

СРАВНЕНИЕ МОДЕЛЕЙ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ ДЛЯ ЗАДАЧИ АВТОМАТИЧЕСКОГО ВЫДЕЛЕНИЯ ТРЕКОВ МОД НА ИОНОГРАММАХ РАДИОЗОНДИРОВАНИЯ ИОНОСФЕРЫ

Масюченко Степан Вадимович

Магистрант

Национальный исследовательский университет "МЭИ"

Щирый Андрей Олегович

канд. техн. наук

Институт земного магнетизма, ионосферы и распространения радиоволн РАН им. Н.В.
Пушкова

г. Москва

1. Введение

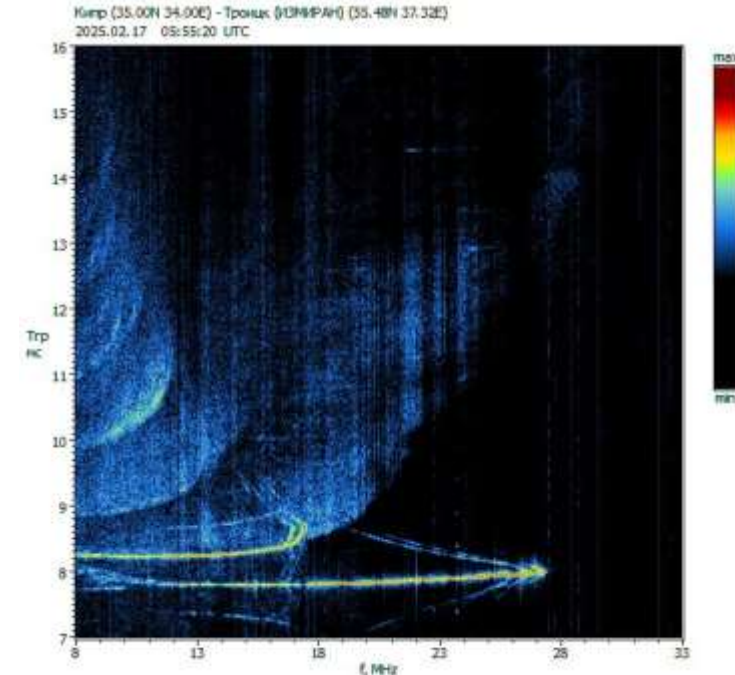
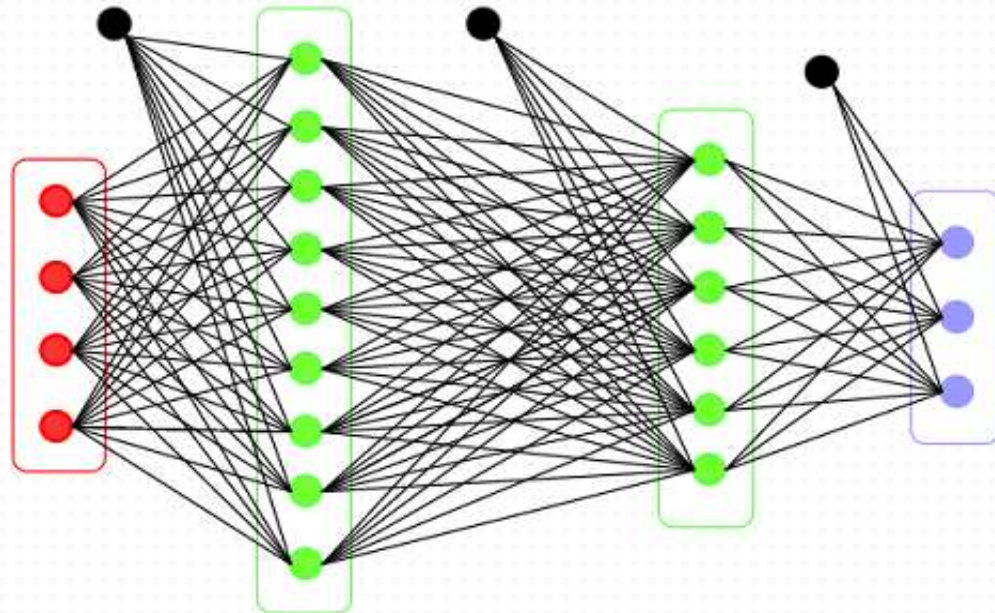
Несмотря на то что современные методы радиозондирования ионосферы обеспечивают получение детализированных ионограмм, их ручная обработка и интерпретация остаётся трудоёмким процессом.

