

Методы автоматической нумерации позвонков на КТ-изображениях

Научный трек иннпрак МФТИ(2023)

23.05.2023
Ли Давид
Б05-912

План презентации

- Напоминание
- Изменение второй модели
- Новые пост процессинги
- АВ тесты
- Что дальше?

Напоминание

Постановка задачи(Напоминание)

- Глобальная задача:
 - Детекция позвонков на КТ-изображениях
 - Straight-forward слишком дорого из-за high-resolution картинок
 - Решение: Interpretable vertebral fractures(A.Zakharov et al.)
- Локальная задача:
 - Автоматический майнинг нетривиальных кейсов
 - Изменение классификационного подхода позвонков

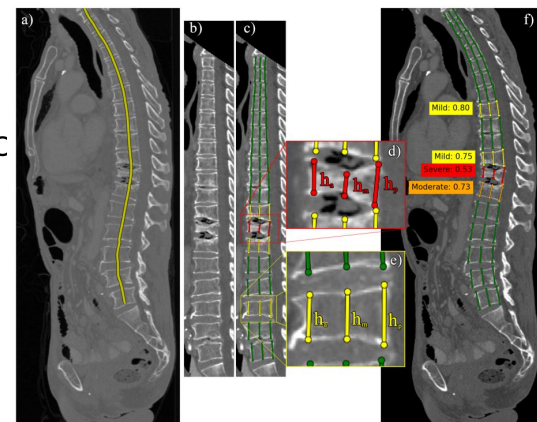


Figure 1: fix inconsistent colors Overview of the proposed model. **Step 1:** a) localizing vertebrae centers in 3D CT (a sagittal projection is shown); b) generating a new 2D image via spine 'straightening'. **Step 2:** c) identifying key-points and the corresponding heights; d-e) a closer look at some vertebrae (colors denote the fracture severity). **Finally:** f) the original image with vertebrae types (right side) and their estimated fracture severities (left side).

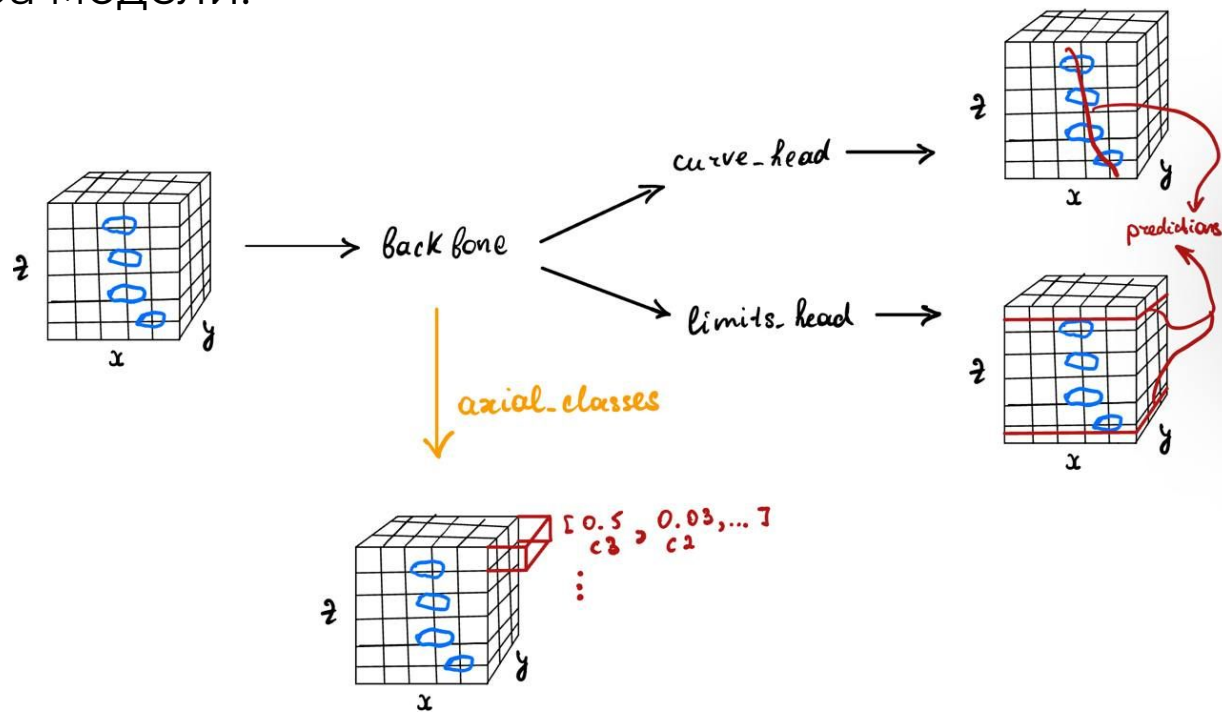
Майнинг метрики(Напоминание)

