МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М. В. ЛОМОНОСОВА
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ
КАФЕДРА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Айрапетьянц Каринэ Арсеновна

Разработка методов искусственного интеллекта для дифференциальной диагностик глиальных опухолей по данным динамических позитронно-эмиссионных исследований.

АННОТАЦИИ СТАТЕЙ

Научный руководитель:
Малоян Нарек Гагикович

Москва, 2021

Оглавление

1. Deep reinforcement learning-based image classification achieves perfect testing set accuracy for MRI brain tumors with a training set of only 30 images

Ссылка

https://arxiv.org/abs/2102.02895

Введение

Задачи классификации и сегментации являются основной областью применения искусственного интеллекта в радиологии и попадают в категорию задач, решаемых с помощью метода глубокого обучения с учителем. Однако, применение данного метода в медицине имеет свои ограничения: для реализации требуется большое количество размеченных данных квалифицированными специалистами; обобщающая способность падает, когда требуется сделать предсказание на изображениях со сканеров, отличных от тех, на которых обучалась сеть, либо на изображениях с других медицинских учреждений. Немаловажен и феномен „черного ящика“, при котором не до конца понятно, как получены результаты и доверие к методу среди специалистов и пациентов падает.

Основная идея

В своих предыдущих работах авторы статьи предложили метод обучения с подкреплением в радиологии и показали, что с его помощью решаются задачи локализации и сегментации пораженной области на изображении. В данной работе для отыскания оптимальной стратегии авторы использовали Марковский процесс принятия решений. Таким образом, черно-белое изображение перекрывается красным, отображая начальное состояние. Далее, на каждом шаге агент совершает действие - предсказание, результатом которого является 0 - нормальное изображение, и 1 - изображение, содержащее опухоль. Если предсказан верный класс, то в следующем состоянии изображение преобразуется в черно-белое с зеленым перекрытием. В противном

случае, изображение остается красным либо с зеленого меняется на красный. За правильное предсказание агент награждается в размере +1, а за неверное штрафуется в размере -1. Основная цель обучения с подкреплением - достичь максимальной суммарной награды. Тренировка основывается на сочетании глубокой Q-сети (DQN) с TD(0) Q-обучением. Также, для сравнения классификации, основанной на глубоком обучении с учителем и обучении с подкреплением, авторы обучили сверточную нейронную сеть с архитектурой, схожей с архитектурой DQN на таком же наборе тренировочных данных.

Данные

В качестве данных для обучения были выбраны 60 двумерных срезов трехмерных изображений из датасета BraTS 2020 Challenge tumor database. Все изображения были сняты в режиме T1-ВИ после введения контрастного вещества. 30 из них были размечены специалистами как нормальные, а оставшиеся 30 - содержат злокачественные глиомы. Далее, 30 изображений из 60 были выбраны в качестве обучающего множества и 30 в качестве тренировочного (в каждом по 15 нормальных и злокачественных изображений).

Результаты

Рассматривая точность на обучающем множестве в зависимости от времени обучения с подкреплением можно видеть постепенное повышение обобщающей способности, а точность в 100% достигается через 200 эпизодов обучения. В то же время, сверточная нейронная сеть быстро переобучается на таком маленьком наборе данных и точность предсказания достигает лишь 57%.

Заключение

Учитывая все вышеизложенное и тот факт, что зачастую медицинские наборы данных очень малы, а в данном исследовании „традиционная“ нейронная сеть быстро переобучается на маленьком датасете, авторы показали, что обучение с подкреплением показывает значительное преимущество в задаче классификации, сегментации и локализации (эти факты показаны в

предыдущих исследованиях). Однако, использование двумерного среза вместо целого трехмерного изображения является ограничением, ровно как и то, что предсказывалось только два класса.

