высоким уровнем таргета. Также регионы с нулевым таргетом на ответе модели имеют большую площадь чем целевом изображении. Например, дороги существенно шире, чем на самом деле.

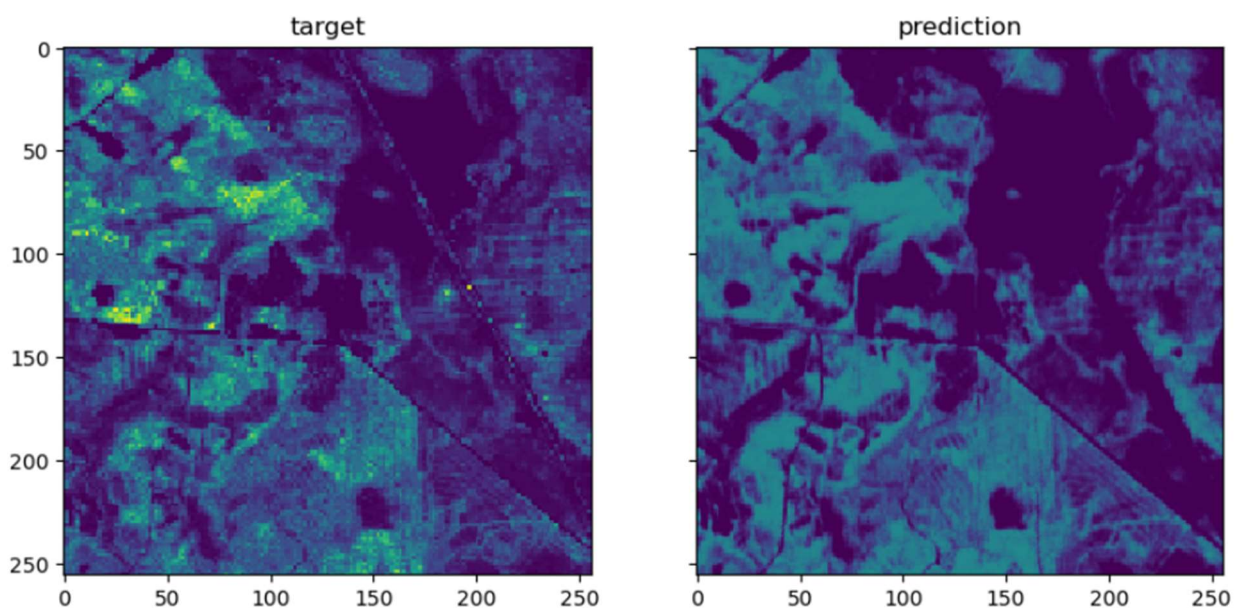


Рисунок 1. Пример целевого изображения и предсказания для участка fed6bb57

Озвученные выводы подтверждаются распределением таргета и прогноза. Мы наблюдаем, что количество пикселей со значениями близкими к нулю существенно выше у модели, чем у реального изображения.