Придумали несколько метрик для предсказания модели чтобы отловить плохие предсказания. Вот наиболее интересные из них:

- Длина позвоночника / кол-во позвонков
- Количество позвонков
- Максимальное расстояние между позвонками
- Максимальная кривизна кривой позвоночника(по короналу и сагитталу)
- Максимальное изменение кривизны кривой позвоночника(по короналу и сагитталу)
- Максимальная вторая производная(по короналу и сагитталу)

Архитектура модели(Напоминание)

- Пост процессинг: MLE
- Архитектура модели:



Новые данные(Напоминание)

- Датасет VerSE

	max	rmse	accuracy	precision	recall	limits_mae	top_limits_mae	bottom_limits_mae	accuracy_verts	accuracy_verts_mle
merged	4.38	2.27	0.99	1.00	0.99	2.98	2.49	0.49	0.89	0.97
verse	3.50	1.81	0.93	0.99	0.92	19.14	12.71	6.43	0.70	0.96

- Новые данные

	max	rmse	accuracy	precision	recall	limits_mae	top_limits_mae	bottom_limits_mae	accuracy_verts	accuracy_verts_mle
index										
merged	4.19	2.15	0.99	1.0	0.99	2.74	2.20	0.54	0.90	0.96
verse	3.38	1.78	0.94	1.0	0.93	26.37	15.61	10.76	0.72	0.95
new_cases	8.80	3.41	0.99	1.0	0.99	1.97	1.25	0.72	0.72	0.97

Вторая модель

Архитектура модели

Было

- Многоклассовый выход
- В качестве пост процессинга: mle

Стало

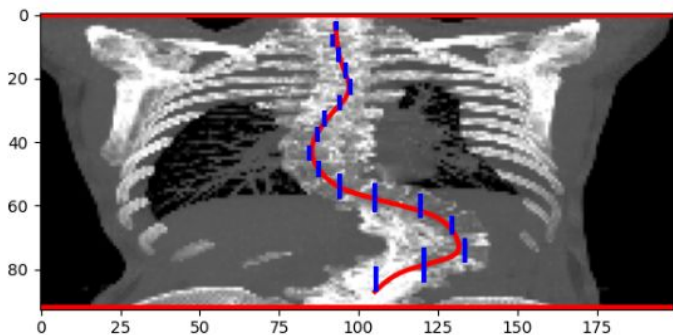
- Одноклассовый выход
- В качестве пост процессинга: mle
- Добавили еще S позвонки

Предположение:

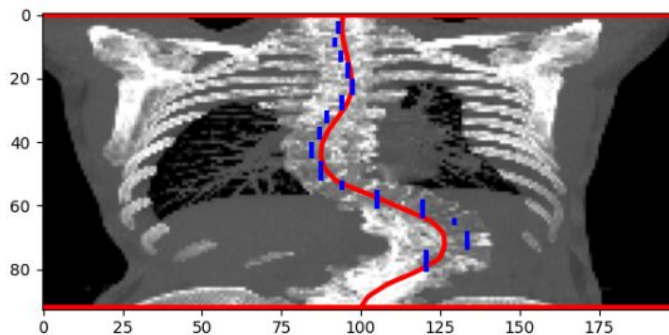
Модель станет лучше работать, т.к. мы сняли с нее довольно сложную задачу в виде классификации. Ей осталось только точно находить точки.