2. Localization of Critical Findings in Chest X-Ray without Local Annotations Using Multi-Instance Learning

Ссылка

https://arxiv.org/abs/2001.08817

Введение

Автоматическое детектирование серьезных нарушений легких, таких как пневмоторакс, пневмония и отек по рентгеновским изображениям широко исследуется на данный момент. Для решения задачи классификации изображений обычно используются сверточные нейронные сети, однако остается открытым вопрос интерпретации и объяснения результата предсказания и в медицине он особенно важен в разрезе доверия данному методу. Существуют методы, которые строят тепловую карту изображения в пикселях, указывая регионы, которые участвовали в прогнозировании. Однако, они применяются уже после классификации и тепловые карты генерируются фильтрами с низким разрешением, а после проецируются обратно до размеров исходного изображения, что может плохо сказаться на локализации в случае медицинских изображений, которые обычно имеют высокое разрешение. Также, широко используются алгоритмы обнаружения объектов и сегментации, выделяющие границу региона, однако они требуют попиксельную маркировку (маску), на основе которой будет строиться предсказание. Здесь и находится ограничение - разметка медицинских изображений требует наличие квалифицированных специалистов и является трудоемкой и времязатратной задачей.

Основная идея

В данной работе авторы ставят задачей устранить два вышеперечисленных недостатка - низкую интерпретируемость и необходимость локальной аннотации на основе технологии многовариантного обучения (multi-instance learning (MIL)). При таком подходе данные разделяют на множество частей, называемых *вариантами*, которые совместно анализируются для понимания

того, какой вклад они внесли в предсказание метки класса. В данном случае в качестве вариантов выступают части изображения. Цель данного подхода в бинарной классификации рентгеновских изображений - предсказать метку для каждой части (0 или 1), что является обучением со слабой разметкой, так как известна метка для целого изображения, но не для его части. Очевидно, что в изображениях, не содержащих аномалий, все части также не будут их содержать, а в изображениях с патологическими нарушениями, хотя бы одна часть должна быть помечена как дефектная. Таким образом, для классификации, части подаются на вход сверточной нейронной сети, которая на выходе выдает вероятность содержания аномалии от 0 до 1 и, так как части не имеют разметки, то в процессе обучения MIL использует механизм, выявляющий связь между меткой, присущей всему изображению и предсказываемой меткой. Полученные значения меток в негативных частях (тех, которые не содержат аномалий), подавляются до 0, а в позитивных - вытягиваются к 1, таким образом происходит непрерывная классификация позитивных и негативных изображений и определяются части, ответственные за принятие сетью решения.

Данные

Авторы использовали три датасета - UWMC (пневмоторакс), RSNA/Kaggle (пневмония), MIMIC-CXR (отек легких).

Результаты

В результате экспериментов, используя сеть VGG16, авторы достигли точности в 0.89, 0.84 и 0.82 для пневмоторакса, пневмонии и отека легких соответственно. Также, в качестве дополнительного эксперимента по классификации пневмоторакса из датасета UWMC было произведено сравнение метода MIL с двумя другими методами классификации на основе модифицированной сети ResNet50 и полносвязной сети (FCN). Получены следующие результаты: 0.96 (ResNet50), 0.93 (MIL), 0.92 (FCN). Также, в статье авторы приводят визуализацию результата работы метода MIL с соответствующей интерпретацией - части, которые содержат патологии с вероятностью, близ-

кой к 1 толстые и обведены темно-красным, части со средним показателем - светло-красные и светлее с меткой близкой к нулю.

Заключение

В данной статье описан и применен метод многовариантного обучения MIL, который одновременно классифицирует изображения и позволяет локализовать патологии без специальной разметки, имея только разделение на классы целых изображений, что позволяет понимать, какая часть изображения внесла больший вклад в результат работы сети. Авторы утверждают, что данный метод масштабируем - его можно использовать для нахождения любого числа патологий на изображении.

3. On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task

Ссылка

https://arxiv.org/abs/1707.01992

Введение

На сегодняшний день большинство исследований в области обработки медицинских изображений нейронными сетями проводятся с использованием двумерных изображений, в то время как трехмерные изображения являются более информативными. Однако, анализ и обработка трехмерных изображений сопряжены с вычислительными трудностями и, в то время как разработка эффективной и рабочей нейронной сети трехмерной архитектуры представляет большой интерес, ее проектирование остается сложной задачей. Целью данной работы является разработка компактной сетевой архитектуры высокого разрешения для сегментации структур объемных изображений. Также демонстрируется возможность оценки неопределенности на воксельном уровне с помощью метода Монте-Карло на предлагаемой сети во время тестирования.

