

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
“ВЫСШАЯ ШКОЛА ЭКОНОМИКИ”

Факультет Санкт-Петербургская школа физико-математических и компьютерных наук
Национального исследовательского университета
«Высшая школа экономики»

Антонова Екатерина Алексеевна

**Детекция основных анатомических структур челюстно-лицевой области на основе
стоматологических КЛКТ**

Курсовая работа

по направлению подготовки 01.04.02 «Прикладная математика и информатика»

образовательная программа «Анализ больших данных в бизнесе, экономике и обществе»

Научный руководитель

доцент НИУ ВШЭ в Санкт-Петербурге

к.ф.-м.н. Минабутдинов Алексей Рафаилович

Подпись студента

Санкт-Петербург 2019

Содержание

Введение	3
Глава 1. Создание искусственного ОПТГ из снимков КЛКТ	7
Глава 2. Применение Faster R-CNN для детекции срезов КЛКТ	14
Заключение	21
Список литературы	23

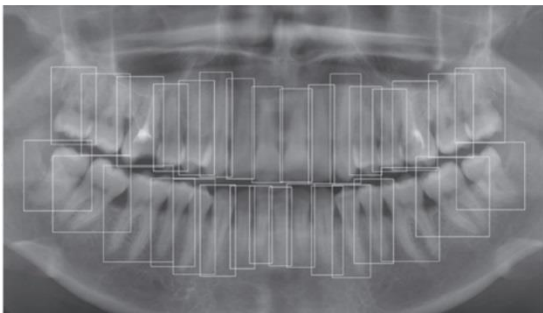
Введение

Рентгенография зубов является важным инструментом, используемым в стоматологии для диагностики и планирования лечения. Интерпретация рентгеновских снимков является трудоемким процессом, она часто подвержена ошибкам из-за индивидуальных особенностей строения челюсти, а также большого разнообразия стоматологических конструкций. Стоматологам в их повседневной практике могут помочь применение автоматизированных инструментов, которые позволяют сэкономить время и улучшить качество стоматологического лечения.

Разработкой методов детекции анатомического строения с помощью компьютерных технологий занимаются уже давно. В течение последнего десятилетия был проведен ряд исследований по анализу рентгеновских снимков челюстей с применением сегментации на уровне пикселей. Они были основаны на традиционных методах компьютерного зрения, таких как пороговые, основанные на гистограмме и т.д. (Lin et al., 2010; Hosntalab et al., 2010) Процесс этот получался достаточно трудоемким. В частности, в своем исследовании Ю. Мики и др. вручную отмечали каждый ограничивающий прямоугольник (контейнер, или bounding box) на изображениях компьютерной томограммы. Нумерация зубов проходила в два этапа: извлечение признаков и классификация. С помощью нейронной сети, наложенной на сегментированные зубы, исследователи определяли ширину, высоту зубов, соотношение их между собой и размер коронки. Далее, чтобы классифицировать зубы, использовались опорные векторные машины (SVM), алгоритм выравнивания последовательностей и нейронные сети прямого распространения (Miki et al., 2017).

В одном из последних исследований, связанных с тематикой данной работы, Д. В. Тузова с соавторами было предложено новое комплексное решение на

основе сверточной нейронной сети (CNN (convolutional neural network)). Она использовалась для обнаружения и нумерации зубов на снимках



ортопантограммы (ОПТГ). ОПТГ представляет собой рентгенологический снимок, отображающий панораму зубов, то есть на нем видно развернутое изображение зубов с челюстями и прилежащими отделами.

Рис. 1. Размеченное изображение ОПТГ способствует точной постановке диагноза, помогает контролировать лечение и избежать многочисленных осложнений.

CNN – это глубокие нейронные сети. Они являются однонаправленными и имеют в своей структуре операцию «свертки», от которой и получили свое название. Архитектура этого вида сетей позволяет переходить от подробного описания изображения, его конкретных особенностей к более абстрактным деталям, через настройку иерархии признаков, фильтруя маловажные и выделяя существенные. Эти архитектуры обычно применяются к решению проблем компьютерного зрения. В исследовании Д.В. Тузова и др. была применена модель сверточной нейронной сети Faster R-CNN, натренированная на 1574 анонимизированных изображениях. Такой подход для детекции зубов дал высокие результаты: чувствительность – 0.9941, точность – 0.9945 (Tuzoff et al., 2019).

Приведенные исследования проводились для анализа 2D-изображений ОПТГ и прикусных снимков (bitewings). До недавнего времени стоматологами обычно использовались рентгеновские дентальные и панорамные снимки зубов. Они давали только плоскостное изображение исследуемой области. Вполне очевидно, что такие снимки не всегда давали возможность получить всю необходимую информацию для постановки точного диагноза. Однако

существует сравнительно новый инструмент для диагностики, который используется в стоматологии – конусно-лучевая компьютерная томография зубных рядов (далее – КЛКТ).

КЛКТ проводится с помощью дентального компьютерного томографа, который дает возможность получить объемное изображение в высоком разрешении. Полученную таким образом 3D-модель можно рассматривать под любым углом и на произвольных срезах, что повышает эффективность последующего лечения. КЛКТ проводится для оценки верхней и нижней челюстей и обнаружения изменений костной структуры. Обычно к этому виду исследования прибегают при необходимости анализа состояния челюстей перед имплантацией или после нее, в случае патологии или аномалии прикорневой зоны, травматического повреждения челюсти, для поиска ретенированных зубов, определения скрытых потенциальных источников воспалений и других сложных для диагностики случаев.