Обработка данных

Были некоторые кейсы с слишком сложными случаями. Иногда даже врачи правильно разметить точки, т.к. эти позвонки были почти горизонтальными. Мы назвали такие позвонки “incorrect” и выкинули их из обучения.



label



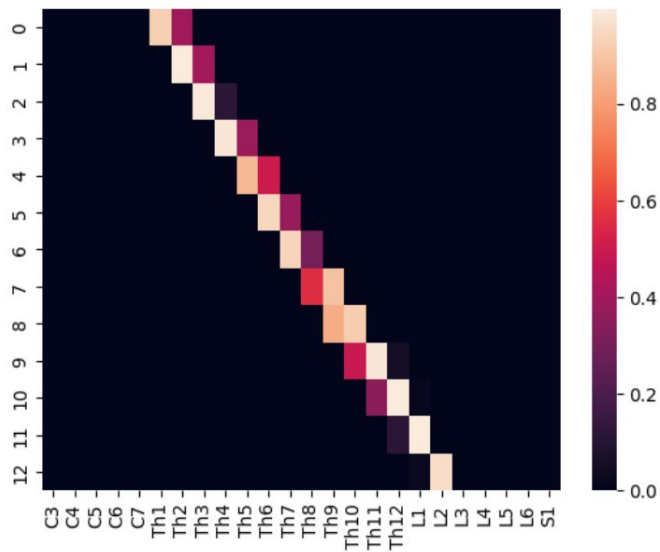
predict

Результаты

	reference	multiclass_mle	binary_mle
auc_0v123	0.942337	0.936485	0.936366
auc_01v23	0.980604	0.973954	0.976978
auc_012v3	0.997067	0.996186	0.993502
types_mae	0.067668	0.044596	0.041942
genant_max	0.028419	0.032533	0.032222
genant_mae	0.028419	0.032533	0.032222
width_mae	0.961009	1.065494	1.054836
width_rmse	1.074773	1.199070	1.187308
points_mae	1.457838	1.434878	1.432495
detection_precision	0.986370	0.981236	0.982638
detection_recall	0.993715	0.990799	0.989242
points_rmse	1.983878	1.909136	1.906749
types_accuracy	0.945917	0.958845	0.960314

Есть проблемы

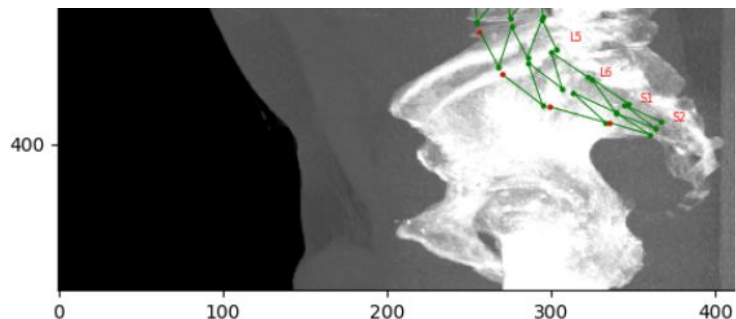
Мы заметили что иногда модель может быть уверена в какой-то последовательности и скорее всего это правильная последовательность, но из-за неуверенности в других позвонках модель выдает неправильную последовательность.



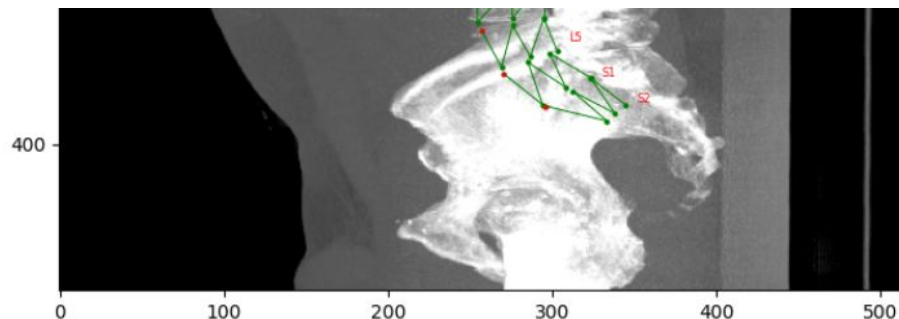
Один позвонок ломает всю последовательность

Есть проблемы

Иногда модель предсказывает дополнительные, не нужные нам позвонки.