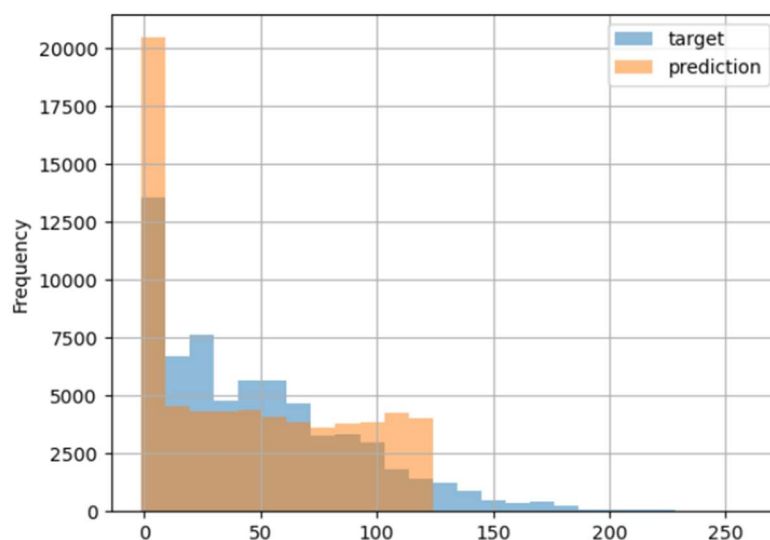


Рисунок 2. Распределение таргета и предсказания для участка fed6bb57

## S2-baseline + стратификация выборки

При анализе результатов работы s2-baseline модели мы обратили внимание на то, что прогнозы модели ограничены сверху. Одной из причин такого поведения может быть тот факт, что пикселей с небольшими значениями существенно меньше, поэтому при формировании батчей для обучения и валидации у нас отдается предпочтение пикселям с небольшими значениями. Для того чтобы поднять качество предсказания пикселей со значительными значениями мы пересобрали батчи для обучения и валидации со стратификацией пикселей по значению таргета. Для каждого участка проводится анализ распределения пикселей по бинам размера 100 в диапазоне от 0 до 1000. Пиксели со значениями больше 1000 помещаются в отдельный бин. Для каждого бина определяется количество пикселей, которое будет выбрано в батч для обучения или валидации.

## Анализ результатов

Добавление стратификации по таргету привело к увеличению качества модели: loss уменьшился до 23.128957, а итоговая метрика уменьшилась до 36.6443.

Детальный анализ снимков, сгенерированных моделью и сравнение их с таргетом показал, что модель стала лучше предсказывать пиксели с большим значением. На рисунке ниже приведен пример целевого изображения и изображения, полученного

от модели. Можно заметить, что значения модели все еще ниже, чем целевое изображение. Остальные артефакты остались без изменений.

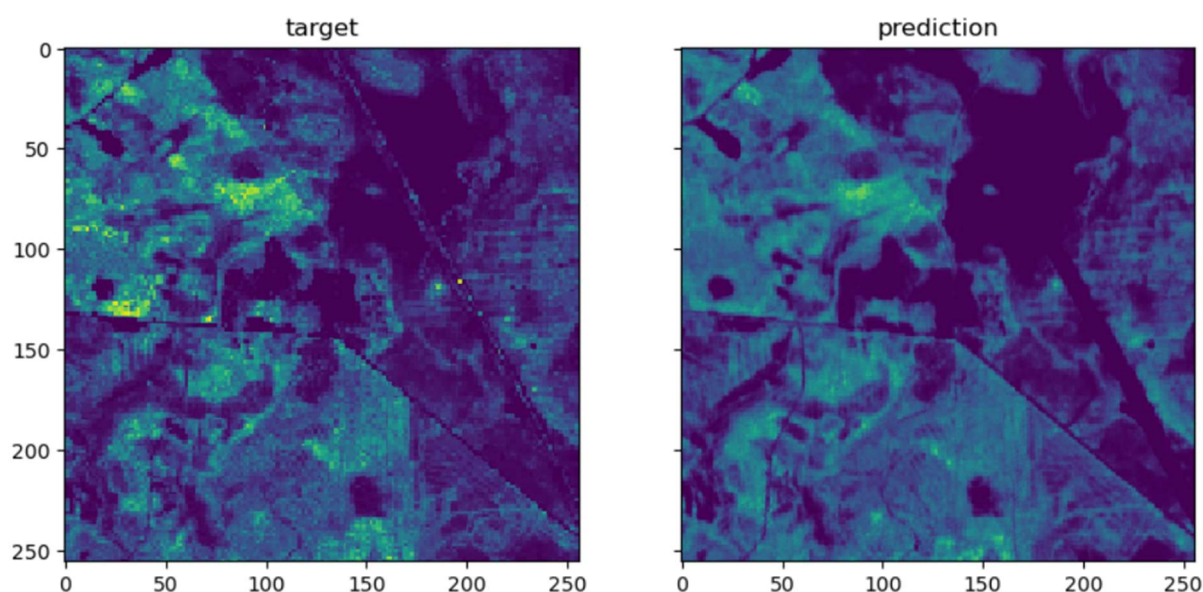


Рисунок 3. Пример целевого изображения и предсказания для участка fed6bb57

Озвученные выводы подтверждаются распределением таргета и прогноза. Мы наблюдаем, что количество пикселей со значениями близкими к нулю все еще выше таргета, но значительно ниже, чем у s2-baseline модели.