Основная идея

Нейронная сеть, предложенная в данной статье состоит из 20 сверточных слоев. В первых семи слоях ядро свертки имеет размерность 3x3x3, они необходимы для локализации низкоуровневых объектов изображения - краев и углов. В последующих сверточных слоях размерность ядра увеличивается в два или четыре раза - они необходимы для локализации более значительных фрагментов. Каждые два сверточных слоя группируются в residual block. Внутри каждого такого блока каждый сверточный слой связан поэлементно с ReLU и со слоем batch нормализации. Сеть обучается от начала и до конца, входные данные предобрабатываются (стандартизация и аугментация).

Данные

Для демонстрации возможности обучения на сложных трехмерных изображениях были выбраны 543 МРТ - изображения, снятых в режиме Т1-ВИ из датасета ADNI.

Результаты

Используя сеть предложенной архитектуры (HC-default) авторы сравнивают качество ее предсказания с предсказаниями, полученными от трех вариаций данной сети: (1) HC-default без residual blocks и с логарифмической функцией потерь (NoRes-entropy); (2) HC-default без residual blocks с коэффициентом Серенсена в качестве функции потерь (NoRes-dice); (3) HC-default с дополнительным dropout слоем (HC-dropout). К тому же, был проведен сравнительный анализ с тремя уже существующими сетями для трехмерной сегментации - 3D U-net, V-net, Deepmedic. Было установлено, что тренировка сетей с логарифмической функцией потерь ведет к низким результатам сегментации. Поэтому, в качестве функции потерь был выбран коэффициент Серенсена (Dice-coefficient). С относительно маленьким количеством параметров, HC-default и HC-dropout превосходят вышеперечисленные модели по метрике Серенсена. Это означает, что предложенная сеть лучше справляется с поставленной задачей.

Заключение

В данной работе была продемонстрирована архитектура трехмерной сверточной сети, которая включает в себя слои свертки и residual block'и. Данная сеть концептуально проще и имеет меньшее количество параметров, чем уже существующие сети для обработки трехмерных изображений. Более того, по сравнению с ними она показала лучшие результаты в задачах сегментации и парцелляции головного мозга.

4. Evidential segmentation of 3D PET/CT images

Ссылка

https://arxiv.org/abs/2104.13293

Введение

Позитронно-эмисионная томография и компьютерная томография (PET/CT) - два мощных инструмента в диагностике онкологических заболеваний. PET - изображения широко используются локализации и сегментации лимфом благодаря чувствительности и метаболической активности опухоли. Из-за низкого разрешения и контрастности, результаты сегментации PET/CT изображений с помощью нейронных сетей не вызывают доверия. В данной работе авторы предлагают модель для сегментации диффузной B-крупноклеточной лимфомы из трехмерных PET/CT изображений, основанной на теории Демпстера-Шафера (BF) *(Теория Демпстера — Шафера математическая теория очевидностей (свидетельств), основанная на функции доверия (belief functions) и функции правдоподобия (plausible reasoning), которые используются, чтобы скомбинировать отдельные части информации (свидетельства) для вычисления вероятности события)* и глубоком обучении.

Основная идея

Архитектура предложенной нейронной сети (ES-Unet) основывается модуле UNet для извлечения признаков (encoder-decoder) и модуле сегментации очевидностей (evidential segmentation - ES), которая основывается на одели evidential neural network и подходе, предложенных в ранних работах, для количественной оценки неопределенности относительно каждого вокселя решения с некоторой степенью доверия по функции массы Демпстера-Шаффера. Основная идея модуля ES - присвоить массу каждому из K классов и всему множеству классов Ω, основываясь на расстоянии между вектором признаков каждого вокселя и центрами прототипа I. В процессе обучения сети минимизируется двусоставная функция потерь, позволяющая увеличить точность по мере Серенсена (Dice score) и уменьшить неопределенность.

Данные

Датасет состоит из 173 изображений, полученных после исследования пациентов, у которых была диагностирована B-крупноклеточная лимфома.

