

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
"ВЫСШАЯ ШКОЛА ЭКОНОМИКИ"  
ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК**

Жуков Федор Викторович

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

**Детекция переломов ребер на снимках компьютерной томографии с помощью  
методов глубокого обучения**

**Detection of Rib Fractures on Computed Tomography Images Using Deep Learning  
Methods**

по направлению подготовки 01.04.02 Прикладная математика и информатика  
образовательная программа «Науки о данных»

Студент

Ф.В. Жуков

Научный руководитель

Доцент

Е.И. Ершов



Соруководитель

Кандидат компьютерных наук, PhD

А.И. Курмуков

Москва 2022

## Аннотация

Перелом ребер важная задача для определения степени тяжести травмы грудной клетки. Для того чтобы качественно определить повреждение нужны опытные радиологи, более того качество также зависит от степени квалификации доктора и количества изображений компьютерной томографии (КТ), которые необходимо обработать. Многие исследования посвящённые этой проблеме так или иначе ставят перед собой цель создать алгоритм, который будет лучше справляться с детекцией и сегментацией переломов ребер. В данном исследовании я изучаю модель глубокого обучения под названием FracNet, разработанную для задач детекции и сегментации переломов ребер на снимках компьютерной томографии, чтобы воссоздать результаты авторов статьи [1]. В датасет RibFrac [17] было включено 7473 аннотированных травматических переломов ребер 900 пациентов, маркировка которых выполнена человеком. Я принял участие в соревновании RibFrac 2020 [18] по анализу переломов ребер. В этом соревновании я занял 342-е место. Результаты у меня получились следующими: Dice = 0.12, IoU = 0.06 (пересечение по объединению), средняя полнота равнялась 0.07, а максимальная полнота равнялась 0.22.

## Annotation

Rib fractures are an important task in determining the severity of a thoracic injury. In order to qualitatively determine the injury, experienced radiologists are needed, moreover, the quality also depends on the doctor's skill and the number of computed tomography (CT) images that need to be processed. Many studies on this problem in one way or another aim to create an algorithm that is better able to detect and segment rib fractures. In this research, I study a deep learning model called FracNet, developed for the task of rib fracture detection and segmentation on CT images, to recreate the results of the authors of [1]. The RibFrac dataset [17] included 7,473 annotated traumatic rib fractures from 900 human-labeled patients. I entered the RibFrac 2020 [18] rib fracture analysis competition. I ranked 342nd in this competition. My results were as follows: Dice = 0.12, IoU = 0.06 (Intersection over union), average recall was 0.07, and maximum recall was 0.22.

Научный руководитель

Курищев А.И.

В.В. Сергеев

## Содержание

Введение	3
Обзор литературы	5
Датасет	5
Модель	8
Архитектура	10
Обучение	10
Метрики	11
Результаты	12
Вывод	14
Литература	15

## Введение

Недавние исследования в области компьютерного зрения и искусственного интеллекта приводят к успехам в анализе медицинских изображений и усовершенствованию технологий глубокого обучения [2, 3, 4]. Глубокое обучение эффективно в задачах 3D и 2D медицинских изображений таких как сегментация [5], детекция [6], классификация [7]. Все исследования в этой области направлены на улучшение результатов, которые уже имеются. Удачным применением технологий глубокого обучения, считается повышение высокопроизводительности алгоритма, точности и уверенности в диагнозе. Все это делается для того, чтобы облегчить труд врача, сделать лучше лечение пациентов и ускорить все процессы непосредственно связанные с работой человека [8].

Количество переломов ребер говорит о тяжести травмы грудной клетки. Такое повреждение для человека представляет разную опасность для здоровья в зависимости от типа перелома. Компьютерная томография предлагает точную оценку наличия переломов ребер, в то время как стандартная рентгенография грудной клетки специфична, но не чувствительна [9, 10, 11]. Диагностика количества переломов ребер также является важным показателем в судебно-медицинской экспертизе для определения степени инвалидности [12].

