Многоструктурная сегментация КЛКТ изображений

Е. М. Шемет

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

lizshem616@gmail.com

Аннотация. Статья посвящена обработке медицинских 3D-изображений для задачи сегментации зубов и челюстных структур. В основе предлагаемого подхода используются две современные нейронные сети: 3D U-Net Swin UNETR, адаптированные для работы с и многоклассовыми медицинскими данными. Основные этапы обработки включают предобработку изображений, обучение моделей и оценку их точности. Особое внимание уделяется процессу предобработки данных, включая нормализацию интенсивности, выравнивание пространственного разрешения и перекодировку меток. В исследовании используется датасет Tooth Fairy 2, содержащий аннотированные 3D-изображения зубных и челюстных структур. Оценка точности сегментации выполняется с использованием метрик Dice и HD 95, что позволяет судить о качестве распознавания сложных анатомических структур и сравнивать эффективность обеих моделей.

Ключевые слова: нейронные сети; КТ; сегментация зубов

I. Введение

Одним из основных неинвазивных методов визуализации в стоматологии является конусно-лучевая компьютерная томография (КЛКТ). Эта технология позволяет получать высококачественные трехмерные изображения анатомических структур, что делает её незаменимым инструментом в диагностике и планировании лечения. Основные преимущества КЛКТ:

- Быстрое получение информации, что значительно ускоряет диагностический процесс.
- Низкая доза облучения по сравнению с традиционной компьютерной томографией, что снижает риски для пациента.
- Хорошая визуализация анатомических структур, особенно твердых тканей, таких как зубы, челюсти и кости черепа.

Однако, несмотря на высокое качество визуализации, процесс анализа изображений КЛКТ часто требует значительного участия экспертов, что делает его трудозатратным для стоматолога. В результате это ограничивает возможность использования всего объема информации, который можно извлечь из трехмерных изображений. Таким образом, автоматизация данной задачи является важным направлением исследований, которое может не только улучшить традиционные методы стоматологического лечения, но и открыть

А. М. Синица

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

amsinitca@etu.ru

новые высокотехнологичные подходы, такие как автоматический анализ и планирование ортодонтических процедур.

Сегментация медицинских изображений является ключевым шагом в автоматизации процессов анализа КЛКТ. Современные методы глубокого обучения, такие как 3D U-Net [1] и Swin UNETR [2], ориентированы на выполнение задач высококачественной сегментации анатомических структур. Swin UNETR, в частности, сочетает подходы сверточных и трансформерных моделей, что делает его особенно эффективным.

Для многих клинических приложений и дальнейшего анализа (например, анализа 3D-сканов в формате сеток с КТ) требуется учитывать весь набор анатомических структур из датасета Tooth Fairy 2. Однако в рамках данного исследования количество целевых классов было уменьшено с 42 до 5 основных анатомических структур.

Целью данной работы является исследование эффективности передовых моделей глубокого обучения, таких как 3D U-Net и Swin UNETR, для сегментации зубных и челюстных структур.

II. Материалы и методы

А. Датасет

Для обучения модели использовался датасет Tooth Fairy 2 [3], [4], созданный для конкурса MICCAI 2024 Tooth Fairy 2. Датасет включает:

- Изображения: Данные КЛКТ представлены в формате .mha. Они имеют одноканальную структуру, где интенсивности пикселей соответствуют значениям шкалы Хаунсфилда.
- Метки: Аннотированные данные, где каждому классу соответствует определенная структура, включая отдельные зубы и окружающие ткани.

Датасет включает 42 аннотированных класса, покрывающих следующие анатомические структуры:

- Крупные структуры: нижняя челюсть, верхняя челюсть, альвеолярные каналы.
- Дополнительные структуры: пазухи, глотка, мосты, коронки, импланты.
- Зубы: Каждый зуб помечен индивидуально, включая центральные резцы, клыки, премоляры, моляры и зубы мудрости. Аннотации включают

как верхние, так и нижние зубы, разделённые на левую и правую стороны.

Для оценки качества датасета были построены boxplot-графики, отображающие средние значения HU (единицы Хаунсфилда) для каждого класса. HU - это измерения плотности, используемые елинины компьютерной томографии. Красные крестики на Рисунке 1 указывают примерные ожидаемые диапазоны НU для каждой анатомической структуры [5], [6]. Как показано на графиках, диапазоны значений HU для всех ожидаемых пределах. классов находятся В Выбросы могут объясняться незначительными ошибками разметки.