Результаты

Предложенная модель превосходит базовую модель UNet, так же как и другие модели (nnUnet, VNet, SegResNet). В частности, ES-Unet превосходит лучшую модель SegResNet на 1.9%, 2.4%, 1.4% по Dice score, Sensitivity и F1 score соответственно.

Заключение

Был разработан фреймворк ES-Unet для сегментации лимфом по трехмерным PET/CT изображениям с количественной оценкой неопределенности. Предложенная архитектура основывается на совмещении модели Unet и модуля ES. Обучение выполняется путем мнимизации двусоставной функции потерь. P азработанная модель справляется с поставленной задачей и превосходит по качеству предсказания уже существующие модели (Unet,nnUnet, VNet, SegResNet).

5. Deep Kernel Representation for Image Reconstruction in PET

Ссылка

https://arxiv.org/abs/2110.01174

Введение

Реконструкция ПЭТ изображения является сложной задачей из-за низкого разрешения и высокого шума в данных. Среди разных методов реконструкции ПЭТ изображений, ядровые методы (kernel methods) шают проблему шума путем интеграции в изображение дополнительной информации. Дополнительную информацию можно получить из составных изображений динамического ПЭТ сканирования или из анатомических изображений (например, МРТ при совместном исследовании ПЭТ/МРТ). В существующих ядерных методах ядро обычно строится при помощи эмпирического подбора векторов признаков и ручного выбора параметров, связанных с методом.

Основная идея

В данной работе описывается эквивалентность между представлением ядра в ядерном методе и обучаемой нейросетевой моделью. Основываясь на этой связи, далее предлагается метод „глубокого ядра“, который изучает обученные компоненты нейросетевой модели на доступных снимках, чтобы достичь автоматизации обучения, основанной на данных для оптимизированной ядерной модели. Далее, обученная ядерная модель применяется для реконструкции ПЭТ изображений и ожидается, что данный метод будет превосходить другие ядерные методы, основанные на эмпирических заключениях. Описываемый метод имеет уникальное преимущество - после обучения модели неизвестный ядерные коэффициенты остаются линейными и легко восстанавливаются по ПЭТ данным. К тому же, для этого не требуется большой набор данных.

Данные

В качестве данных были использованы снимки динамического ПЭТ сканирования с помощью сканера GE 690, данные пациентов из UC Davis Medical Center со сканера GE Discovery ST PET/CT.

Результаты

В работе наряду с предложенным методом приводятся существующие методы восстановления изображений, а далее с их помощью моделируются данные. Смоделированные данные были восстановлены с помощью четырех различных методов: (1) стандартная ML-EM реконструкция; (2) существующий ядерный метод без обучения; (3) предлагаемый метод глубокого ядра с онлайн обучением для извлечения признаков; (4) метод реконструкции DIP. Так, к примеру изображения, восстановленные с помощью ML-EM метода получились очень шумными, DIP метод привел к сильному сглаживанию, а восстановленные изображения с помощью описанного метода показали более четкие контуры и более низкий шум в левом и правом желудочке и миокарде.

Заключение

Таким образом, авторы разработали новый ядерный метод для реконструкции ПЭТ изображений, который показывает более оптимальное обучение ядра, чем в эмпирических методах. Результаты компьютерного моделирования и реального набора данных показывают, что предложенный метод превосходит существующие ядерные и нейросетевые методы реконструкции ПЭТ изображений.

6. Multimodal PET/CT Tumour Segmentation and Prediction of Progression-Free Survival using a Full-Scale UNet with Attention

Ссылка

https://arxiv.org/abs/2111.03848

Введение

Опухоли головного мозга и шеи являются пятыми по распространенности онкологическими заболеваниями в мире. Сегментация новообразований в области головы и шеи и предсказание исхода болезни важны для диагностики, лечения и мониторинга заболевания.Ручная сегментация новообразований, локализованнных в голове и шее является более сложной задачей по сравнению с другими частями тела, потому что опухоль показывает похожие значения интенсивности с соседними тканями, и человеческому глазу трудно отделить больную ткань от здоровой по КТ-изображению. На данный момент комбинация ПЭТ/КТ играет ключевую роль в диагностике новообразований. В данной работе решается задача сегментации опхолей головы и шеи с помощью сверточной нейронной сети, а также задача предсказания выживаемости пациентов с помощью регресионной модели.