Детекция переломов на КТ-изображениях является сложной задачей, если применять обычные аксиальные тонкие снимки. Каждое ребро уникально и имеет свою форму, в основном с диагональным ходом. Перелом в местах сгиба самый отсутствующий тип перелома в медицинской практике из-за сложного внешнего вида. Перелом без смещения может быть не обнаружен, если плоскость сканирования компьютерной томографии параллельна ему, как это видно на рисунке 1. Сама диагностика переломов ребер трудоемкая и утомительная задача, которая требует кропотливого изучения изображения в разных плоскостях снимок за снимком.

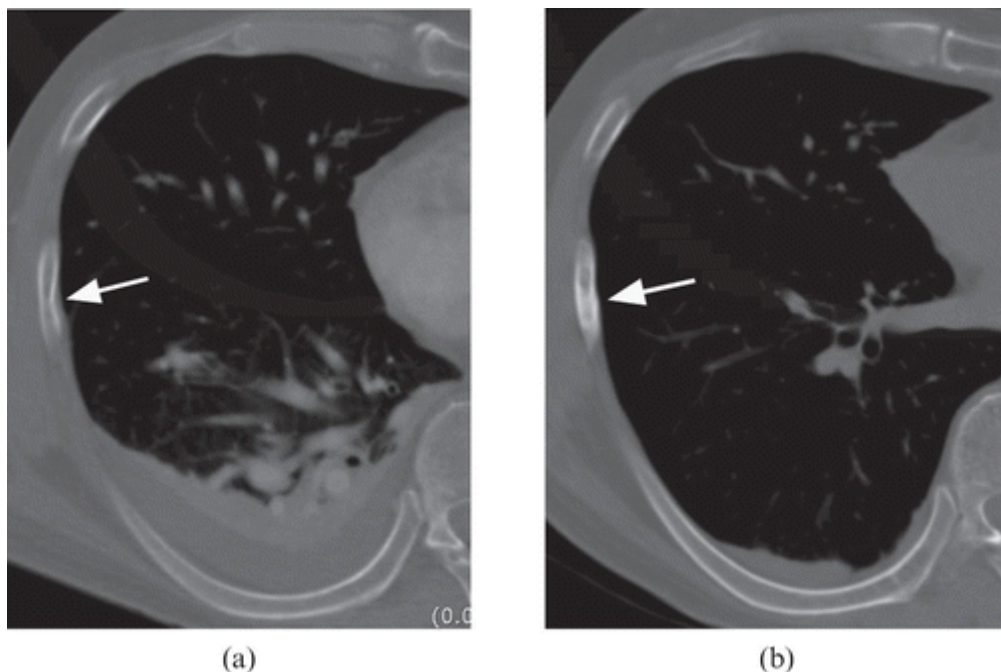


Рисунок 1. (a) Перелом правого седьмого ребра (стрелка) пропущен на первичной компьютерной томографии. Он расположен на боковой дуге и имеет форму пряжки. (b) На повторном КТ-снимке через 2 дня видно образование мозоли (стрелка) на том же месте. (Иллюстрации взяты с [20]).

В моей работе я повторяю результаты авторов статьи [1], используя автоматическую систему FracNet для детекции и сегментации переломов ребер на КТ-сканах.

## Обзор литературы

В качестве основы для моей работы я использовал статью работу [1]. В данном исследовании авторы создали автоматическую систему, которая сегментирует и детектирует переломы ребер на снимках компьютерной томографии.

Существует не так много Мало исследований посвященных алгоритмам глубокого обучения в данной области.

Авторы разработали и предложили систему FracNet на основе 3D UNet [13], обученную на огромном датасете RibFrac, состоящем из 7 473 сегментаций переломов ребер на уровне вокселей, из 900 снимков компьютерной томографии грудной клетки. Маркировка данных происходила при участии человека, что обеспечивала высокое качество аннотаций. На тестовом наборе датасета разработанная система FracNet достигла чувствительности обнаружения 92,9% (при 5,27 ложных срабатываний на исследование) и коэффициента сегментации Dice 71,5%, что в разы превосходит другие модели, в основе которых 3D-варианты FCN [14] и DeepLab v3+ [15], ResNet-18 [16]. Для выявления медицинской ценности разработанной системы были проведены наблюдательные исследования с участием двух опытных радиологов. Интересно, что совместная работа человека и компьютера значительно повысила чувствительность детекции по сравнению с диагностикой только с помощью компьютера и только с помощью человека, а также сократила клиническое время по сравнению с диагностикой только с помощью человека. Модель предложенная авторами достигла более высокой чувствительности обнаружения, чем эксперты-люди.