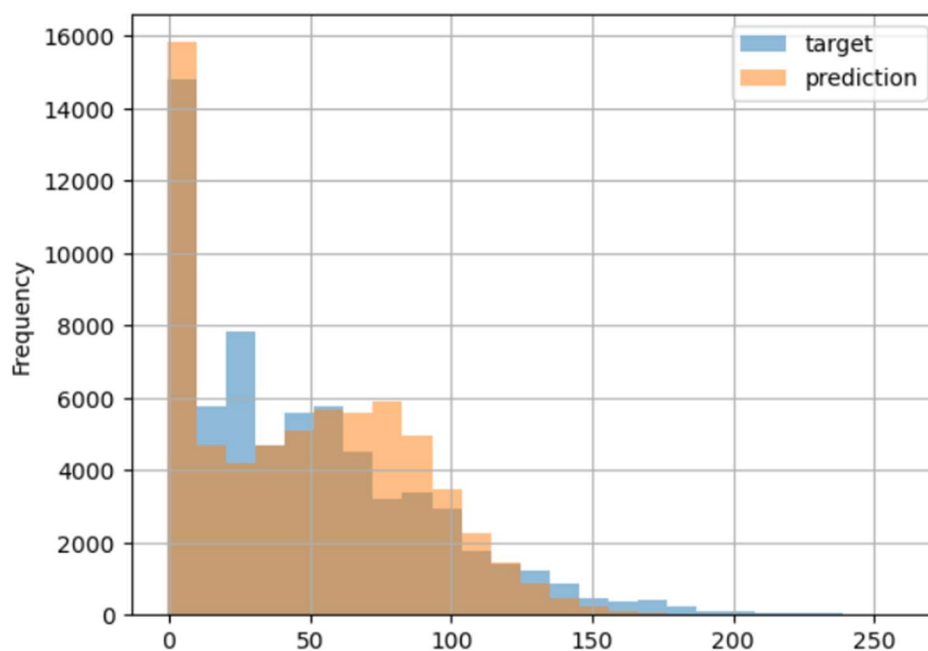


Рисунок 4. Распределение таргета и предсказания для участка fed6bb57

## Magnitude-Signal модель

Несмотря на то, что добавление стратификации улучшило модель, мы наблюдаем, что распределение таргета и распределение

предсказания все еще имеют отличия. Для того чтобы улучшить качество предсказаний в широком диапазоне целевой переменной предлагается использовать на выходе два линейных слоя: один для определения уровня ответа(амплитуда), а второй для его корректировки(сигнал). Тогда ответ модели будет произведением амплитуды и сигнала.

## Анализ результатов

Применение описанной модели привело к ухудшению качества модели: loss увеличился до 24.481165, а итоговая метрика увеличилась до 37.8875.