В отношении 3D-снимков КЛКТ работ по детекции практически не проводилось. Одним из немногих исключений является статья А. Закиров и др. В этом исследовании они предприняли попытку решения задачи сегментации с помощью модели сверточной сети V-Net, однако препроцессинг был выполнен вручную: разметка bounding box (2D и 3D контейнер), и после нее выделение маски для сегментации. Точность детекции зубов в данном случае составила 0.96 (Zakirov et al., 2018).

Так или иначе, работы по детекции зубов сталкиваются с одной и той же тяжелой проблемой – «холодного старта». Разметка зубов вручную является трудоемким процессом и требует больших ресурсов. Задача данной работы заключается в том, чтобы приблизиться к автоматическому сегментированию. Детекция анатомических структур (челюстей и зубов) проводится в отношении 3D-снимков КЛКТ, как наиболее новом и точном инструменте для анализа

стоматологических проблем. Данные, которое находятся в нашем распоряжении, являются частично размеченными и неразмеченными, что усложняет решение поставленной задачи. Для решения данной проблемы потребуются совместить накопленный богатый опыт при анализе 2D-снимков ОПТГ и первые удачные результаты использования сверточных сетей для работы с 3D-снимками КЛКТ.

Глава 1. Создание искусственного ОПТГ из снимков КЛКТ

Алгоритмы машинного обучения направлены на изучение важных функций из набора доступных данных и применения этих знаний для дальнейшей интерпретации ранее неисследованных примеров. В компьютерном зрении, основная проблема заключается в извлечении функций, которые были бы полезны для семантических проблем высокого уровня, таких как распознавание объекта. В методах машинного обучения используются два основных подхода к обучению: обучение с учителем, когда аннотированные данные используются для представления основополагающей информации для обучения, и без учителя, когда алгоритм направлен на выявление кластеров данных без использования аннотаций.

Одной из главных проблем решения задачи детекции анатомических структур, в частности зубов, на 3D-снимках КЛКТ является дорогостоящая разметка снимков. Каждый 3D-снимок делится на большое количество 2D-срезов (в среднем по 500), на которых и происходит разметка. Если таких снимков насчитывается несколько тысяч, то можно себе представить масштаб работы, которую необходимо проделать вручную. Поэтому главная задача состоит в том, чтобы научиться выделять главные анатомические структуры на снимках, не затрачивая дополнительные ресурсы на разметку данных. Для задачи детекции зубов на 2D-снимках ОПТГ уже есть готовые решения и натренированная нейронная сеть. Подобные разработки уже активно применяются в исследованиях.

Возможный подход к решению задачи холодного старта заключается в том, чтобы взять данные КЛКТ и уже имеющиеся инструменты для работы с 2D-снимками ОПТГ, и выделить из снимков КЛКТ многообразие, которое можно было бы перевести в аналогичный ОПТГ вид (искусственный ОПТГ). К преобразованному изображению возможно будет применить уже существующие

методики, не затрачивая ресурсы на разметку снимков. Конечно, информация, полученная при анализе с 2D-среза, недостаточна для того, чтобы полностью восстановить 3D bounding box. Эту проблему еще предстоит решить. Для решения задачи было бы интересно разработать соответствующие байесовский подход. Например, информацию с полученной искусственного ОПТГ можно было бы попытаться объединить с информацией с других проекций. У нас появится много априорной информации, которую можно будет в дальнейшем учитывать разными способами.

В нашем распоряжении имеется около 70 3D-снимков КЛКТ. Для того, чтобы перевести эти снимки в аналогичный ОПТГ вид, были использованы следующие подходы:

- 1) Квантильная регрессия. Это статистический метод, предназначенный для оценки заданного уровня квантиля. Квантиль порядка τ – это числовое значение, вероятность попасть левее которого равна τ . Так же, как классические методы линейной регрессии, основанные на минимизации сумм квадратов и позволяющие оценить модели для условных средних функций, квантильная регрессия предлагает механизм для оценки моделей условной медианной (как частный случай квантили 0.5) или других условных квантилей. Квантильная регрессия – это проблема линейного программирования, которая решается симплекс-методом. Квантильная регрессия способна обеспечить более полный статистический анализ стохастических отношений среди случайных величин.
- 2) Параметрическое представление кривой
$$\gamma: t \in T \subset \mathbb{R} \rightarrow \bar{r}(t) = (x_1(t), x_2(t)) \in \mathbb{R}^2.$$
$$\gamma: \begin{cases} x_1 = f_1(t) \\ x_2 = f_2(t) \end{cases}, \quad t \in T$$
- 3) Натуральная параметризация. Параметр $s \in T$ называется натуральным параметром кривой γ , если длина

$l_{[s_1, s_2]} = |s_2 - s_1| \Leftrightarrow \|\vec{v}(s)\| = 1$, для любых $s \in T$. Параметризация кривой естественным параметром известна как ее естественная (натуральная) параметризация.

- 4) Интерполяция на регулярной сетке в произвольных измерениях. В нашем случае трилинейная интерполяция. Это метод многомерной интерполяции на трехмерной регулярной сетке. Он аппроксимирует значение функции в промежуточной точке (x, y, z) в пределах локальной осевой прямоугольной призмы, используя данные функции в точках решетки.

Задачу построения искусственного ОПТГ можно разделить на две части. Для нахождения кривой, задающейся линией зубов на снимках («зубной дуги»), была сформулирована задача квантильной регрессии со значениями 0.95. После идентификации линии зубов, получаем множество точек, образующих кривую. Мы находим естественный параметр кривой и работаем с кривой в терминах естественной параметризации, так как это позволяет считать расстояние между отдельными фрагментами снимка. Параметр на кривой называется натуральным, если разность его значений для двух соседних точек равна модулю длины дуги между этими точками.