## Датасет

Набор данных состоит из исследований собранных с января 2017 года по декабрь 2018 путем поиска в электронных медицинских картах и радиологических информационных системах больниц пациентов с травматическими переломами ребер, обнаруженных на снимках компьютерной томографии грудной клетки и брюшной полости. Всего в набор данных включили 7473 травматических перелома ребер 900 пациентов, средний возраст которых 55 лет. Возраст пациентов в основном был в диапазоне 21-94 года. Среди них было 580 мужчин и 329 женщин. Датасет можно найти на сайте [17].

Как показано на рисунке 2 (а), условия включения следующие: (1) пациенты с тонкослойными КТ-изображениями травмы грудной клетки (1–1,25 мм), содержащими все ребра, и (2) тонкослойные КТ-изображения без респираторных артефактов, снижающих точность диагностики.

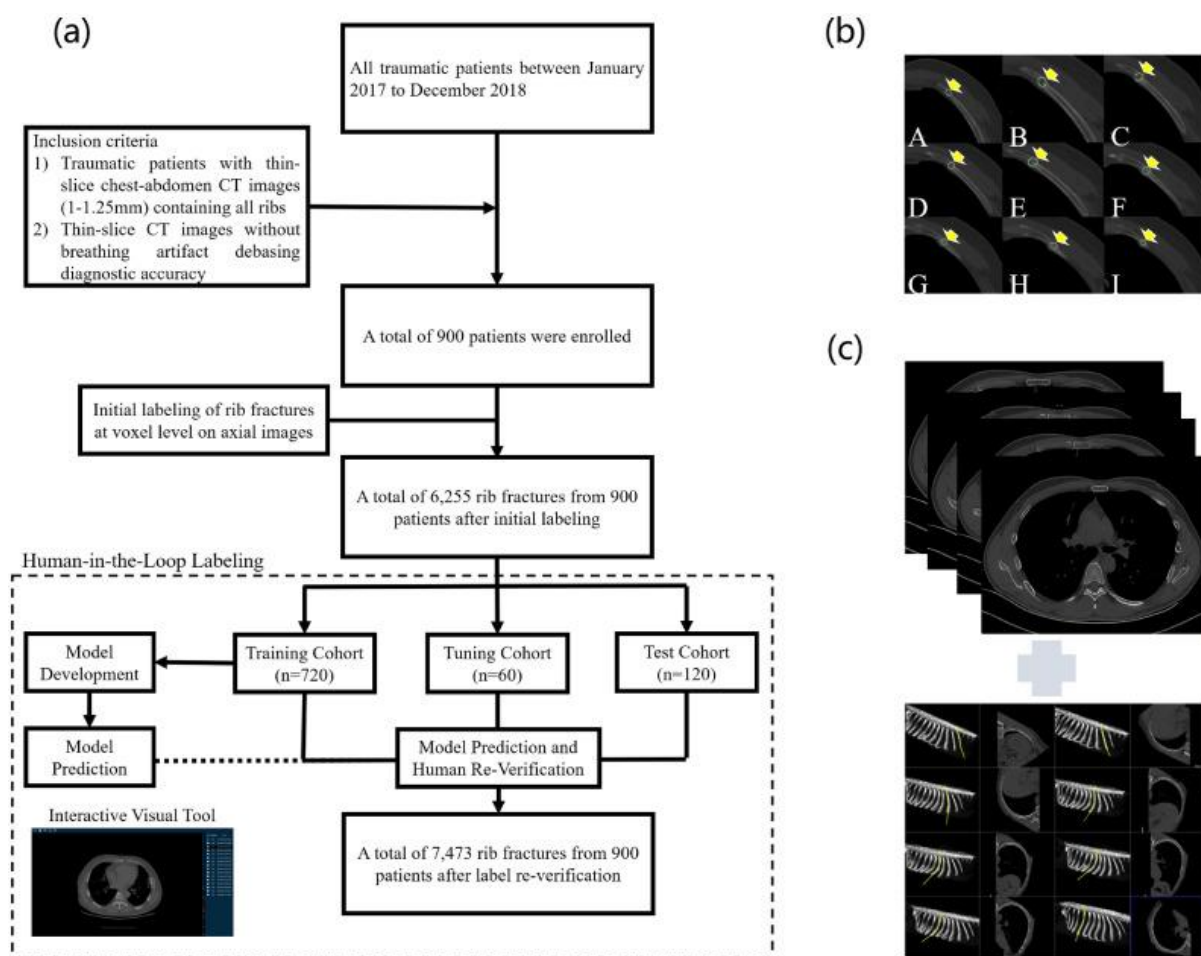


Рисунок 2. (a) Блок-схема набора данных для датасета RibFrac, включая маркировку переломов ребер с помощью человека. (b) Иллюстрация ручной маркировки переломов ребер. (c) Верификация ручной маркировки с помощью осевых изображений (сверху) и ручное искривление изображений плоскостного реформирования (внизу). (Иллюстрации взяты с [1]).