В данной работе предлагается подход, основанный на применении моделей машинного обучения, для задачи сегментации и классификации треков мод на ионограммах.

Планируется построить систему, состоящую из набора моделей, которая будет выполнять следующие функции:

1. Сегментировать трек и выделить полезный сигнал на ионограмме.
2. Корректно классифицировать все треки мод, учитывая их взаимное расположение.

В данной работе, будет показаны метрики доработанной модели сегментации, а так же MVP модели классификации треков.



2. МОДЕЛЬ СЕГМЕНТАЦИИ ТРЕКОВ

2.1 НАБОР ДАННЫХ

Входной набор данных, содержит изображения, на которых вручную размечены пиксельной сегментационной маской, полезный сигнал (треки мод).

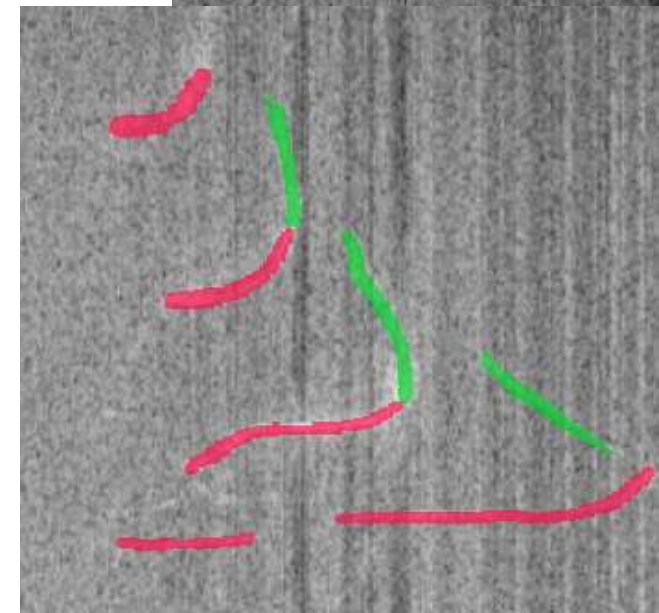
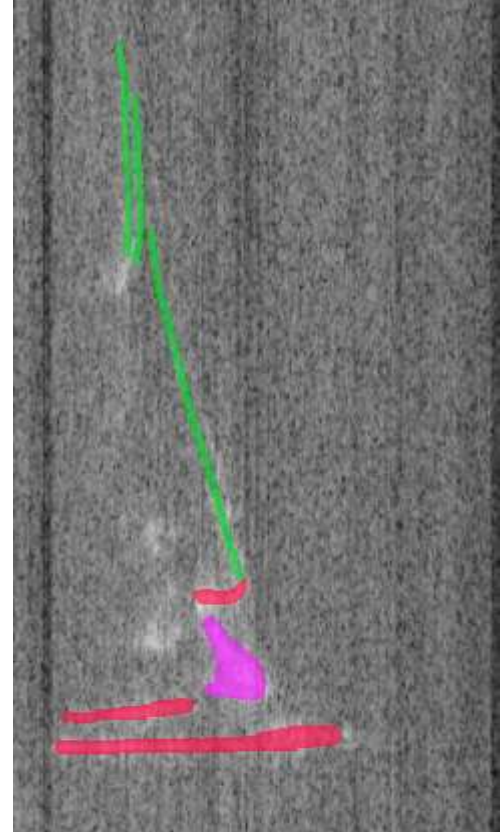
Всего было выделено 3 класса треков мод:

1. L - нижний трек (Красный)
2. U - верхний трек (Зелёный)
3. N - Неопределённый трек (Фиолетовый)

На текущий момент, в наборе данных содержится 1170 ионограмм. При обучении моделей, было принято решение не выделять отдельно тестовый набор данных, так как выделение дополнительного тестового набора данных, привело бы к значительном сокращению обучающей выборки, что негативно сказалось бы на качестве модели. Финальная оценка, будет проводится на валидационном наборе данных.

В дальнейшем будут предоставлены результаты моделей, которые обучались на 2 вариациях набора данных : первый вариант включает три класса (L/U/N), а второй - только один класс (Т - трек). Вторая вариация датасета, была получена заменой всех классов на один из первой вариации.