Рассмотрим семейство кривых, получаемых из кривой γ_0 с помощью сдвига каждой точки кривой вдоль ее единичной нормали. В качестве начальной кривой γ_0 возьмем параболу $x_2 = x_1^2$. Натуральную параметризацию γ_0 можно получить задав параметрически кривую γ_0 :

$$\gamma_0: \begin{cases} x_1 = t \\ x_2 = t^2 \end{cases} \Leftrightarrow x_2 = f(x_1) = x_1^2, \quad t \in \mathbb{R}.$$

Касательный вектор к кривой γ_0 имеет вид: $\vec{v}(t) = \frac{d\vec{r}}{dt} = \left(\frac{dx_1}{dt}(t); \frac{dx_2}{dt}(t) \right)$, где

$$\frac{dx_1}{dt} = \dot{x}_1, \quad \frac{dx_2}{dt} = \dot{x}_2; \quad \|\vec{v}\| = \sqrt{\dot{x}_1^2 + \dot{x}_2^2} - \text{длина касательной вектора.}$$

Определим новую параметризацию $t \in \Gamma \rightarrow \tau \in \mathbf{T}; t = \varphi(\tau)$, такую что:

$\varphi'(\tau) \neq 0$, например $\varphi'(\tau) > 0$. Тогда $\vec{r}(t) = \vec{r}(\varphi(t)) = (\vec{r} \circ \varphi)(t)$.

Возьмем в качестве нового параметра естественный параметр кривой – s ;

Тогда вектор касательной задается следующим выражением:

$$w(s) = \left(\frac{dx_1(\varphi(s))}{ds}, \frac{dx_2(\varphi(s))}{ds} \right).$$

Обратная функция к φ имеет следующий вид.

Теорема 1.

Пусть $\gamma_0: \begin{cases} x_1 = t \\ x_2 = t^2 \end{cases}, t \in \mathbb{R}$. Тогда $\varphi^{-1}(t) = \frac{1}{2}t\sqrt{1+4t^2} + \frac{1}{8}\arctg 2t$.

Доказательство:

Парабола $x_2 = x_1^2$ в натуральной параметризации задается таким образом:

$$\gamma_0: \begin{cases} x_1 = \varphi(s) \\ x_2 = \varphi^2(s) \end{cases}, \quad \text{где } t = \varphi(s).$$

Вектор касательной к γ_0 имеет вид: $w(s) = \left(\frac{d\varphi}{ds}, 2\varphi(s) \frac{d\varphi}{ds} \right) = \varphi'(s) \cdot (1, 2\varphi(s))$.

Условие нормализации $\|w(s)\| = 1$ приводит к дифференциальному уравнению на функцию φ : $\varphi'(s)^2 + 4\varphi(s)^2\varphi'(s)^2 = 1 \Leftrightarrow \varphi'(s)^2 = \frac{1}{1+4\varphi(s)^2}$.

Его общее решение запишем в виде квадратуры: $s = \int \sqrt{1 + (2\varphi)^2} d\varphi + C$.

Так как в качестве кривой была взята парабола, интеграл считается явно.

$$\int \sqrt{1 + (2\varphi)^2} d\varphi = \frac{1}{2}\varphi\sqrt{1 + (2\varphi)^2} + \frac{1}{4}\ln \left| 2\varphi + \sqrt{1 + 4\varphi(s)^2} \right| + \ln C.$$

Таким образом, $s = \frac{1}{2}\varphi\sqrt{1 + (2\varphi)^2} + \frac{1}{4}\ln \left| 2\varphi + \sqrt{1 + 4\varphi(s)^2} \right| + \ln C$.

Константа C определяется из условия: $t = 0 \Rightarrow x_1 = 0, x_2 = 0 \Leftrightarrow \vec{r}(0) = \vec{0} \Rightarrow C = 1$.

В итоге приходим к искомому выражению для $s = \varphi^{-1}(t)$:

$$s = \frac{1}{2}t\sqrt{1+4t^2} + \frac{1}{8}\arctg 2t \quad \blacksquare$$

Для получения натуральной параметризации необходимо решить уравнение из Теоремы 1, что делается численно.

Построим нормаль к кривой γ_b .

Кривая γ_b определяется таким соотношением:

$$\gamma_b: \begin{cases} x_1 = t \\ x_2 = t^2 + b\|\bar{n}\| \end{cases} \Leftrightarrow x_2 = x_1^2 + b\|\bar{n}\|, \text{ где нормаль определяется} \\ \text{следующим образом: } \vec{N} \perp \vec{v} \Leftrightarrow \langle \vec{N}, \vec{v} \rangle = 0.$$

$$l: \dot{f}_1(\tilde{x}_1 - x_1) + \dot{f}_2(\tilde{x}_2 - x_2) = 0, n: \tilde{x}_2 - x_2 = -\frac{1}{\dot{f}_2(x_2)}(\tilde{x}_1 - x_1).$$

$$\vec{N} = \left(\dot{f}_1, -\frac{1}{\dot{f}_2} \right) - \text{вектор нормали (ненормированный) к } \gamma_b.$$

В итоге выражение для кривой γ_b следующее:

$$\gamma_b: \begin{cases} \tilde{x}_1(t) = x_1(t) + bn_1(t) \\ \tilde{x}_2(t) = x_2(t) + bn_2(t) \end{cases} \Leftrightarrow \begin{cases} \tilde{x}_1(t) = t + b \frac{2t}{\sqrt{4t^2 + 1}} \\ \tilde{x}_2(t) = t^2 - b \frac{1}{\sqrt{4t^2 + 1}} \end{cases}$$

Рассмотрим семейство Γ , состоящее из цилиндрических поверхностей, имеющими в основании параболу. Итоговый искусственный ОПТГ получается как среднее по параметру b в границах от γ_1 до γ_2 , которые подбираются эмпирически.