Разметкой данных участвовали 5 опытных рентгенологов (5-20 лет опыта). Обломки поврежденной кости были включены в объем переломов, как показано на рисунке 2 (b). Так как повреждения могут быть незаметными из-за того, что линия перелома отсутствует или она параллельна плоскости, в которой происходит детекция, на рисунке 2 (c) для обеспечения более высокой точности маркировки применялись осевые изображения совместно с изображениями плоскостной реформации, выполненными вручную.

Первоначальная модель глубокого обучения, разработанная по тем же принципам, что и FracNet, была основана на обучающей части датасета RibFrac. Первоначальная система использовалась для предсказания переломов на обучающем, валидационном и тестовом наборах RibFrac.

Как можно видеть на табл. 1 датасет состоит из 3 наборов: обучение (720 штук), валидация (60 штук) и тест (120 штук).

Когорты	Доступность	Количество КТ сканов	Количество Кт слайсов	Количество переломов
Обучение	Всего	720	265,302	6,156
	Публичный	420	154,127	3,987
	Внутренний	300	111,175	2,169
Валидация	Публичный	60	22,562	435
Тест	Публичный	120	44,619	882

Таблица 1. Набор данных RibFrac.



## Модель

Решение задачи состоит из 3 шагов (рисунок 3): предварительная обработка, предсказание по скользящему окну и постобработка.