Было проведено сокращение 42 исходных классов до 5 основных:

- Фон, неопределенные классы и нерелевантные области (классы 0, 3–10, 40): класс 0
- Нижняя челюсть (класс 1): класс 1
- Верхняя челюсть (класс 2): класс 2
- Нижние зубы (классы 31–38, 41–48): класс 3
- Верхние зубы (классы 11–18, 21–28): класс 4

Эта трансформация выполняется в специальном методе, где каждый пиксель оригинальной аннотации получает новый класс В соответствии предопределённым отображением. Для ускорения обработки используется векторизация, что позволяет применять трансформацию ко всему массиву эффективно.

Чтобы обеспечить надёжность оценки качества, датасет был случайным образом разделён на обучающую и тестовую выборки в соотношении 80/20.



Рис. 1. Боксплоты со средними значениями НU для каждого класса

В. Модель сегментации

Как уже упоминалось ранее, в качестве базовых архитектур для сегментации использовались модели 3D U-Net и Swin UNETR. Они обучались с использованием оптимизатора AdamW и функции потерь Tversky loss.

3D U-Net основан на сверточных слоях, которые обеспечивают локальную обработку данных, что делает его эффективным для крупных и хорошо визуализируемых структур. В то же время Swin UNETR сочетает сверточные блоки и механизмы трансформеров, что особенно полезно для сегментации мелких и

сложных объектов. Сравнение этих моделей между собой позволяет глубже понять их применимость к задаче сегментации зубных структур.

Предложенный конвейер сегментации включает следующие этапы:

- Загрузка изображений и меток: данные загружаются из формата .mha с использованием ITKReader.
- Предобработка данных, включаюшая нормализацию интенсивностей с учетом ненулевых пикселей: интенсивности пикселей нормализуются так, чтобы ИХ значения соответствовали диапазону [0,1], что улучшает сходимость модели, И пространственное выравнивание: для обеспечения одинакового пространственного разрешения данные масштабируются до вокселей размером 1.0 \times 1.0 \times 1.0 мм.
- Прогон модели с использованием метода скользящего окна (размер окна 96 × 96 × 96 пикселей).

Для повышения устойчивости и способности нейросетей к обобщению применялись методы аугментации данных. В частности, к изображениям применялся случайный гауссовский шум с вероятностью 10%. Кроме того, для обеспечения сбалансированного представления положительных и отрицательных классов в датасете использовалась случайная обрезка областей интереса на основе центрального пикселя класса.

Модель сегментации 3D U-Net, использованная в данной работе, имеет следующие гиперпараметры:

- Пространственные размеры: 3D
- Входные каналы: 1
- Выходные классы: 5
- Каналы: (32, 64, 128, 256)
- Шаги свёртки: (2, 2, 2, 2)
- Число остаточных блоков: 2

Для модели Swin UNETR использовались следующие гиперпараметры:

- Входные каналы: 1
- Выходные классы: 5
- Размер изображения: (96, 96, 96)
- Размер признаков: 48
- Drop rate: 0.0
- Attention drop rate: 0.0
- Dropout path rate: 0.0

Оптимизатор AdamW использовался для обучения с шагом обучения 1е-4 и коэффициентом весового затухания 1е-5. AdamW улучшает сходимость и предотвращает переобучение, сочетая преимущества оптимизатора Adam с более эффективным применением L2-регуляризации.

Также применялось обучение с смешанной точностью для уменьшения нагрузки на память и

повышения производительности при обучении и тестировании модели. В режиме смешанной точности используется 16-битный формат чисел с плавающей запятой (FP16), что снижает объем используемой памяти. Однако некоторые вычисления выполняются с 32-битной точностью (FP32), что сохраняет точность модели и ускоряет процесс обучения.

Для оптимизации работы модели использовалась функция потерь Tversky с параметрами $\alpha = 0.3$ и $\beta = 0.7$, как рекомендовано в оригинальной публикации. Она предназначена для измерения различий между двумя наборами, такими как предсказанные и эталонные маски, используемые в сегментации.

$$T(\alpha, \beta) = \frac{\sum_{i=1}^{N} p_{0i}g_{0i}}{\sum_{i=1}^{N} p_{0i}g_{0i} + \alpha \sum_{i=1}^{N} p_{0i}g_{1i} + \beta \sum_{i=1}^{N} p_{1i}g_{0i}},$$
(1)

где p_{ij} – это выходное значение Softmax слоя предсказания, p_{0i} – вероятность, что воксель і является фоном, а p_{1i} – вероятность, что воксель і является объектом. Также, $g_{0i} = 1$ и $g_{0i} = 0$ для вокселей переднего и заднего плана соответственно, и наоборот для g_{1i} . Параметры α и β регулируют чувствительность функции потерь к разным типам ошибок. Во время обучения использовалось уменьшение суммы (sum reduction), что означает, что потери суммировались по всем примерам в батче, а не усреднялись.