В обоих случаях, набор данных будет разделён по следующему соотношению : train/val - 994/176 (0,15/0,85). Набор данных хранится в формате YOLO.

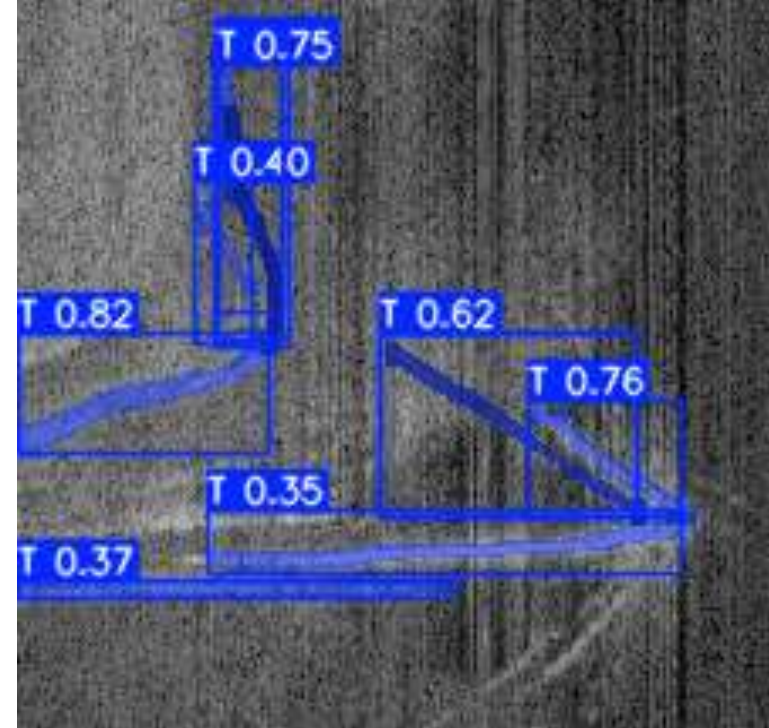


2.2 ОПИСАНИЕ МОДЕЛИ И РЕЗУЛЬТАТЫ

При обучении использовалась модель YOLO26m-seg, обновлённые блоки которой, извлекают более глубокие признаки чем все предыдущие модели семейства.

Стоит учитывать, что метрики новой модели считались на большем количестве изображений - 176 против 20 у старых моделей.

Модель обученная на 1 классе, показала рекордные значения метрик, в дальнейшем, будет использоваться именно она, т.к. в текущей логике реализации системы, классификацией занимается отдельная модель.



НС	Набор данных	map (box)	map_50 (box)	map (L-class) (box)	map (U-class) (box)	map (N-class) (box)	map (mask)	map_50 (mask)	map (L-class) (mask)	map (U-class) (mask)	map (N-class) (mask)
Mask RCNN	273 изобр.	0.29	0.61	0.36	0.31	0.20	0.15	0.44	0.26	0.01	0.18
YOLO_v11_n	273 изобр.	0.31	0.57	0.32	0.43	0.18	0.16	0.48	0.23	0.12	0.14
YOLO_v11_s	273 изобр.	0.31	0.67	0.33	0.46	0.14	0.16	0.52	0.24	0.11	0.13
YOLO_v11 m	273 изобр.	0.32	0.58	0.37	0.34	0.14	0.17	0.55	0.23	0.15	0.13
YOLO_v11_s	546 изобр. ауг.	0.36	0.66	0.72	0.78	0.47	0.21	0.60	0.27	0.20	0.17
Mask RCNN	546 изобр. ауг.	0.35	0.67	0.36	0.34	0.34	0.18	0.49	0.26	0.01	0.28
YOLO_v26_m	1170 и. (3 class)	0.34	0.614	0.43	0.35	0.24	0.24	0.59	0.31	0.2	0.2
YOLO_v26_m	1170 и. (1 class)	0.39	map_50 (box) - 0.751				0.272	map_50 (mask) - 0.684			

3. МОДЕЛЬ КЛАССИФИКАЦИИ ТРЕКОВ

3.1 НАБОР ДАННЫХ (НА ОСНОВЕ ДАННЫХ ИЗМИРАНА)

Скриптом из данного набора, по фильтру ручной разметки, было отобрано 2466 ионограмм.

