



Задача «Классификация опор контактной сети и других объектов в массиве точек лазерных отражений для обеспечения безопасности движения железнодорожного транспорта»

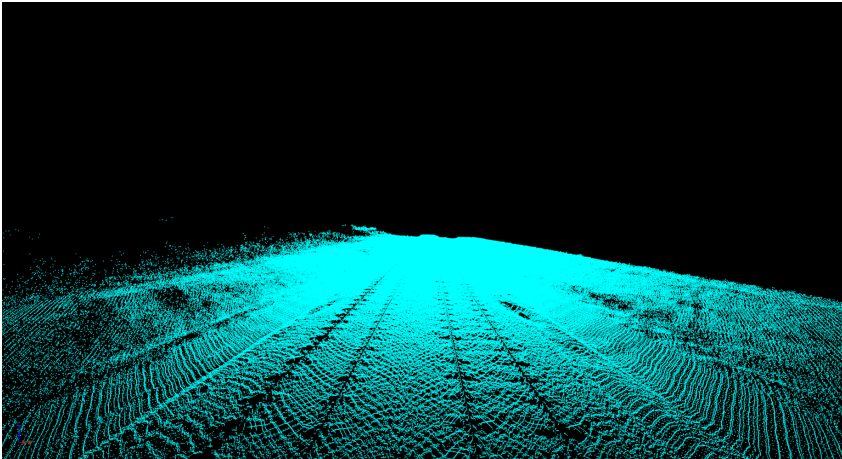
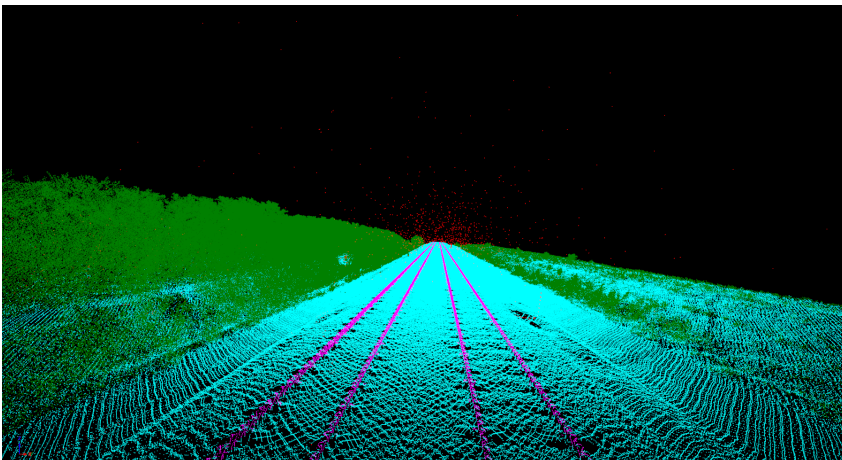
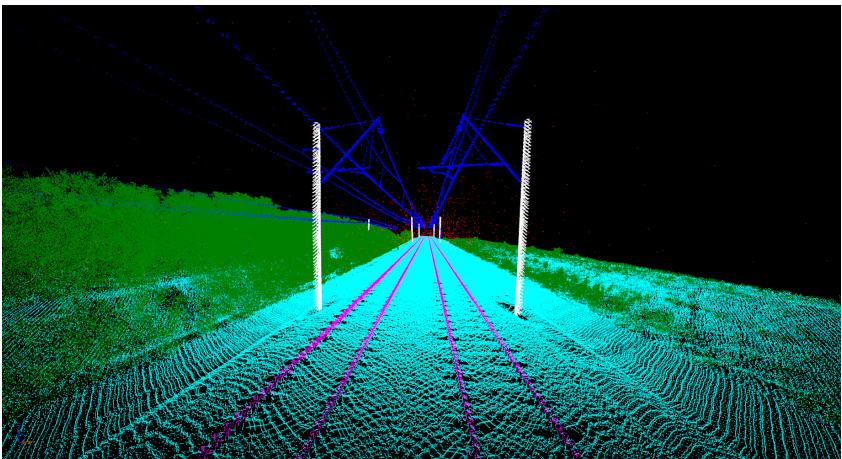
Введение

Российский железнодорожный комплекс — один из крупнейших в мире. Общая протяженность эксплуатационной длины железнодорожных путей превышает 120 тысяч километров. В прошлом году в нашей стране поезда и электрички перевезли более 1 миллиарда пассажиров. Именно поэтому к проектированию и эксплуатации железнодорожной инфраструктуры предъявляются высокие требования по обеспечению безопасности.