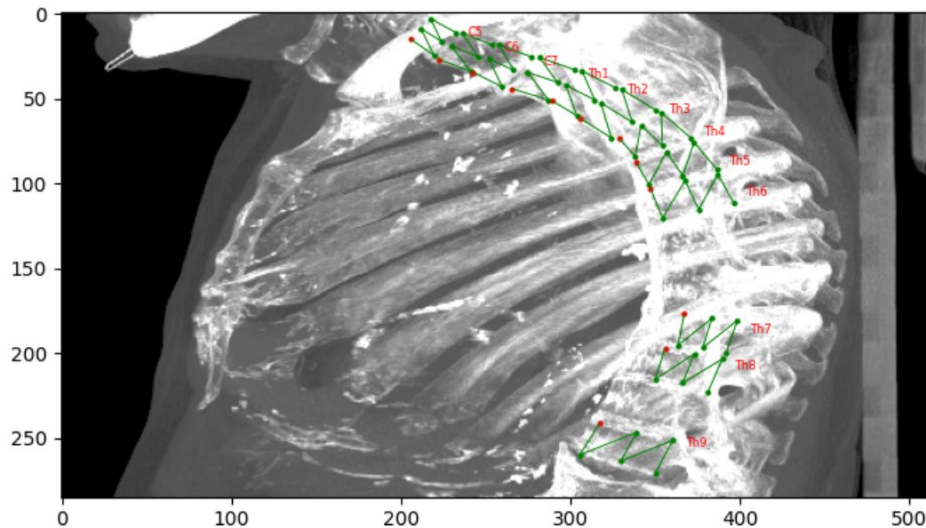
predict



label

Есть проблемы

Предполагаемые последовательности слишком зафиксированы. Что будет если модель пропустит несколько позвонков?

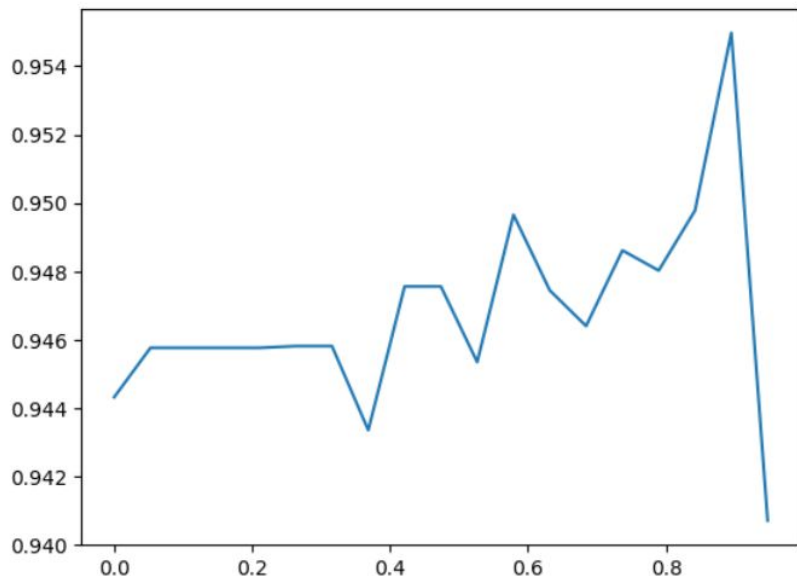


Модель не поняла, что пропустила позвонки и дала им соответствующие номера.

Пост процессинги

Possible sequences

Первую проблему можно решить учитывая только наиболее уверенные позвонки. Если число слишком маленькое (< 5), то передаем процессинг в mle, т.е. метод некоторое обобщение mle



Зависимость метрики от threshold (простите за ужасный график, все что нашел в архиве:())

FYI:

Мы взяли $\text{thr} = 0.8$

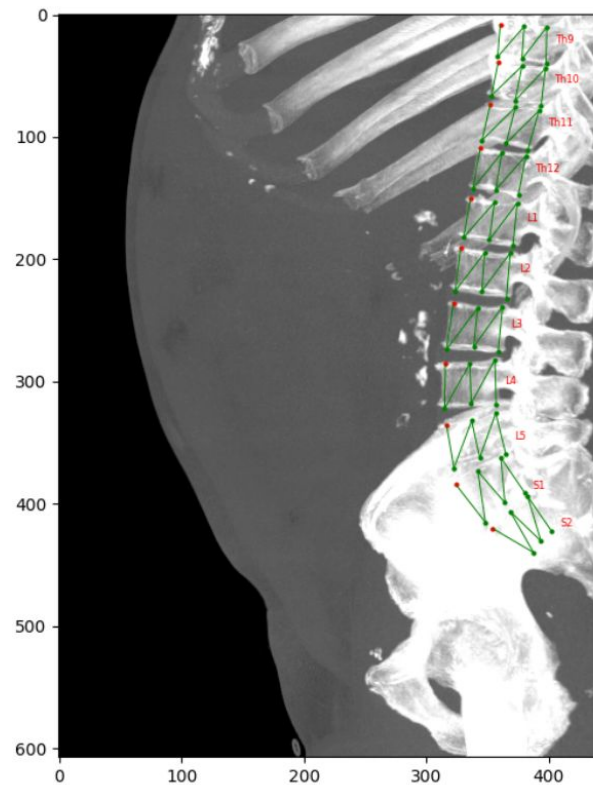
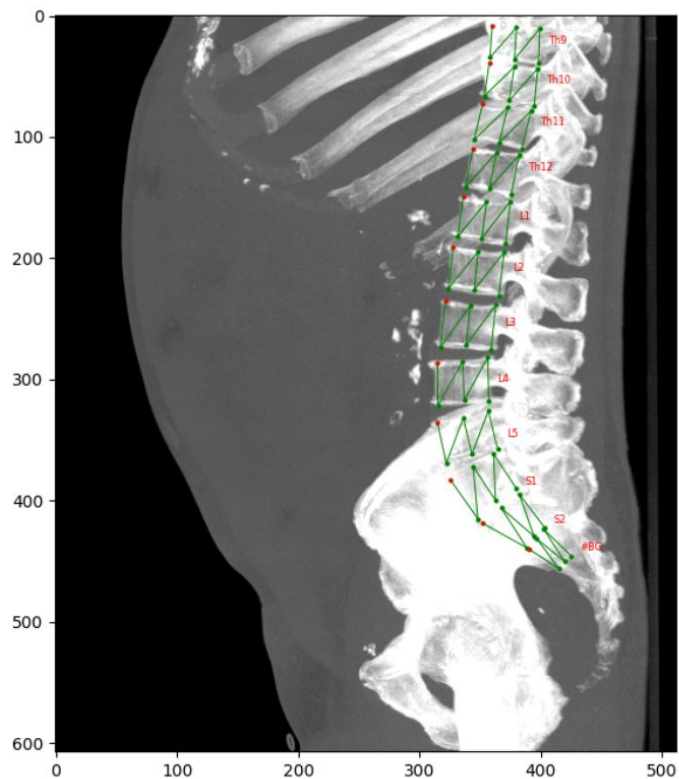
Possible sequences expand edges

Заметим что вторую проблему можно решить этим пост процессингом, ведь мы используем только позвонки с наибольшей вероятностью. Для этого введем класс #BC

Результаты

	reference	multiclass_mle	multiclass_possible_sequence	binary_mle	binary_possible_sequence	binary_possible_sequence_expand_edges
auc_0v123	0.942337	0.936485	0.936438	0.936366	0.936366	0.936366
auc_01v23	0.980604	0.973954	0.973943	0.976978	0.976969	0.976969
auc_012v3	0.997067	0.996186	0.996203	0.993502	0.993499	0.993499
types_mae	0.067668	0.044596	0.040002	0.041942	0.037180	0.036051
genant_max	0.028419	0.032533	0.032535	0.032222	0.032224	0.032224
genant_mae	0.028419	0.032533	0.032535	0.032222	0.032224	0.032224
width_mae	0.961009	1.065494	1.065085	1.054836	1.054811	1.054811
width_rmse	1.074773	1.199070	1.198653	1.187308	1.187282	1.187282
points_mae	1.457838	1.434878	1.434804	1.432495	1.432491	1.432491
detection_precision	0.986370	0.981236	0.981161	0.982638	0.982638	0.982638
detection_recall	0.993715	0.990799	0.990725	0.989242	0.989242	0.989242
points_rmse	1.983878	1.909136	1.908999	1.906749	1.906748	1.906748
types_accuracy	0.945917	0.958845	0.963775	0.960314	0.965574	0.966702