Набор данных был приведён к следующему формату:

1. Папка с изображениями, сохранёнными в цветовой гамме gray.

2. Папка с аннотациями, в которой хранится необходимая информация, для каждого изображения.

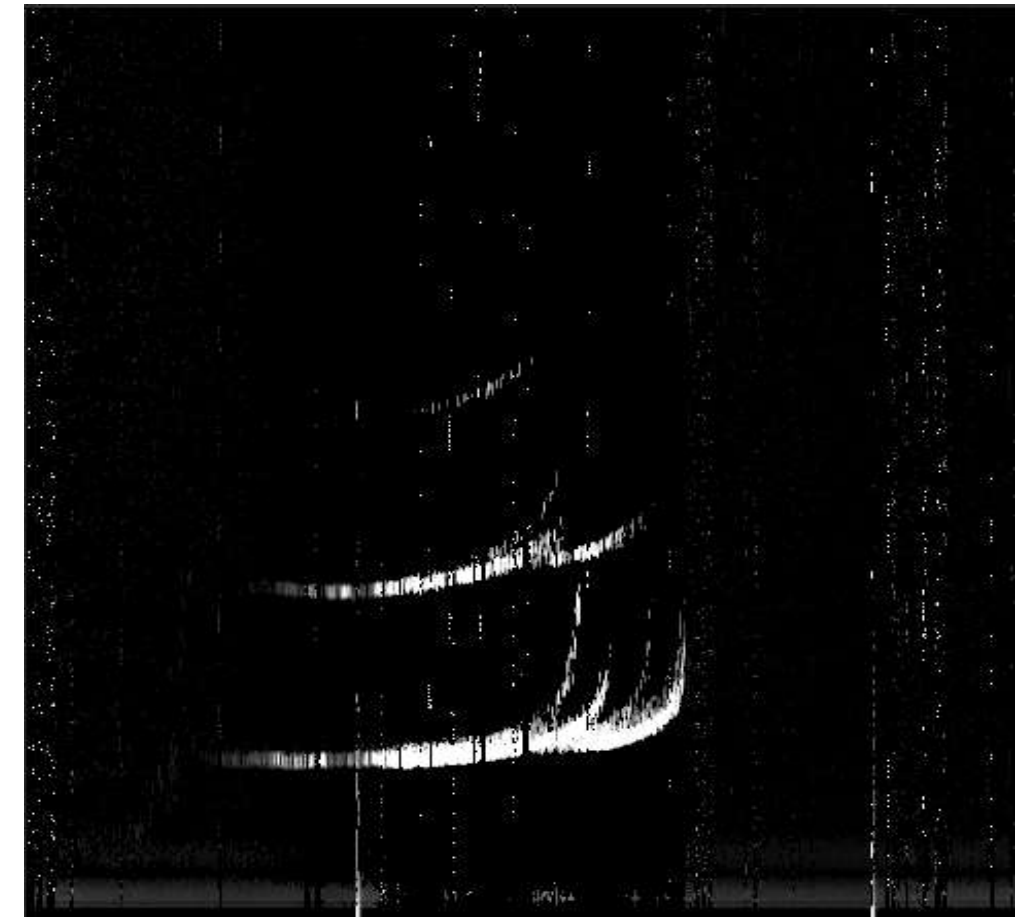
На основе разметки ломанной линии, которая изначально была в данных измирана, были сформированы BBOX в следующем формате : нормализованные от 0 до 1 координаты (x1, y1, x2, y2), где (x1, y1) - координаты нижнего левого угла бокса, (x2, y2) - координаты верхнего правого угла бокса.

В датасете содержится 6 различных классов треков : 'F1o', 'F2o', 'F1x', 'Ex', 'F2x', 'Eo'.