Реализация:

1. Прочитав том 3D КЛКТ, выбираем эмпирически нужный нам 2D-срез, расположенный в середине тома.
2. Для нахождения кривой, огибающей расположение зубов, используем квантильную регрессию, с квантилем 0,95. Таким образом задается и описывается зубная дуга. (Рис. 2).
3. Получаем множество точек и коэффициенты для квадратного трехчлена.

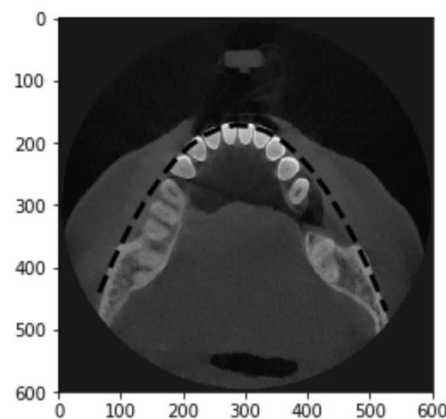


Рис. 2. Дуга, описывающая расположение зубов.

4. Считаем длину параболы, через интеграл $\int_{l_1}^{l_2} \sqrt{1 + (2ax + b)^2} dx$ и убираем лишние темные зоны.
5. Определяем интерполирующую функцию на основе данных. Иными словами, далее мы численно интерполируем изображение, чтобы избежать дискретных вокселей.
6. Через цикл генерируем множество параллельных кривых, так как каждое изображение должно быть сгенерировано из другой параллельной прямой. Вложенным циклом идем по длине кривой, используя метод Ньютона, чтобы найти натуральную параметризацию. Интерполируем координаты.
7. Каждую получившуюся кривую добавляем в массив данных. Итоговый искусственный ОПТГ получается, как усреднение массива.