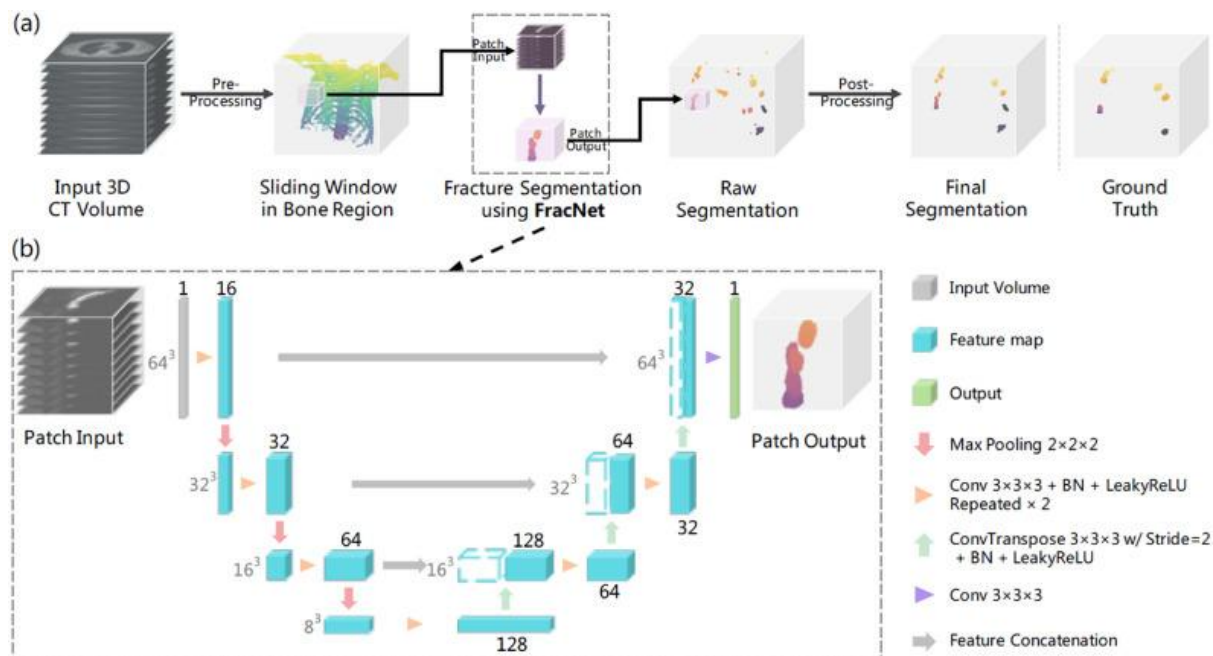


Рисунок 3. (а) Схема детекции переломов ребер по снимкам КТ. Была разработана трехмерная сверточная нейронная сеть под названием FracNet для сегментации переломов в режиме скользящего окна. Псевдоцвет на рисунке используется для лучшей визуализации бинарных изображений костей и результатов сегментации. (б) Архитектура нейронной сети FracNet основывается на 3D UNet [13]. (Иллюстрации взяты с [1]).

Предварительная обработка вокселей состоит из фильтрации и пороговая обработка. Это необходимо для более быстрой и качественной детекции. Изначальное расположение костей относительно друг друга не было изменено, потому что в датасет включены только тонкослойные компьютерные томограммы. Насыщенность входных вокселей была сокращена до скользящего окна и нормализована в диапазоне  $[-1, 1]$ .

Следующий этап это предсказание по скользящему окну. Стандартная маркировка с помощью ограничительной рамки может быть неэффективна, потому что ребра имеют вытянутую форму. При такой ситуации ограничительные рамки могут не содержать много деталей. Это задача решалась в виде трехмерной сегментации и для выполнения этой задачи была разработана 3D-сеть FracNet на основе Unet [13]. В режиме скользящего окна с шагом 48 я подавал сканы компьютерной томографии в сеть, обрезанные до размера 32X32X32 для более быстрого обучения. Далее, среди всех патчей предсказаний выделялись области перекрытия нескольких предсказаний и в

них сохранялись максимальные значения. Предполагаемое место сегментации получалось путем сборки мест с предсказаниями. Далее, максимальное значение сохранялось в области, в которой пересекались несколько предсказаний.

Для эффективной постобработки и уменьшения количества ложно-положительных предсказаний малые прогнозы размера менее 200 вокселей были отфильтрованы. Были также отфильтрованы области с позвоночником. Результаты сегментации были бинаризованы с низким порогом равным 0,1 для того, чтобы создать предполагаемые места повреждений, далее сложили части, которые связаны на бинарной сегментации. Местам переломов были приписаны вероятности вычисляемые как среднее оценок необработанной сегментации по всем вокселям внутри предполагаемого места перелома.

## Архитектура

Модель FracNet основана на 3D UNet [13]. Архитектура состоит из энкодера-декодера (рисунок 3 (b)). Энкодер состоит из нескольких уровней понижающей дискретизации. Каждый уровень состоит из 3D-свертки, батч-нормализации, нелинейности и максимального пулинга. После каждого этапа понижающей дискретизации карта признаков уменьшалась в два раза, а количество каналов удваивалось. Декодер представляет из себя восстановление карты признаков путем серии транспонированных свёрток. Признаки из энкодера повторно использовались объединяясь с признаками одних и тех же уровней энкодера. После восстановления исходного размера карты признаков использовался слой свертки размером  $1 \times 1 \times 1$  для уменьшения выходного канала до 1. Выходной сигнал активированный сигмоидной функцией обозначал либо фон равным нулю, либо поражением равным единице.