Распределение классов: F2o: 33.5%/F2x: 33.2%/Eo: 18.7%/F1o: 7.5%/F1x: 6.2%/Ex: 0.9%

Имеется явный дисбаланс некоторых классов. Но было принято решение, оставить текущее распределение.

```
{"image_name": "460290",  
  "image_width": 480,  
  "image_height": 497,  
  "tracks": [  
    {"id": 0,  
      "bbox": [0.05, 0.76, 0.10, 0.47],  
      "class": "F2o"},  
    {"id": 1,  
      "bbox": [0.10, 0.76, 0.18, 0.50],  
      "class": "F2x"}]}
```



3.2 CNN Region-based Transformers МОДЕЛЬ

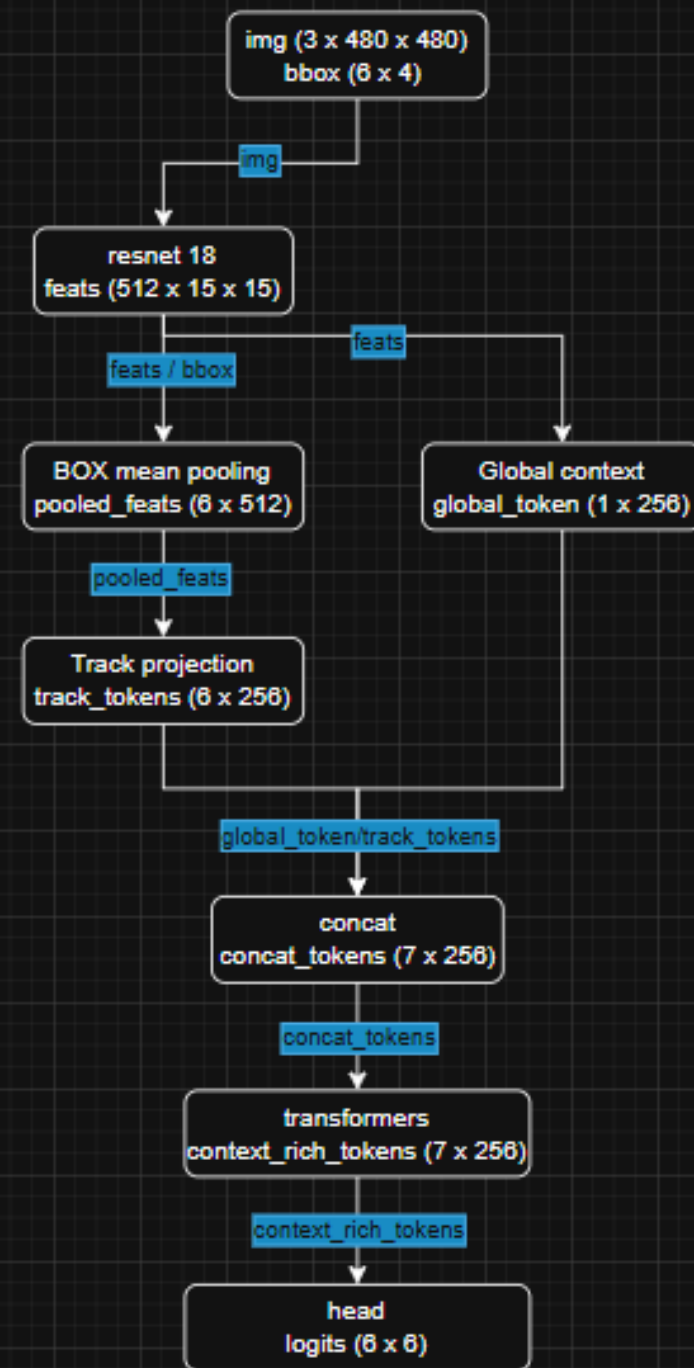
На вход модель получает изображение и набор bbox, которые в дальнейшем необходимо проклассифицировать. На выходе, для каждого bbox, модель выдаёт распределение вероятностей на каждый из существующих классов.

Модель является гибридом архитектуры CNN и Transformers. Архитектуру можно классифицировать как CNN Region-based Transformers.

Основное преимущество модели, при классификации она учитывает взаимную встречаемость треков друг с другом.

Так же модель косвенно через признаки, учитывает взаимное расположение треков, но в дальнейшем, планируется добавить эмбединги, которые будет добавлять в исходные координатную и физическую информацию о каждом треке. Это требует доработки датасета.