Для обучения модели Swin UNETR использовались предварительно обученные веса [8], полученные в результате самонастраиваемого предобучения (selfsupervised pre-training) энкодера Swin UNETR (3D Swin Transformer) на выборке из 5050 КТ-сканов из общедоступных датасетов.

С. Оценка качества

Качество сегментации оценивалось с использованием метрик, предложенных в конкурсе MICCAI 2024 Tooth Fairy 2: коэффициент сходства Dice (DSC) и расстояние Хаусдорфа 95-го перцентиля (HD95). Эти метрики позволяют количественно оценить точность и согласованность предсказанных масок с эталонной разметкой из датасета ToothFairy2.

Коэффициент сходства Dice (DSC) используется для оценки степени перекрытия предсказанной модели сегментации и эталонной разметки. Рассчитывается по следующей формуле:

$$DSC = \frac{2TP}{2TP + FP + FN},$$
 (2)

где значение DSC варьируется от 0 до 1, где 1 означает полное совпадение предсказанной и эталонной масок.

В качестве второго показателя качества используется 95-й процентиль расстояния Хаусдорфа (HD95). Эта широко применяется метрика в медицинской сегментации изображений, так как измеряет расстояние между двумя наборами точек, что особенно важно в клинических приложениях. В данной работе 95-й перцентиль, что использовался позволяет исключить влияние выбросов.

Значения DSC и HD95 рассчитываются для каждого класса отдельно. Средние значения по всем классам используются для общей оценки точности сегментации.

III. РЕЗУЛЬТАТЫ



Рис. 2. 2D и 3D-представление слева направо: исходное изображение, метки, предсказание модели Unet и предсказание модели Swin-UNETR

Место	Dice Нижняя Челюсть	HD95 Нижняя Челюсть	Dice Верхняя Челюсть	HD95 Верхняя Челюсть	Dice Нижние Зубы	HD95 Нижние Зубы	Dice Верхние Зубы	НD95 Верхние Зубы
1st	0,9886	NA	0,9624	31,8114	0,9905	0,4908	0,9736	10,5555
2nd	0,9899	NA	0,9661	1,309	0,9922	0,4908	0,94	30,5749
4th	0,9847	NA	0,9166	34,0577	0,9873	0,5987	0,953	11,0443
5th	0,9849	NA	0,9164	33,8867	0,9516	20,8298	0,9445	30,565
6th	0,9835	NA	0,9255	5,3236	0,9815	0,6715	0,9398	30,5715
Swin UNETR	0,9668	1,8308	0,8855	1,8207	0,9194	1,4207	0,8727	2,2343
3D Unet	0,9497	2,0602	0,8219	2,1635	0,8782	2,2445	0,8459	2,9424

ТАБЛИЦА I.	ПОКАЗАТЕЛИ DSC И HD95: РЕЗУЛЬТАТЫ ПОБЕДИТЕЛЕЙ MICCAI 2024 И РЕЗУЛЬТАТЫ НАШИХ МОДЕЛЕЙ
------------	--

Результаты оценки качества моделей Swin UNETR и 3D UNet по метрикам коэффициента сходства Dice (DSC) и 95-го процентиля расстояния Хаусдорфа (HD95), рассчитанные как в среднем для всех классов, так и отдельно для каждой анатомической структуры, приведены в табл. 1. Оценка производительности проводилась на тестовой выборке.

Сравнение средних значений и показателей по каждому классу демонстрирует, что Swin UNETR превосходит 3D UNet по всем метрикам. Особенно значительное улучшение наблюдается в классах верхней челюсти и верхних зубов, где Swin UNETR достигает более высоких значений DSC и показывает меньшие ошибки локализации (HD95).

Хотя 3D UNet также продемонстрировала высокие результаты, модели с более сложной архитектурой, такие как Swin UNETR, обеспечивают более точное соответствие сегментируемым структурам, что делает Swin UNETR более предпочтительной для задачи сегментации зубных структур.