Основная идея

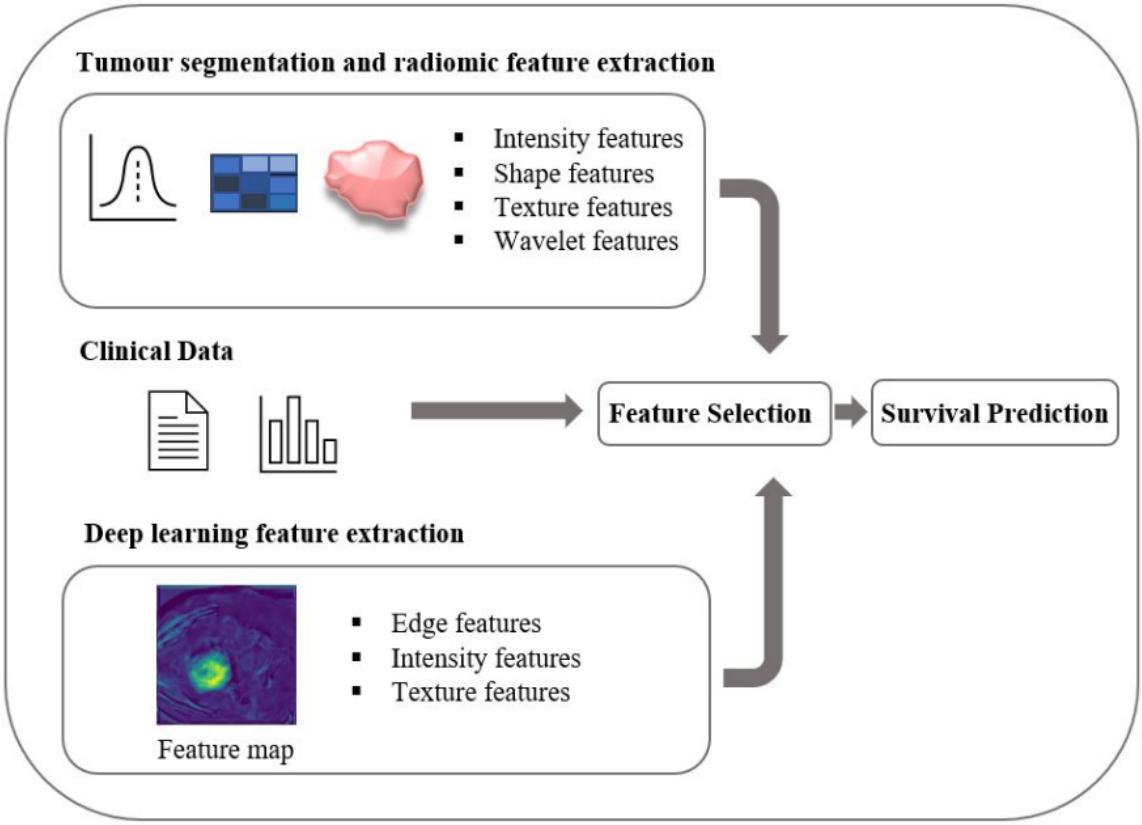
Авторы предлагают производить сегментацию опухолей головы и шеи по ПЭТ/КТ изображениям, используя полномасштабную сеть архитектуры 3D UNet3++ с механизмом, имитирующим когнитивное внимание (attention mechanism). Предложенная модель, NormResSE-UNet3+ была обучена с гибридной функцией потерь, состоящей из Log Cosh Dice и Focal loss. Далее, предсказанные карты сегментации дополнительно уточняются с помощью механизма постпроцессинга - Conditional Random Fields, чтобы уменьшить число ложноположительных ответов и улучшить сегментацию границы опухоли. Для решения задачи предсказания выживаемости предлагается регресионная модель CoxPH относительной опасности, использующая комбинацию клини-

ческих, radiomics (признаки, полученные из медицинских изображений с помощью определенных методов) признаков, а также признаков, полученных при глубоком обучении на ПЭТ/КТ-изображениях.