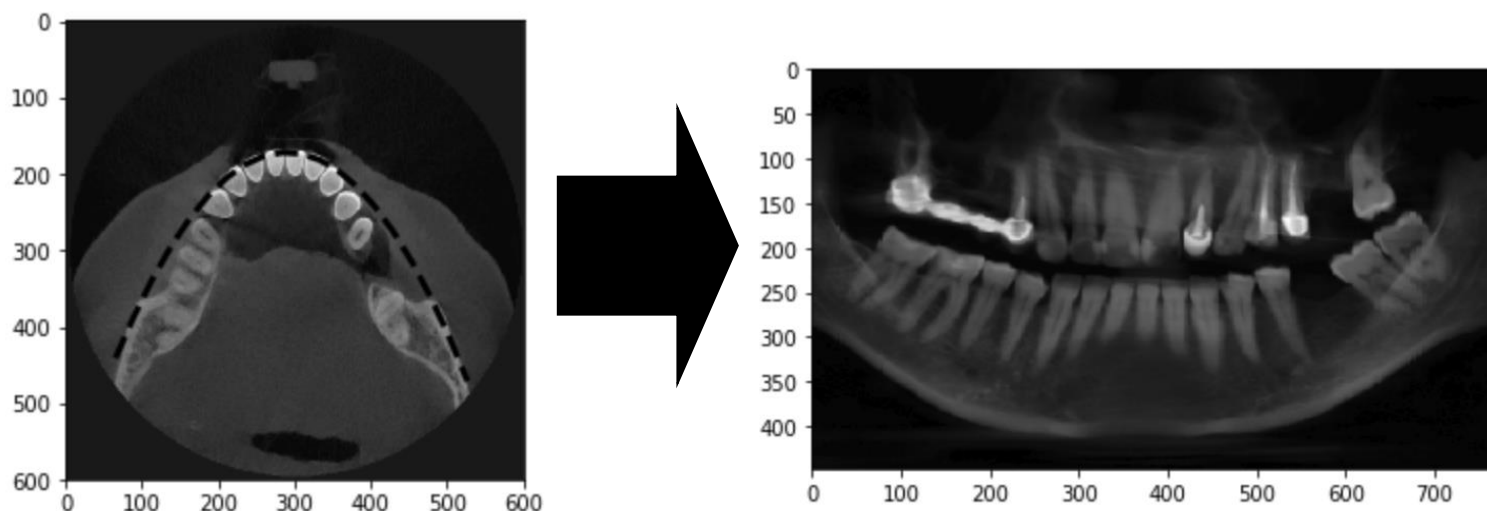


Рис. 3. Искусственный ОПТГ

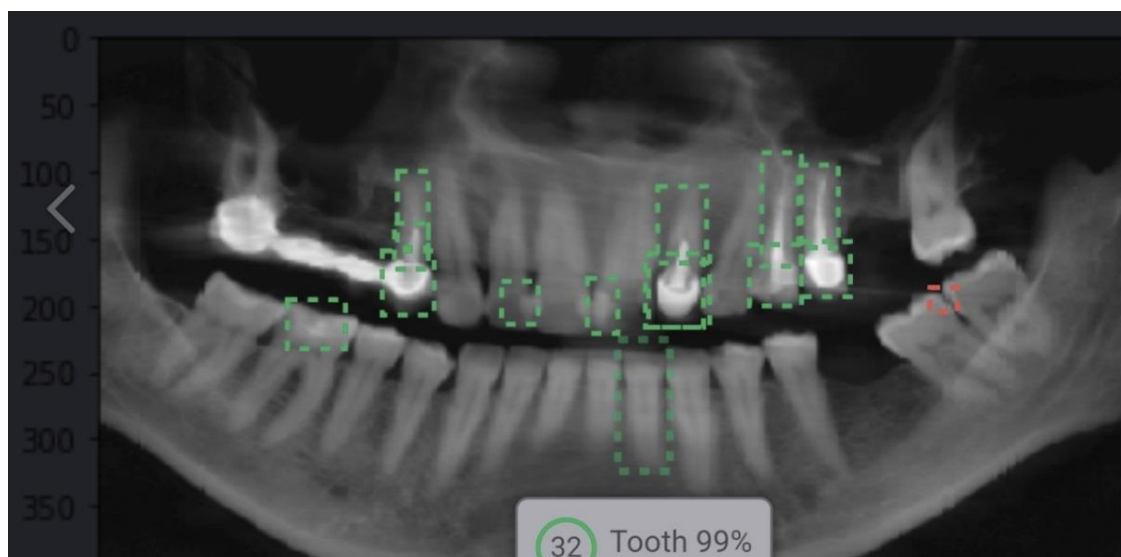


Рис. 4. Детектирование анатомических структур на искусственном ОПТГ с помощью нейронной сети Dental.AI

В результате приведенных операций были получены искусственные ОПТГ из 3D-снимков КЛКТ. Образцы были загружены в нейронную сеть Dental AI. В итоге почти все зубы были найдены и пронумерованы автоматически. Кроме того, нейронная сеть смогла найти большинство патологий. (См. Рис. 4). Однако в отличие от результатов на настоящих ОПТГ, где точность детекции составляет 0,99, можно отметить очевидные проблемы с распознаванием крайних структур на челюсти, прежде всего третьих маляров (зубов «восьмерок») или, например, мостовых конструкций, если они находятся на краю снимка. Как уже отмечалось выше, невозможно получить полное многообразие данных для восстановления 3D контейнера из 2D-панорамы. Однако данная панорама располагает важной информацией, например высотой зубов, их расположением, которую можно использовать как априорную для дальнейших преобразований. Несмотря на ограниченные результаты, следует признать, что опыт преобразования структур оказался позитивным. Полученные данные, как представляется, могут стать важным этапом для дальнейших исследований в этом направлении, использовании более подходящих для данной задачи нейронных сетей.

Глава 2. Применение Faster R-CNN для детекции срезов КЛКТ

Второй возможный подход к решению проблемы детекции зубов на снимках КЛКТ состоит в применении методов глубокого обучения. Глубокое обучение – это класс обучаемых алгоритмов искусственного интеллекта (ИИ), который позволяет компьютерной программе автоматически извлекать и изучать важные функции входных данных для дальнейшей интерпретации ранее невидимых образцов. Основное отличие методов глубокого обучения заключается в том, что они могут учиться на основе ввода необработанных данных, например пикселей изображений, без необходимости ручной разработки функций. Глубокие сверточные нейронные сети (CNN) – один из самых популярных наборов методов глубокого обучения, который обычно применяется для задач распознавания изображений. Архитектуры CNN используют конкретные характеристики ввода данных изображения, такие как пространственные отношения между объектами, для эффективного представления и изучения иерархических особенностей с использованием нескольких уровней абстракции.

Архитектуры CNN имеют корни в имитации зрительной коры. В 50-х и 60-х годах XX века, изучая мозг кошек, Хьюбел и Визель обнаружили, что зрительная кора кошки организована иерархически, где простые клетки реагируют на простой стимул, такой как ориентация. Сложные клетки реагируют как на ориентацию, так и на движение, и так далее с растущей сложностью (Hubel, Wiesel, 1959; Hubel, Wiesel, 1962). В 1982 году Фукусима и Мияке предложили модель нейронной сети для задачи распознавания рисунка, вдохновленные структурой зрительной коры. Эта ранняя модель была похожа на архитектуру современной CNN. Тем не менее, учеными в тот момент «обучение с учителем» не использовалось (Fukushima, Miyake, 1982). Наконец, в 1998 году Ян Лекун представил архитектуру сверточной нейронной сети, которая обучалась, за счет градиентного спуска. Этот метод применялся для задачи чтения документов

(LeCun, 1998). С тех пор было разработано много новых архитектур, однако основные идеи остаются прежними.

Подобно традиционным нейронным сетям прямого распространения, CNN состоят из нейронов, в них есть веса, которые нужно обучить, и использовать как линейные, так и нелинейные преобразования для извлечения признаков из данных. Однако в отличие от сетей прямого распространения, где нейроны полностью связаны друг с другом, и их количество на каждом уровне определяет количество параметров для изучения, архитектура CNN организована особым образом. Изучаемые веса распределяются между нейронами одной карты признаков. Для задач компьютерного зрения, архитектура CNN использует определенные свойства изображений, что позволяет гораздо более эффективно обрабатывать исходный ввод на уровне пикселей, обучая значительно меньшее число параметров. В дополнение к уменьшенному количеству параметров такая архитектура нейронной сети менее чувствительна к проблеме переобучения.

Архитектура CNN представлена последовательностью уровней разных типов. Первые слои обычно объединяют сверточные слои и пулинг. Сверточные слои очень важны для преобразования входных данных. Их параметры образуют набор обучаемых фильтров, где каждый фильтр имеет фиксированный размер, например, $5 \times 5 \times 3$. На сверточном уровне нейронная сеть проходит через фильтр по данным, получаемым с входа, и вычисляет точечные произведения между записями фильтра и входными данными. Результаты обычно передаются через нелинейную функцию, создающую карту объектов. Смысл процесса заключается в том, что сеть изучает отличительные шаблоны для каждого применяемого фильтра, которые пространственно коррелированы и могут быть найдены во входных данных.