## Обучение

Для того чтобы установить баланс между положительными и отрицательными экземплярами, положительные экземпляры  $32 \times 32 \times 32$  были вырезаны из области размером  $96 \times 96 \times 96$  в центре перелома ребра. Отрицательные экземпляры были получены из областей костей без переломов. Дисбаланс присутствовал в силу скудного набора аннотаций во всем датасете. Каждый батч состоял из 12 наблюдений. В качестве аугментаций использовались повороты на случайный трехмерный угол. Функция потерь представлена комбинацией Dice loss и Binary Cross-Entropy. Сеть обучалась с помощью оптимизатора Adam и скорость обучения изменялась в зависимости от эпохи. Она линейно увеличивалась с 0,00001 до 0,1 во время первой эпохи, а потом точно так же уменьшалась до 0,001 во время 100 эпох.

## Метрики

Всего может быть четыре результата работы модели. Первый, модель предсказала перелом и перелом действительно есть, такие результаты называются истинно-положительными (TP). Второй, модель предсказала перелом, а при этом перелома нет, такие результаты называются ложно-положительными (FP). Третий, модель не предсказала перелом, но перелом есть, такие результаты называются ложно-отрицательными (FN). Четвертый, модель не предсказала перелом и перелома действительно нет, такие результаты называются истинно-отрицательными (TN). На основе этих результатов строится одна из самых популярных метрик для сегментации в области анализа медицинских изображений Dice (формула 1). Также применялась метрика пересечение по объединению (IoU, формула 2).

$$Dice = \frac{2 \times TP}{TP + FP + FN + TN} \quad (1)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

Для детекции использовались метрики полноты (формула 3), точности (формула 4) и F1-score (формула 5), которая является средним гармоническим между полнотой и точностью.

$$Полнота = \frac{TP}{TP + FN} \quad (3)$$

$$Точность = \frac{TP}{TP + FP} \quad (4)$$

$$F1-score = \frac{Точность \times Полнота}{Точность + Полнота} \quad (5)$$

## Результаты

Изначальный датасет имеет размер 34 гигабайта, но Для своих экспериментов я использовал 30% данных в силу ограниченных вычислительных ресурсов. Тренировка модели происходила достаточно плавно как можно видеть на рисунке 4 (100 эпох) и рисунке 5 (195 эпох), на которых представлены кривые функции потерь обучения и валидации соответственно.

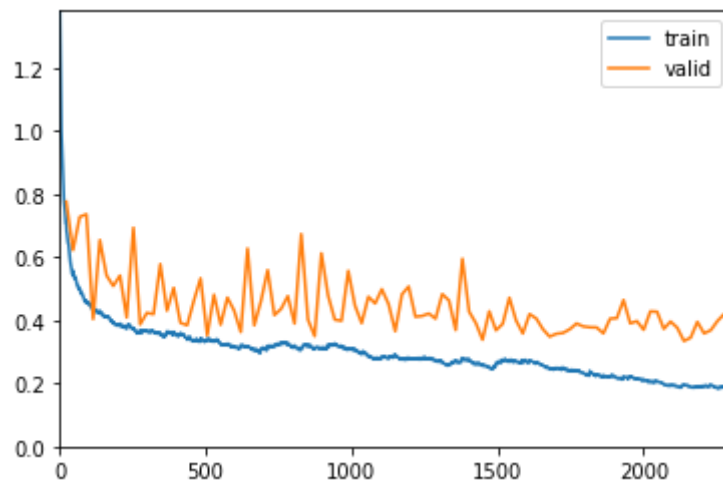


Рисунок 4. Функции потерь обучения (синий) и валидации (оранжевый). 100 эпох.