Предобработка данных

Для задачи сегментации была использована трилинейная интерполяция (trilinear interpolation) ПЭТ и КТ-изображений. Интесивность ПЭТ (заданная в SUV) была нормализована с помощью Z-score, а интенсивность КТ (заданная по шкале Хаунсфилда), приведена к [-1,1].

Данные для предсказания выживаемости были обработаны с учетом пропущенных значений. Каждый пропущенный признак - это функция от существующих признаков. Пропущенные признаки восстанавливаются итеративно с помощью Lasso регрессии и 5-fold кросс-валидации на клинических, радиомических признаках и признаках, полученных из 3D-UNet.

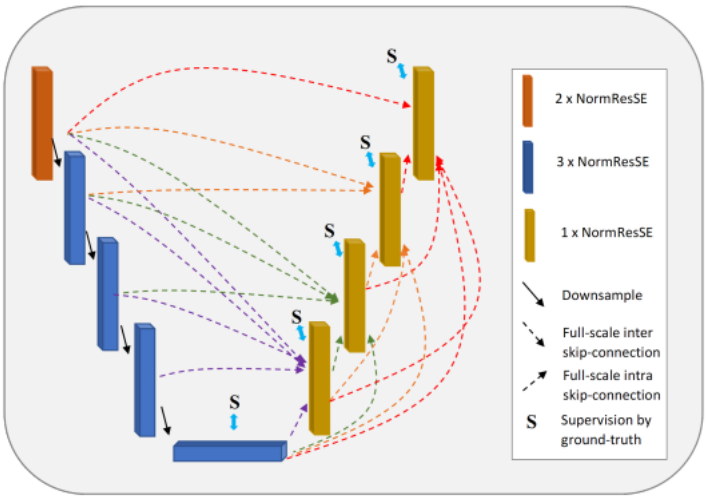


Пайплайн для задачи предсказания выживаемости. Состоит из трех шагов: сбор клинических данных, изображений и препроцессинга. Затем, выбираются извлеченные признаки и производится предсказание выживаемости.

Модель для сегментации

Архитектура предложенной сети NormResSE-UNet3+:

- На вход подается тензор, размерности 2x144x144x144, состоящий из конкатенации ПЭТ и КТ изображений.
- Энкодер состоит из residual squeeze-and-excitation блоков, первый блок из которых содержит 24 фильтра. Размерность выхода энкодера - 384x3x3x3
- Путь декодирования состоит из полномасштабных соединений и модуля, содержащего правильную разметку изорбражений (ground truth).
- У декодера одноканальный выход размерности 1x144x144x144



Архитектура NormResSE-UNet3+

Данные

Данные были предоставленны организаторами соревнования HECTOR - MICCAI. Всего тренировочных примеров - 224 из 5 центров: CHGJ, CHMR, CHUS,CHUP,CHUM.

Результаты

Было обучено несколько моделей нейронных сетей для задачи сегментации опухолей головы и шеи. Результаты сведены в единую таблицу.

Cross-validation fold	NormResSE-UNet3+		NormResSE-UNet3+ + CRF	
	DSC	HD	DSC	HD95
Fold 1	0.792	3.18	0.822	3.11
Fold 2	0.693	3.43	0.702	3.41
Fold 3	0.728	3.32	0.749	3.29
Fold 4	0.736	3.31	0.738	3.30
Fold 5	0.742	3.29	0.756	3.28
Ensemble	0.738	3.30	0.753	3.28

Количественные результаты сегментации

Результаты предсказания выживаемости. Предложенная регрессионная модель CoxPH показала лучший результат:

Survival Models	C-index
CoxPH Regression (clinical)	0.70
CoxPH Regression (clinical + PET radiomics)	0.67
CoxPH Regression (clinical + CT radiomics)	0.68
CoxPH Regression (clinical + PET/CT radiomics)	0.72
CoxPH Regression (clinical + deep learning features)	0.76
CoxPH Regression (clinical + CT radiomics + deep learning features)	**0.82**
Random Survival Regression (clinical)	0.59
Random Survival Regression (clinical + PET radiomics)	0.60
Random Survival Regression (clinical + CT radiomics)	0.61
Random Survival Regression (clinical + PET/CT radiomics)	0.59
Random Survival Regression (clinical + CT radiomics + deep learning features)	0.58
DeepSurv (clinical)	0.60
DeepSurv (clinical + PET radiomics)	0.68
DeepSurv (clinical + CT radiomics)	0.69
DeepSurv (clinical + PET/CT radiomics)	0.73
DeepSurv (clinical + PET/CT radiomics + deep learning features	0.65

Количественные результаты предсказания выживаемости.