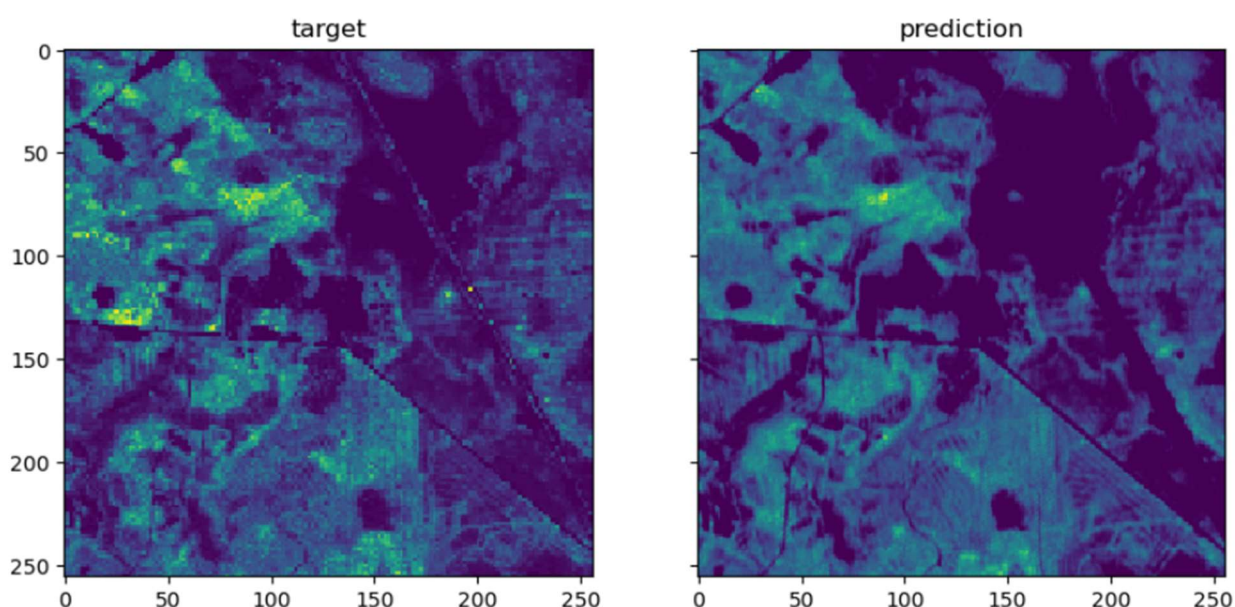


Рисунок 5 Пример целевого изображения и предсказания для участка fed6bb57

Анализ полученных изображений позволяет сделать вывод о том, что такая модель строит заниженные прогнозы для небольших значений таргета.

Распределение таргета и прогнозных значений иллюстрирует, что в области небольших значений целевой переменной модель отдает предпочтение прогнозам близким к нулю. При этом следует отметить, что распределение предсказаний достаточно хорошо повторяет распределение таргета для диапазона значений от 100 до 150.

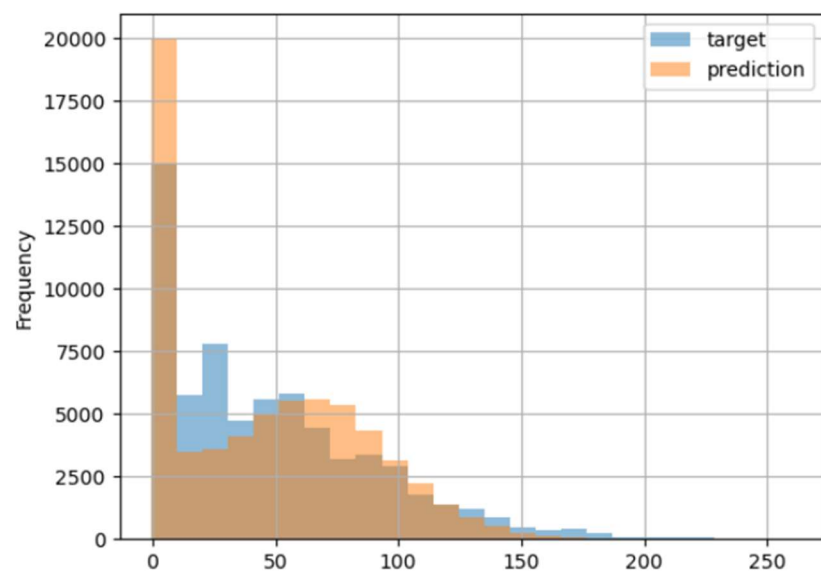


Рисунок 6 Распределение таргета и предсказания для участка fed6bb57