В связи с совершенствованием технологий геодезической съемки, инновационным решением при диагностике, строительстве и эксплуатации железнодорожных путей и придорожной инфраструктуры является построение цифровых моделей пути (ЦМП) на основе данных лазерного сканирования. Результат лазерного сканирования — это высокоточный массив точек лазерных отражений (ТЛО), на основе которого проводятся все необходимые измерения и последующий анализ данных. Определение опор контактной сети (ОКС) в массиве ТЛО — это одна из обязательных подзадач процесса определения габаритов приближения подвижного состава к ОКС.

В настоящий момент определение взаимного расположения объектов железнодорожной инфраструктуры происходит при натурном обследовании — в связи с чем высок риск получения недостоверной информации, что критично для обеспечения безопасности движения.

Стоит понимать, что точки не расположены случайным образом, они правдоподобно отражают положение объектов в пространстве. Для того, чтобы лучше понять, как распределены классы и какие ТЛО им принадлежат, рассмотрим следующие картинки. В них пошагово добавляются новые классы для наглядности.

	<p>Точки, принадлежащие земле</p>
	<p>Точки, принадлежащие земле, растительности, рельсам, шуму</p>
	<p>Точки, принадлежащие земле, растительности, рельсам, шуму, столбам, проводам контактной сети</p>

Условие задачи

Участникам чемпионата предстоит создать свое решение по прогнозированию класса ТЛО с использованием параметра этой точки.

Описание входных значений

train.csv — файл, содержащий данные для тренировки. Структура датасета:

- Параметр «id» — уникальный идентификатор ТЛО.
- Параметр «Easting» — направление на восток, условная координата X ТЛО в метрах
- Параметр «Northing» — направление на север, условная координата Y ТЛО в метрах
- Параметр «Height» — отметка (высота), условная координата H ТЛО в метрах
- Параметр «Reflectance» — отражательная способность ТЛО
- Параметр «Class» — признак класса точки

test.csv - содержит данные ТЛО для предсказания

sample_solution.csv — пример файла для отправки

Пояснение к данным

Столбец «Class» хранит в себе тип класса точки, где:

0 – точки земли

1 – точки опор

3 – точки растительности

4 – точки рельсов

5 – точки элементов контактной сети

64 – точки шумов

Метрика

В качестве метрики выступает Recall по столбцу Class.

Recall считается как:

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

где: TP (True Positive) — количество верно угаданных значений одного класса

FN (False Negative) — количество неправильно угаданных значений класса

Правила чемпионата:

1. С момента открытия датасета до момента завершения приема решений репозиторий участника, в котором он ведет разработку по задаче текущего чемпионата, должен оставаться закрытым.
2. Участник обязан открыть доступ к репозиторию на чтение по ссылке (которая была прикреплена в ЛК в поле «Ссылка на код (гитхаб)») не позднее чем в течение 12 часов с момента окончания дедлайна отправки решений на региональном чемпионате.
3. Согласно п. 5.8 Положения в процессе верификации решений организаторы и технические эксперты, проверяющие решения участников, в праве назначить интервью с участниками чемпионата. Участник получит приглашение и ссылку на интервью не позднее чем за 12 часов до публикации итогового лидерборда. Пропуск интервью участником является поводом для дисквалификации.
4. Организаторы вправе исключить участника из призовых позиций лидерборда за непредоставление одного из артефактов решения задачи: тизера, скринкаста, презентации, ссылки на репозиторий.
5. Организаторы вправе дисквалифицировать участника в случае выявления плагиата кода или несоблюдения Положения конкурса.
6. Участник, получивший 2 дисквалификации за сезон конкурса, попадает в чёрный список с дальнейшим отстранением от участия в чемпионатах до конца сезона.