1. Вход: изображение (3 x 480 x 480) + 6 боксов (6 x 4, недостающие - нули).
2. ResNet-18 (предобученная): вход 480 x 480. Выход карта признаков (512x15x15).
3. Box mean pooling:
 - Масштабирование боксов с 480 до 15.
 - Маска (15 x 15) для каждого из 6 боксов.
 - Усреднение признаков по маске.
 - Выход: (6 x 512).
4. Track projections: линейный слой + нормализация. Выход : (6 x 256).
5. Global context: из карты (512 x 15 x 15). Выход : глобальный эмбединг (1 x 256).
6. Concat: (6 x 256) + (1 x 256). Выход : (7 x 256).
7. Transformer (энкодер): (7 x 256). Выход : контекстуальные эмбединги (7 x 256).
8. Head: классификация каждого из 7 токенов. Выход : логиты по классам для каждого бокса.



3.3 РЕЗУЛЬТАТЫ МОДЕЛИ

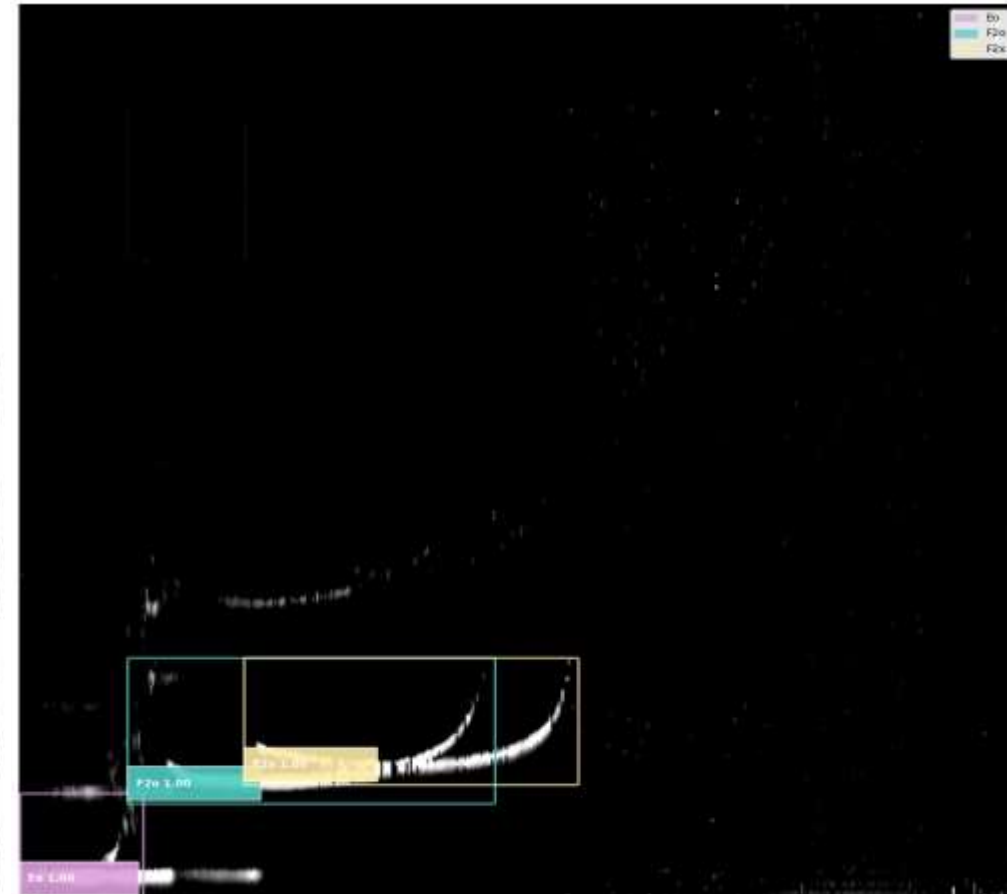
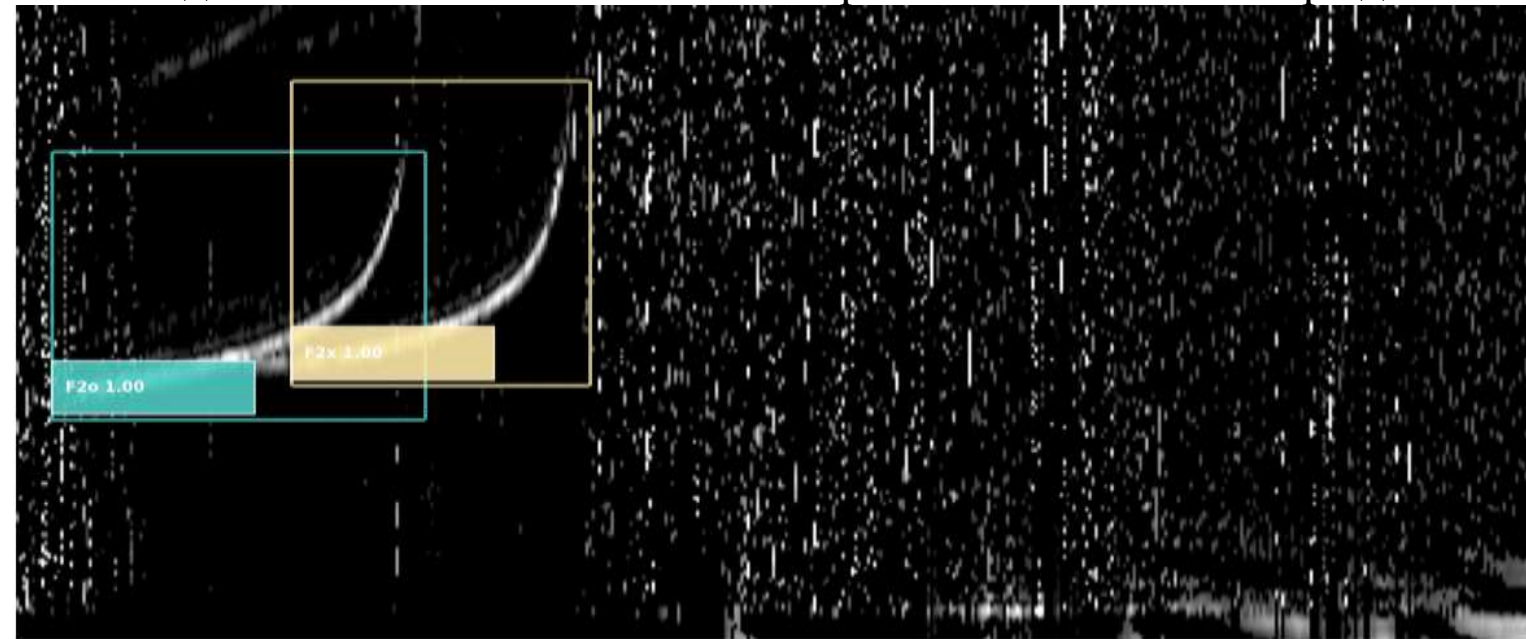
Модель обучалась на данных разбитых следующим образом : train/val/test - 1725/370/370 - 70/15/15.