Для изображения сеть может узнать фильтры, которые активируются, когда они видят некоторый тип визуального шаблона, например, линии определенной ориентации. Слои пулинга используются для уменьшения пространственного

размера входного представления. Эта процедура позволяет, с одной стороны, уменьшить количество параметров для обучения. С другой стороны, она не позволяет нейронной сети переобучаться. Итогом сверточного слоя, с пулингом или без, является трехмерный тензор, где количество применяемых фильтров определяет глубину. Тензор, который мы получаем на выходе, может служить входом для следующего сверточного слоя, делая таким образом сеть глубже. Последние слои CNN часто представлены полносвязными слоями, где каждый нейрон соединен со всеми нейронами предыдущего слоя. Последний полносвязный слой, называемый выходным слоем, который выводит итог нейронной сети.

В течение последнего десятилетия были предприняты усилия для разработки архитектур на основе CNN для решения многих задач компьютерного зрения, включая классификацию изображений, локализацию объектов, обнаружение объектов и сегментацию (Goodfellow, Bengio, Courville, 2016; LeCun, Bengio, Hinton, 2015; Schmidhuber, 2015; Hosang et al., 2017).

В стоматологии проблема классификации объектов рассматривается в одних работах для обнаружения ориентира (Arik, Ibragimov, Xing, 2017), в других для классификации зубов на компьютерной томограмме (КТ), наконец, третьи используют этот вопрос для классификации зубов по панорамным снимкам (Tuzoff et al., 2018). Мики с соавторами делали исследования на архитектуре AlexNet CNN (Miki et al., 2016; Krizhevsky, Sutskever, Hinton, 2012). Группа исследователей во главе с Д.В. Тузовым использовала VGG-16 в качестве базовой CNN для решения задачи классификации зубов (Tuzoff et al., 2018; Simonyan, Zisserman, 2015).

Задача обнаружения направлена не только на классификацию отдельного объекта на изображении, но и на вывод координат контейнера (bounding box). Эта задача обычно рассматривается как задача регрессии, когда сеть производит как оценки для класса, так и действительные числа, которые можно перевести на

контейнер (bounding box). CNN, используемые для обнаружения объекта, обычно основаны на тех же архитектурах, что и CNN для классификации, с той лишь разницей, что последний уровень этих сетей выводит регрессор и классификатор. Для обучения модели обычно используется функция ошибок.

Задача обнаружения объектов направлена на обнаружение различного количества объектов на одном изображении и отметкой их через контейнер (bounding box). Один из наиболее популярных подходов к детекции состоит из двух этапов: 1) нахождение области интереса (RoIs); 2) получение меток классов и координат для контейнера (bounding box) с использованием найденных областей. В ранних архитектурах подобного типа для генерации RoI использовались внешние алгоритмы без обучения, например, выборочный поиск (Girshick et al., 2015). В 2017 году была предложена современная архитектура Faster R-CNN, которая объединяет нахождение области интереса с обнаружением объекта (Ren et al., 2017). Представителем такой архитектуры, в которой существует однофазный подход, является YOLO (Redmon, Farhadi, 2017) и SSD (Liu et al., 2016).

В 2018 году был предложен метод обнаружения и нумерации зубов на панорамных изображениях (Tuzoff et al., 2018). Д.В. Тузов с соавторами предложил комплексное решение для обнаружения и классификации зубов на основе CNNs. Зубы были обнаружены на изображениях и пронумерованы в соответствии с двузначным стандартом ПИИ (ISO, 2016). Набор данных из 1574 изображений был использован для обучения и тестирования модели. Набор исходных данных был разделен на: 1352 изображений, составляющих обучающую выборку, 222 изображений для тестовой выборки. Пять экспертов предоставили аннотации к изображениям. Авторы использовали современную архитектуру Faster R-CNN для модуля детекции (Ren, He, Girshick, Sun, 2017).

Для решения задачи детекции анатомических структур в условиях холодного старта, попробуем применить нейронную сеть, на основе модели,

предложенной коллективом авторов во главе с Д.В. Тузовым. В качестве архитектуры нейронной сети была использована та же самая Faster R-CNN. Она представляет собой единую унифицированную сеть двух модулей: сеть региональных предложений (RPN) и детектор объекта. RPN предлагает участки, на которых могут быть найдены интересующие нас объекты. Детектор объекта использует эти предложения для дальнейшей локализации объекта и классификация. Оба модуля совместно используют слои свертки базовой CNN, которая обеспечивает компактное представление исходного изображения, известного как карта объектов.

Необходимые признаки выучиваются на этапе обучения, а не добавляются вручную, что является ключевым отличием по сравнению с классическим компьютерным зрением. Модуль находит зубы на изображении, выводя контейнеры (bounding box). Для создания региональных предложений RPN «проходится» по карте признаков и в каждом окне местоположения, выделяет потенциальный контейнер (bounding box), именуемый «якорь». Для каждого якоря RPN оценивает вероятность того, что он будет содержать объект или фон, и сжимает контейнер специальным регрессором. Лучшие предложения в рейтинге затем используются в качестве входа для сети обнаружения объектов. Детектор объекта уточняет оценку класса области, чтобы определить зуб или фон для задачи обнаружения, и генерирует окончательные координаты контейнера (bounding box).