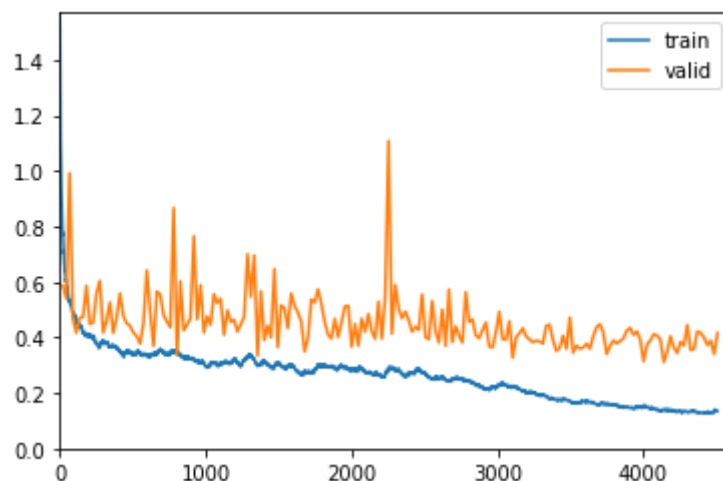


Рисунок 5. Функции потерь обучения (синий) и валидации (оранжевый). 195 эпох.

В целом переобучение не наблюдалось, значение на валидации всегда было выше чем значение на обучении. На валидации модель была сильно неустойчива, в то время как обучение проходило достаточно гладко.

Я принял участие в соревновании RibFrac 2020 [18] по анализу переломов ребер. В этом соревновании я занял 342-е место. Результаты у меня получились следующими: Dice = 0.12, IoU = 0.06 (пересечение по объединению), средняя полнота равнялась 0.07, а максимальная полнота равнялась 0.22. Эти результаты достаточно сильно отличаются от участников. Также я сделал предсказания для обучающей части датасета RibFrac, чтобы сравнить их с маркированными данными. В итоге среднее значение Dice получилось 0.13 со стандартным отклонением 0.034. Точность имела среднее значение 0.63 и стандартное отклонение 0.16. Среднее значение F-score 0.73.

Намного лучше результаты получились у создателей данного метода описанного в статье [1]. Исходя из рисунка 6 (а) метод достиг чувствительности детекции в 92% со средним количеством ложно-положительных срабатываний меньше 6 на скан. Авторам удалось достигнуть большого успеха в задаче сегментации, где Dice = 87,3%. Иллюстрация сегментации с помощью FracNet представлена на рисунке 6 (b). Качество сегментации непосредственно зависит от формы кости. Вытянутые форму чаще хуже сегментируется, в то время как более округлые лучше. На рисунке 6 (c) это хорошо видно.