Странное наблюдение, что модель уже на начальных эпохах имеет достаточно высокие метрики (на 3 эпохе val ассурасу составлял 0.9733, а пиковая точность была достигнута на 17 эпохе - val ассурасу 0.9843). Но скорее всего это происходит из-за схожести и простоты образов в наборе данных, а так же из-за большого количества схожих изображений. Набор данных не отсматривался вручную. За основу была взята разметка из набора данных измиран.

На тестовом наборе данных получились следующие метрики :

Accuracy	Precision	Recall	F1-score
0.9865	0.96	0.96	0.96

Модель показала очень высокие метрики на тестовом наборе данных.



4 ЗАКЛЮЧЕНИЕ

Новый набор данных и новая модель YOLO26-m сегментации, показали себя наилучшим образом. Для задачи выделения треков (1 класс), метрики составили : $\text{map}_{50}(\text{mask}) - 0.684$, $\text{map}_{50}(\text{box}) - 0.751$, что является очень высоким показателем, достаточным для дальнейшего использования результата модели. В дальнейшем можно ещё расширить набор данных, если будет необходима более точное выделение масок треков (на данный момент $\text{map} - 0.272$).

Эксперимент с данными ИЗМИРАН и кастомной архитектурой CNN-Region based Transformers, можно считать успешным. Было выявлено, что архитектура способна уловить взаимное расположение треков, и достаточно точно предсказывает класс трека внутри bbox. В дальнейшем необходимо очистить текущий набор данных от похожих изображений, дообучить модель на вертикальных ионограммах, и расширить набор классов, для того чтобы модель могла работать с наклонными ионограммами. Так же, можно добавлять чёткие координатные признаки треков в эмбединги, т.к. текущая реализация, косвенно учитывает координатное расположение треков друг относительно друга.

После доработки модели классификации, можно объединить эти две модели в одну систему, которая будет решать задачу сегментации и классификации треков.