Результаты



Skipping post processing

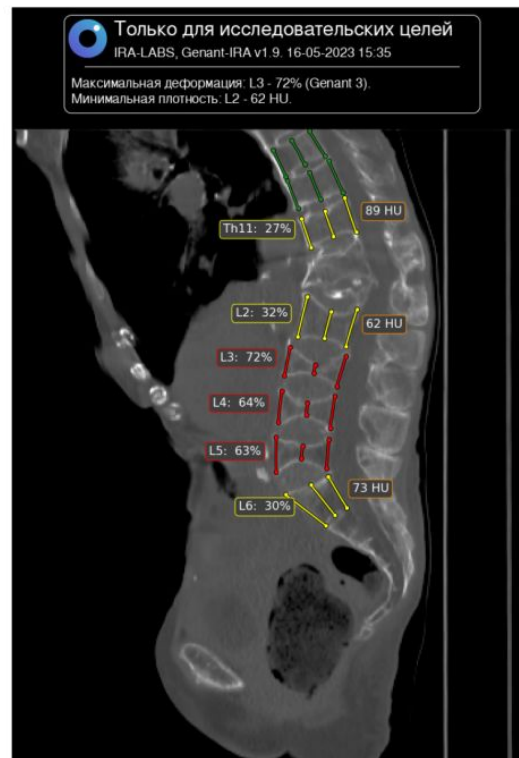
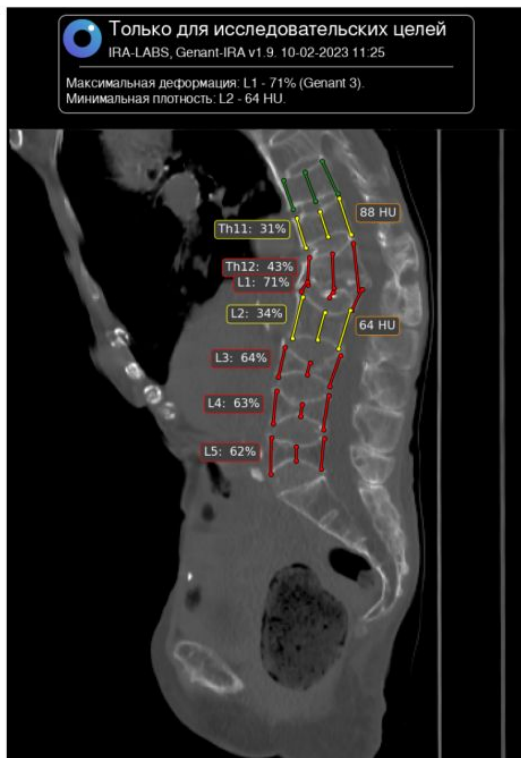
Эту проблему можем решить с помощью расстояний между позвонками. Вопрос в том как именно понять сколько позвонков мы пропустили. Для этого мы найдем позвонки между которыми аномальные расстояния(выше изначального трешхолда) и переберем все возможные классы позвонков для этих позвонков.

Другими словами разобьем цепь позвонков на компоненты связности в зависимости от расстояний между ними и для них переберем всевозможные номера.

Результаты

	reference	multiclass_mle	multiclass_possible_sequence	binary_mle	binary_possible_sequence	binary_possible_sequence_expand_edges	binary_skipping_seq
auc_0v123	0.942337	0.936485	0.936438	0.936366	0.936366	0.936366	0.936366
auc_01v23	0.980604	0.973954	0.973943	0.976978	0.976969	0.976969	0.976969
auc_012v3	0.997067	0.996186	0.996203	0.993502	0.993499	0.993499	0.993499
types_mae	0.067668	0.044596	0.040002	0.041942	0.037180	0.036051	0.033566
genant_max	0.028419	0.032533	0.032535	0.032222	0.032224	0.032224	0.032224
genant_mae	0.028419	0.032533	0.032535	0.032222	0.032224	0.032224	0.032224
width_mae	0.961009	1.065494	1.065085	1.054836	1.054811	1.054811	1.054811
width_rmse	1.074773	1.199070	1.198653	1.187308	1.187282	1.187282	1.187282
points_mae	1.457838	1.434878	1.434804	1.432495	1.432491	1.432491	1.432491
detection_precision	0.986370	0.981236	0.981161	0.982638	0.982638	0.982638	0.982638
detection_recall	0.993715	0.990799	0.990725	0.989242	0.989242	0.989242	0.989242
points_rmse	1.983878	1.909136	1.908999	1.906749	1.906748	1.906748	1.906748
types_accuracy	0.945917	0.958845	0.963775	0.960314	0.965574	0.966702	0.967268

Результаты



Union post processing + итоговые результаты

Объединим все наши методы.