Анализ качества сегментации различных классов выявил следующие закономерности:

- Лучшая сегментация наблюдается для нижней челюсти. Это объясняется её большим представлением в данных. Значения Dice максимальны, а HD95 минимальны.
- Худшие результаты сегментации наблюдаются для верхней челюсти и верхних зубов. Оба этих класса встречаются в датасете значительно реже.

Для визуальной оценки качества сегментации на Рисунке 2 представлены результаты моделей 3D UNet и Swin UNETR на данных ToothFairy2. Анализ 2D и 3Dвизуализаций предсказаний моделей показывает, что Swin UNETR демонстрирует лучшие результаты по сравнению с 3D Unet. Также можно заметить, что Swin UNETR выявляет те классы, которые 3D Unet не способен обнаружить.

Для оценки конкурентоспособности обученных моделей рассмотрены результаты топ-5 участников [9] конкурса MICCAI 2024 Tooth Fairy 2 Challenge, представленные в табл. 1. Победители MICCAI 2024 демонстрируют более высокие значения DSC, как в среднем, так и по каждому классу. Лучшие модели достигают точности сегментации выше 96 %, что значительно выше, чем у Swin UNETR (91.1 %). Однако по метрике HD95 наши модели превосходят победителей: в большинстве случаев HD95 у Swin UNETR значительно ниже, чем у топовых конкурентов.

IV. ЗАКЛЮЧЕНИЕ

В данной работе проведен анализ и сравнение моделей 3D UNet и Swin UNETR для задачи сегментации зубных и челюстных структур на КЛКТ-изображениях. Результаты обучения показывают, что обе модели эффективно справляются с сегментацией зубных и челюстных структур, однако Swin UNETR превосходит 3D UNet по всем оцениваемым метрикам, особенно в сегментации верхней челюсти и верхних зубов. Полученные результаты подчеркивают важность качественной предобработки включая ланных. выравнивание нормализацию интенсивности И разрешения. пространственного Кроме того. использование функции потерь Tversky и оптимизатора AdamW также способствовало улучшению качества моделей.

Анализируя результаты, можно заметить, что качество сегментации сильно зависит от представления структур в датасете: сегментация нижней челюсти и нижних зубов оказалась более точной, чем сегментация верхней челюсти и верхних зубов. Это связано с неравномерным распределением данных в датасете ToothFairy2. Сравнение с победителями конкурса MICCAI 2024 показало, что наши модели уступают лидерам по метрике DSC, но превосходят их по HD95.

Дальнейшее повышение качества сегментации возможно за счет увеличения размера обучающей выборки и оптимизации методов аугментации данных.

Список литературы

- Çiçek Ö., Abdulkadir A., Lienkamp S. S., Brox T., Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016: Proceedings. Part II. Springer, 2016. 424–432 p.
- [2] Hatamizadeh A., Nath V., Tang Y., Yang D., Roth H. R., Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images // International MICCAI Brainlesion Workshop. Springer, 2021. 272–284 p.
- [3] Cipriano M., Allegretti S., Bolelli F., Pollastri F., Grana C. Improving Segmentation of the Inferior Alveolar Nerve through Deep Label Propagation // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022. 21105–21114 p.

- [4] Cipriano M., Allegretti S., Bolelli F., Di Bartolomeo M., Pollastri F., Pellacani A., Minafra P., Anesi A., Grana C. Deep Segmentation of the Mandibular Canal: A New 3D Annotated Dataset of CBCT Volumes // IEEE Access. 2022. Vol. 10. 11500–11510 p.
- [5] Chougule V., Mulay A., Ahuja B. Clinical Case Study: Spine Modeling for Minimum Invasive Spine Surgeries (MISS) Using Rapid Prototyping // Bone (CT). 2018. Vol. 226. 3071 p.
- [6] Choi H. J., Lee H. J., Kang S. G. The Clinical Significance of Hounsfield Number of Metallic and Non-Metallic Foreign Bodies in the Soft Tissue // Soonchunhyang Medical Science. 2010. 226–230 p.
- [7] Loshchilov I., Hutter F. Decoupled Weight Decay Regularization // International Conference on Learning Representations (ICLR), 2019.
- [8] Tang Y., Yang D., Li W., Roth H. R., Landman B., Xu D., Nath V., Hatamizadeh A. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. 20730–20740 p.
- [9] Bolelli F., Lumetti L., Vinayahalingam S., Di Bartolomeo M., van Nistelrooij N., Marchesini K., Anesi A., Grana C. ToothFairy2 Challenge: Multi-Structure Segmentation in CBCT Volumes, 2024.