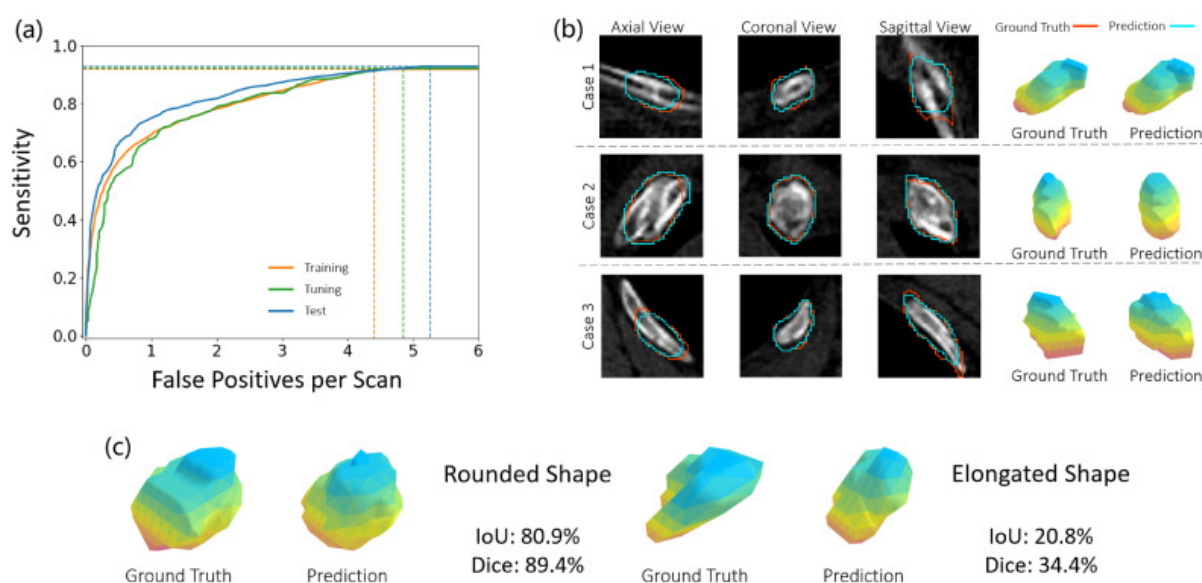


Рисунок 6. (а) Кривые FROC эффективности обнаружения FracNet на обучающей, валидационной и тестовой когортах RibFrac. (b) Иллюстрация предсказанной сегментации на тестовых когортах RibFrac. (c) Сравнение метрик сегментации (IoU и Dice) для округлой и вытянутой формы. (Иллюстрации взяты с [1]).

## **Вывод**

Изменив размер обрабатываемых картинок и длительность обучения мне удалось наблюдать стабильно улучшающиеся значения метрик и плавность тренировки, что является хорошими признаками любой модели глубокого обучения. В дальнейшем, я бы хотел усовершенствовать архитектуру модели добавив в неё больше уровней состоящих из 3D-свертки, батч-нормализации, нелинейности и максимального пулинга.

## Литература

- [1] [Liang Jin, Jiancheng Yang, Kaiming Kuang et al. Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet.](#)
- [2] [Wei Zhao, Jiancheng Yang, Yingli Sun et al. 3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas.](#)
- [3] [Hai-tao Zhang, Jin-song Zhang, Hai-hua Zhang et al. Automated detection and quantification of COVID-19 pneumonia: CT imaging analysis by a deep learning-based software.](#)
- [4] [Anna Majkowska, Sid Mittal, David F. Steiner, et al. Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation.](#)
- [5] [Nicholas Bien, Pranav Rajpurkar, Robyn L. Ball, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet.](#)
- [6] [Xuanang Xu; Fugen Zhou; Bo Liu; Dongshan Fu; Xiangzhi Bai. Efficient Multiple Organ Localization in CT Image Using 3D Region Proposal Network.](#)
- [7] [Zhao X, Xie P, Wang M et al. Deep learning-based fully automated detection and segmentation of lymph nodes on multiparametric-mri for rectal cancer: a multicentre study.](#)
- [8] [Topol EJ. High-performance medicine: the convergence of human and artificial intelligence.](#)
- [9] [Talbot BS, Gange Jr., CP, Chaturvedi A, Klionsky N, Hobbs SK, Chaturvedi A. Traumatic rib injury: patterns, imaging pitfalls, complications, and treatment.](#)
- [10] [Urbaneja A, De Verbizier J, Formery AS et al. Automatic rib cage unfolding with CT cylindrical projection reformat in polytraumatized patients for rib fracture detection and characterization: Feasibility and clinical application.](#)
- [11] [Jin L, Ge X, Lu F, et al. Low-dose CT examination for rib fracture evaluation: a pilot study.](#)



- [12] [Kolopp M, Douis N, Urbaneja A et al. Automatic rib unfolding in postmortem computed tomography: diagnostic evaluation of the OpenRib software compared with the autopsy in the detection of rib fractures.](#)
- [13] [Falk T, Mai D, Bensch R et al. U-Net: deep learning for cell counting, detection, and morphometry.](#)
- [14] [Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation.](#)
- [15] [Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation.](#)
- [16] [Hara K, Kataoka H, Satoh Y. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?.](#)
- [17] [RibFrac Dataset](#)
- [18] [MICCAI 2020 RibFrac Challenge: Rib Fracture Detection and Classification.](#)
- [19] [Dice Coefficient, Intersection over union](#)
- [20] [S H Cho, Y M Sung and M S Kim. Missed rib fractures on evaluation of initial chest CT for trauma patients: pattern analysis and diagnostic value of coronal multiplanar reconstruction images with multidetector row CT.](#)