	reference	old_model	new_model
auc_0v123	0.942337	0.936438	0.939606
auc_01v23	0.980604	0.973943	0.979654
auc_012v3	0.997067	0.996203	0.995995
types_mae	0.067668	0.049092	0.027978
genant_max	0.028419	0.032535	0.031689
genant_mae	0.028419	0.032535	0.031689
width_mae	0.961009	1.065085	1.043078
width_rmse	1.074773	1.198653	1.173439
points_mae	1.457838	1.434804	1.414823
detection_precision	0.986370	0.981161	0.981808
detection_recall	0.993715	0.990725	0.990190
points_rmse	1.983878	1.908999	1.885366
types_accuracy	0.945917	0.954425	0.973224

АВ тесты

Проведем АВ тестирование

Возьмем много кейсов. И запустим на них наши модели(старую и новую). Из них отберем “интересные” кейсы которые отправим врачам. Они скажут какая из моделей справилась лучше или же не справились обе или же обе справились хорошо.

“Интересные кейсы”

Возьмем 4000 кейсов и отберем их по следующим метрикам:

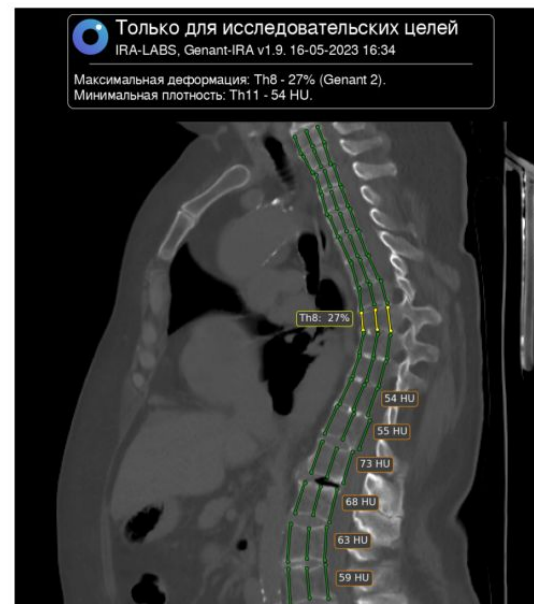
- разное количество позвонков с Генант 3 (деформация $> 40\%$)
- позвонки с различающимися Генант 3 и разницей больше какого-то трешхолда
- разное количество найденных позвонков
- отличие нумерации
- сильно расхождение в кривой позвоночника по метрике **max**