VGG-16 Net16 использовался в качестве базовой CNN для обоих RPN и обнаружение объектов. VGG-16 представляет собой 16-слойную CNN архитектуру. Она получила название в честь исследовательской группы, которая разработала эту сеть (Visual Geometry Group, Отдел инженерных наук, Оксфордский университет).

Во время обучения веса моделей были предтренированы на набор данных ImageNet, который использовался для основного CNN. Все слои CNN были

настроены, так как набор данных значительно отличается от ImageNet. Начальный learning rate был выбран как 0,001 с дальнейшим экспоненциальным затуханием. Модель обнаружения зубов была реализована, как настраиваемая версия Faster. Реализация Python R-CNN с помощью TensorFlow бэкенд.

Сложной задачей компьютерного зрения в частности является построение качественной тренировочной выборки. Чтобы решить данную проблему, выбрана разметка (горизонтальных) сечений КЛКТ, что позволяет использовать существующие инструменты разметки совместно с эффективными алгоритмами анализа 2D-изображений. Для выделения основных структур используются габаритные контейнеры (bounding box). Разметка на срезах была неполной. Было размечено 354 горизонтальных срезов, 50 из которых были отнесены к тестовой выборке (bounding box, в случае 2D – прямоугольники, см. иллюстрации).

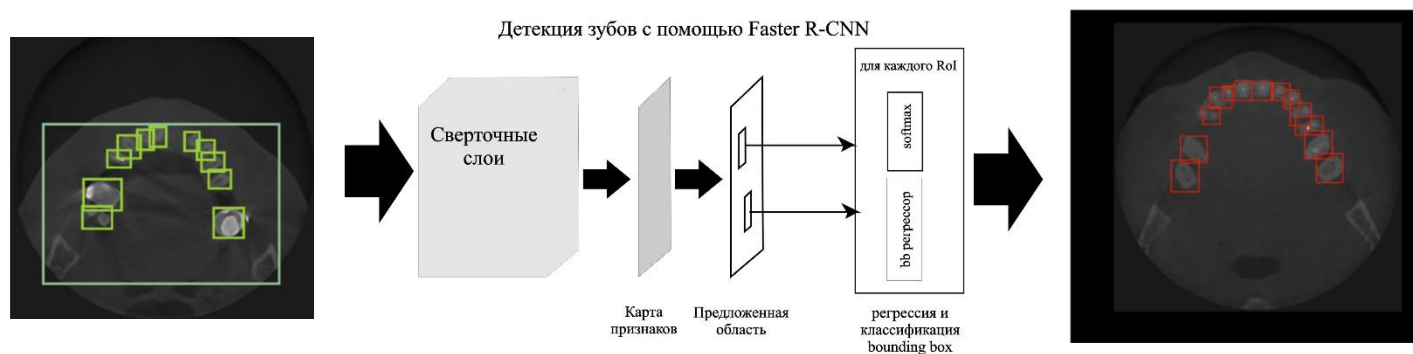


Рис. 5. Схема сверточной нейронной сети на основе архитектуры Faster R-CNN

Каждое изображение было проанализировано системой. Тестовые данные не загружались в систему на этапе обучения. Аннотации, сделанные системой и экспертами, были сравнены для оценки производительности. В тех случаях, когда система и эксперт предоставил одинаковые аннотации, оба были сочтены правильными.

На основе неполной разметки были получены следующие результаты:
Текущие результаты работы алгоритма:

Табл. 1:

	Текущие показатели (КЛКТ)	Целевые показатели (ОПТГ)
Precision	0.87	0,9945
Recall	0.98	0,9941

Таким образом, точность нейронной сети, построенная по не полностью размеченным данным, составляет 87%. В среднем на 12-15% ниже в сравнении с нейронной сетью, обученной на 1352 панорамных рентгеновских снимках (ОПТГ). Для задачи объединения 2D габаритных контейнеров в 3D были испробовано различные алгоритмы кластеризации, однако положительного результата получить не удалось. Лучше всего работает примитивный алгоритм, который пробегает сверху-вниз и сильно пересекающиеся 2D контейнеры относит к одному (будущему) 3D контейнеру. Следует признать, что для достижения результатов экспертного уровня для этих задач новые системы, скорее всего, потребуют гораздо большее количество размеченных изображений для обучения, чем используется в этой работе. Эти направления исследований могут потребовать расширения количества классов для Faster R-CNN, реализующих методы сегментации для более точной локализации зубов, экспериментирования с новыми архитектурами и сетями. Важное преимущество использования подхода CNN заключается в том, что шаги по улучшению могут вводиться постепенно, и результаты предыдущих шагов могут быть повторно использованы: доработка существующей модели, обучение новых моделей на уже аннотированных данных, сегментация или локализация объектов внутри ранее обнаруженных границ.

Заключение

В данном исследовании была поставлена задача детекции основных анатомических структур на снимках КЛКТ. Для ее решения были применены два метода, которые можно использовать для решения этой задачи, в условиях холодного старта. Иными словами, главное внимание было сосредоточено на актуальных проблемах компьютерного зрения. В значительной степени данный проект опирался на исследования последних лет, имевших определенные успехи в данной сфере распознавания изображений.

Первый метод состоял в том, чтобы выделить из снимков КЛКТ многообразие, которое можно было бы перевести в аналогичный ОПТГ вид (искусственный ОПТГ). И проанализировать на существующей нейронной сети, натренированной на анализ снимков ОПТГ. Данную задачу удалось решить, после чего образцы были загружены в нейронную сеть Dental AI. Точность составила 91%. В итоге почти все зубы были найдены и пронумерованы.