Заключение

Авторы предложили модель NormResSE-UNet3+ для сегментации опухолей головы и шеи по мультимодальным ПЭТ/КТ изображениям, в основе которой лежит архитектура UNet3+ с комбинированными слоями сжатия (squeeze-and-excitation), чтобы использовать возможности модели непрерывно фокусироваться на релевантных областях интереса, что помогло улучшить точность сегментации.

7. Glioma Segmentation with Cascaded Unet

Ссылка

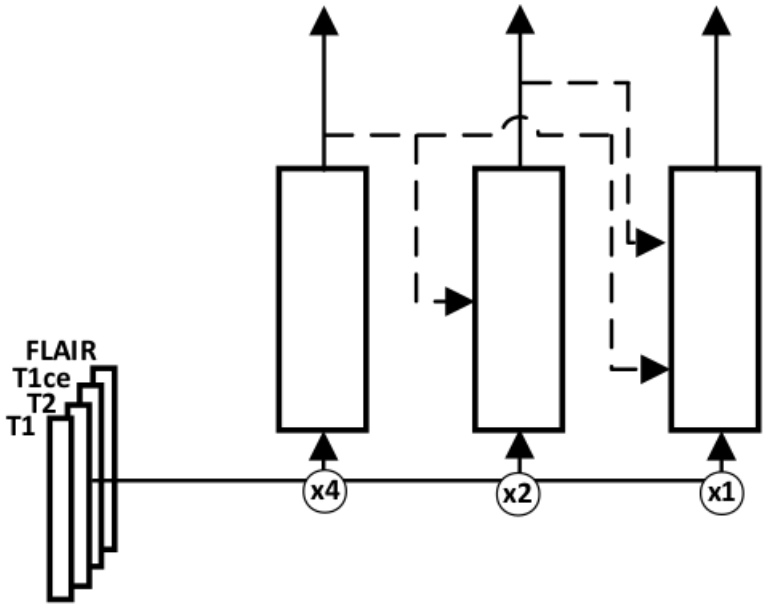
https://arxiv.org/abs/1810.04008

Введение

Точная сегментация и реконструкция медицинских 3D изображений способны дать больше необходимой информации о прогрессировании заболевания и позволяют терапевту спланировать успешный курс лечения для больного. В данной работе авторы представляют каскадный вариант популярной сети UNet, который итеративно улучшает результаты сегментации, полученные на предыдущих шагах.

Основная идея

Предложенный метод может быть представлен как цепь классификаторов C_i, одинаковой топологии F, у каждого из которых свой собственный набор параметров W_i для оптимизации в течение обучения. Результат вычисления i-го шага представляется следующим образом: $Y_i = F(X_i, Y_{i-1}, Y_{i-2}, W_i)$. Каждый из базовых блоков C_i - это сеть архитектуры UNet, измененная для задачи сегментации глиом. В сравнении со стандартной архитектурой UNet, в предложенной модели используется несколько энкодеров, которые раздельно обрабатывают входные данные. Также, предложен метод объединения их выхода: в UNet i-й выход декодера зависит от выхода соответствующего энкодера и выхода предыдущего декодера - $d_i^t = f(e_i^t, d_{i-1}^t)$. Раскрывая первую свертку f, получаем - $d_i^t = g(W_{i,e}^t e_i^t + W_i, d^t d_{i-1}^t)$. Далее предлагается объединить контекст, полученный на более низких слоях, добавляя соответствующий выход y^t, поэтому $d_i^t = g(W_{i,e}^t e_i^t + W_{i,d}^t d_{i-1}^t + W_{i,y}^t y^{t-i})$.



Схематическое представление метода, описанного в статье. T1, T2, T1ce, FLAIR - входные модальности МРТ-изображения, x4,x2 - понижающий фактор входа сети. Пунткирные линии - соединения между блоками C_i.