Примеры

A

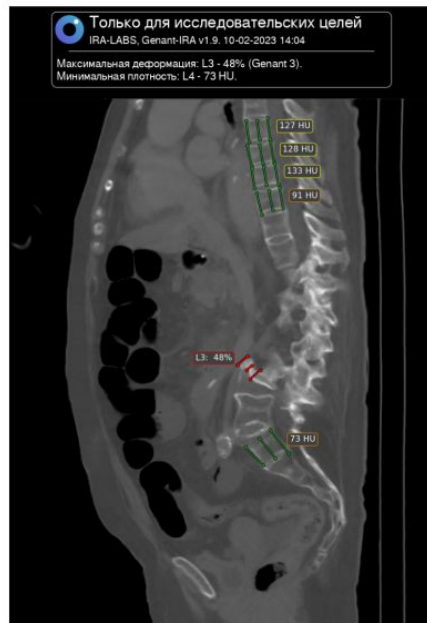


B

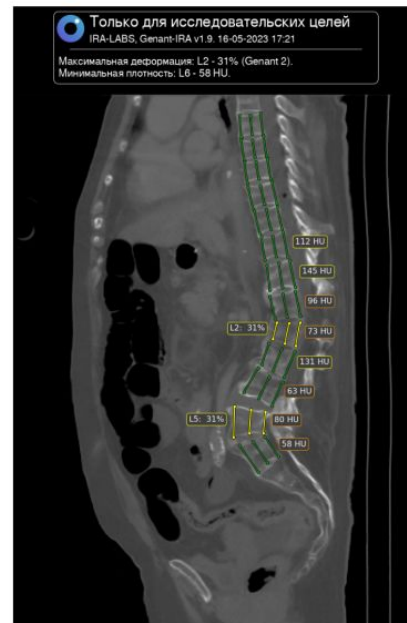


Примеры

А

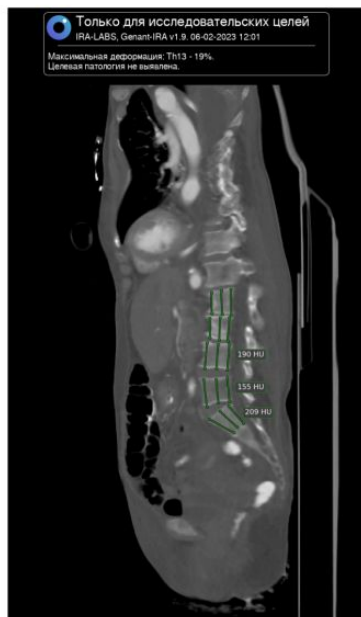


В

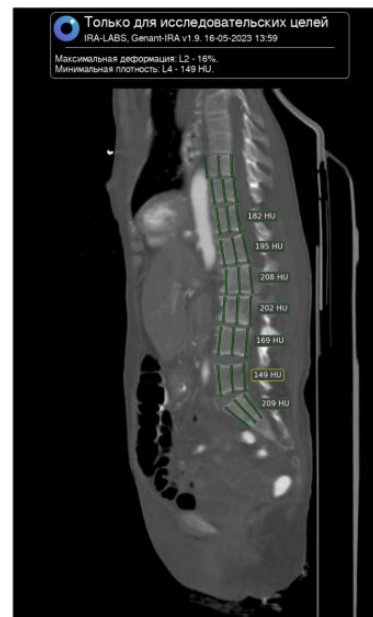


Примеры

А



В



Примеры

А

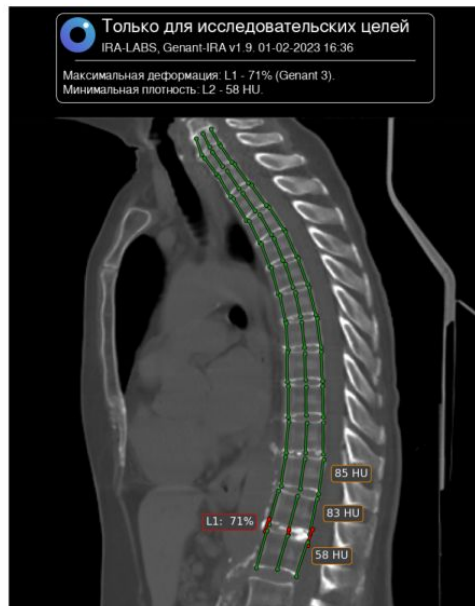


В

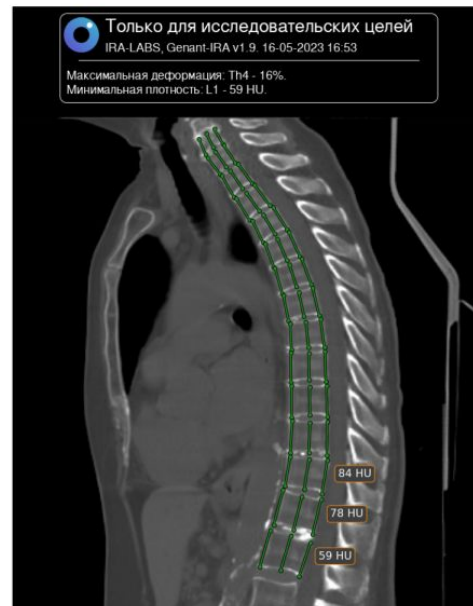


Примеры

A



B



Примеры

A



B



Что дальше?

Что будет с развитием модели?

Сейчас врачи размечают выборку для АВ теста. Если эффект будет стат.значимый то мы зарелизим эту модель и будем разрабатывать новую версию. Релиз модели означает что во всех поликлиниках с которыми мы сотрудничаем будет использоваться новая модель. Сейчас в Москве нашими алгоритмами обрабатывается > 50% КТ-снимков. Скорее всего если вы будете делать снимок позвоночника, то как раз наша модель его обработает:)

Что будет с развитием модели?

Дальнейшее развитие модели предполагается с помощью реализации end-to-end модели. Ожидается что она:

- Будет более устойчивее работать
- Покажет результат получше

Важно понимать что у нас есть возможность сделать этот шаг потому что **появились достаточные вычислительные мощности.**

THANKS FOR YOUR ATTENTION

NO QUESTIONS PLEASE

makeameme.org