Второй подход заключался в применении глубоких сверточных сетей для анализа многослойных КЛКТ-снимков. В результате удалось подтвердить гипотезу о том, что на основе CNN сеть может быть обучена для обнаружения зубов на срезах КЛКТ даже при неполной разметке данных. Точность нейронной сети, построенной по не полностью размеченным данным, составляет 87%. Можно сделать вывод, что применение данного метода имеет высокий потенциал для достижения результатов экспертного уровня. Вероятно, использование гораздо большего количества изображений принесет значительно лучшие результаты.

Перспективой для развития данного исследования является совмещение первого и второго методов. В дальнейшем планируется извлечение информации из искусственных ОПТГ для доработки 3D-многообразий, полученных с

помощью нейронной сети. Можно предположить, что такой подход уменьшит недостатки и использует все достоинства обоих способов работы со снимками.

Кроме того, дальнейшей насущной задачей является улучшение качества указанных алгоритмов для достижения целевых показателей (текущий state-of-the-art, см. табл. 1), полученных для аналогичной задачи анализа панорамных снимков (ОПТГ). Для этих целей планируется итеративно улучшить и уточнить разметку, использовать нейросетевые алгоритмы для агрегирования полученных 2D - bounding boxes в 3D- bounding boxes.

Многообещающей альтернативой алгоритмам детекции являются алгоритмы сегментации. В этом случае для выделения объектов используются не габаритные контейнеры, а классифицируются непосредственно пиксели (в 2D) и воксели (в 3D) изображений. В силу крайней трудоемкости получения тренировочной выборки (разметка попиксельно слой за слоем 3D-изображения представляется недоступной) задача автоматической сегментации основных анатомических структур представляет как практический, так и научный интерес. Если удастся достигнуть требуемого качества анализа, то в перспективе открывается возможность для перехода к очередному этапу, например, использования новых методов для поиска кариеса, апикального периодонтита, периимплантита, ретенции, убыли костной ткани, эндодонтических осложнений, кист, новообразований и т.д.

Список литературы

1. Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., Zheng X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. [Электронный ресурс] URL: <https://www.tensorflow.org/> (дата посещения 04.06.2019);
2. Arik S.Ö., Ibragimov, B., Xing, L. Fully automated quantitative cephalometry using convolutional neural networks // Journal of Medical Imaging. 2017. № 4. 014501.
3. Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L. ImageNet: A large-scale hierarchical image database // The Conference on Computer Vision and Pattern Recognition (CVPR). 2009. P. 248-255;
4. Fukushima K., Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position // Pattern Recognition. 1982. Vol. 15. № 6. P. 455-469.
5. Girshick R. Fast R-CNN // IEEE International Conference on Computer Vision (ICCV). 2015. P. 1440-1448;
6. Girshick R., Donahue J., Darrell T., Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015. № 38. P. 142-158;
7. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016;
8. Hosntalab M., Zoroofi R.A., Tehrani-Fard A.A., Shirani G. Classification and numbering of teeth in multi-slice CT images using wavelet-Fourier descriptor // International Journal of Computer Assisted Radiology and Surgery. 2010. Vol. 5. № 3. P. 237-249;
9. Huang J., Rathod V., Sun C., Zhu M., Korattikara A., Fathi A., Murphy K. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;

10. Hubel D. H., Wiesel T. N. Receptive Fields of Single Neurons in the Cat's Striate Cortex. *Journal of Physiology*. 1959. Vol. 148. № 3. P. 574-591.
11. Hubel D. H., Wiesel T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex // *The Journal of Physiology*. 1960. Vol. 160. № 1. P. 106-154.
12. ISO 3950:2016 Dentistry--Designation system for teeth and areas of the oral cavity
13. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // *Annual Conference on Neural Information Processing Systems (NIPS)*. 2012. № 2. P. 1097-1105.
14. LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. 2015. № 521. P. 436-444;
15. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // *Proceedings of the IEEE*. 1998. Vol. 86. № 11. P. 2278-2323.
16. Lin P.L., Lai Y.H., Huang P.W. An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information // *Pattern Recognition*. 2010. Vol. 43. № 4. P. 1380-1392;
17. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y. SSD: Single Shot MultiBox Detector // *Computer Vision – ECCV*. 2016. P. 21- 37;
18. Miki Y., Muramatsu C., Hayashi T., Zhou X., Hara T., Katsumata A., Fujita H. Classification of teeth in cone-beam CT using deep convolutional neural network // *Comput Biol Med*. 2017. № 80. P. 24-29;
19. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. P. 6517-6525;
20. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39. №6. P. 1137-1149;

21. Schmidhuber J. Deep learning in neural networks: An overview // Neural Networks. 2015. Vol. 61. P. 85-117;
22. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition // arXiv: 1409.1556 [cs.CV] 10 April 2015
23. Tuzoff D.V., Tuzova L.N., Bornstein M.M., Krasnov A.S., Kharchenko M.A., Nikolenko S.I., Sveshnikov M.M., Bednenko G.B. Tooth detection and numbering in panoramic radiographs using convolutional neural networks // Dentomaxillofac Radiol. 2019. Vol. 48. № 4.
24. Tuzoff D.V. , Tuzova L.N., Kharchenko M.A. Report on tooth detection and numbering in panoramic radiographs using CNNs // 21st Conference on Medical Imaging with Deep Learning (MIDL). 2018;
25. Zakirov A., Ezhov M., Gusarev M., Alexandrovsky V., Shumilov E. Dental pathology detection in 3D cone-beam CT // arXiv:1810.10309v1 [cs.CV] 24 October 2018.
26. Николенко С., Кадури́н А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. Санкт-Петербург, 2018.