Данные

BraTS 2018

Результаты

Результат сегментации оценивался по метрике Dice, отдельно вычисленной для следующих частей опухоли: WT (whole tumor) - вся опухоль, ET (enchancing tumor) - усиливающаяся часть опухоли и TC (tumor core) - ядро опухоли.

Method	WT	ET	TC
UNet	0.901	0.767	0.797
ME UNet	0.904	0.763	0.823
C ME UNet	0.906	0.772	0.836

Результаты без аугментации выходов

Method	WT	ET	TC
UNet	0.901	0.779	0.837
ME UNet	0.907	0.784	0.827
C ME UNet	0.908	0.784	0.844

Результаты с аугментацией выходов

Заключение

В данной работе был предложен алгоритм автоматической сегментации опухолей головного мозга по МРТ-зображениям, который решает также проблему мультимодального входа и показывает хорошие резльтаты по сравнению с моделью UNet.

8. CaraNet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects

Ссылка

`https://arxiv.org/abs/2108.07368`

Введение

На данный момент разработано достаточное количество архитектур сверточных сетей для решения задачи сегментации медицинских, которые показывают хорошие результаты. Однако, только малая часть исследований учитывает размер интересующих объектов на изображении и поэтому многие модели показывают плохой результат при сегментации объектов малого размера, что сильно влияет при диагностике заболевания. В данной работе предлагается нейросетевая модель Context Axial Reserve Attention Network (CaraNet), которая способна улучшить результаты сегментации малых объектов по сравнению с уже существующими моделями.

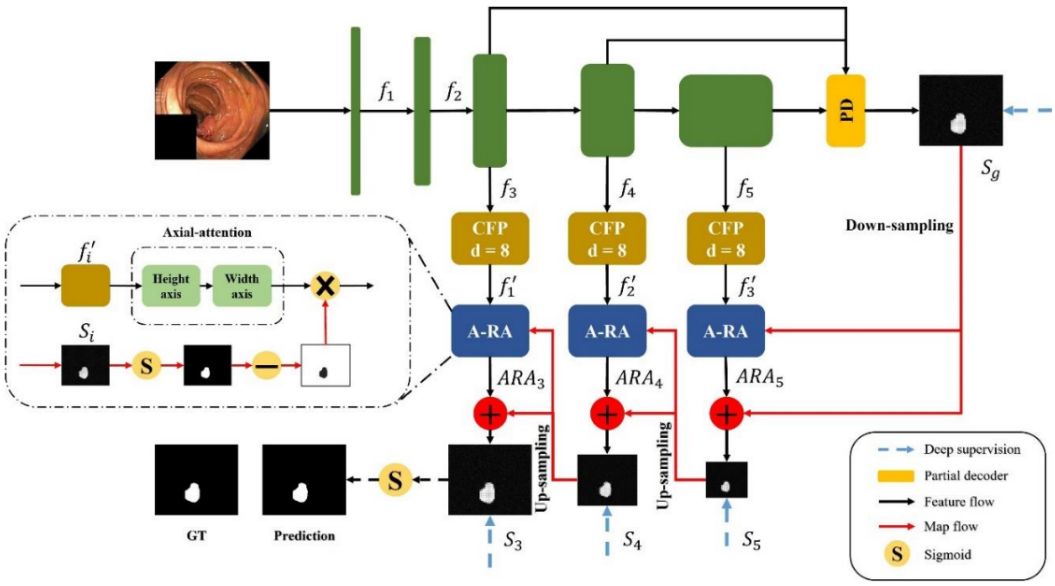
Основная идея

В архитектуре CaraNet используется параллельный частичный декодер (parallel partial decoder) для генерации высокоуровневой семантической карты и набор операций (Context and Axial Reverse Attention) для идентификации глобальных и локальных признаков.
Модули CaraNet:

- *Parallel partial decoder.* Эксперименты показали, что низкоуровненвые признаки вычислительно более сложны и вносят меньший вклад в улучшение результатов сегментации. Поэтому, авторы используют параллельный частичный декодер $p_d(\cdot)$ для извлечения высокоуровневых признаков $PD = p_d(f_3, f_4, f_5)$ и получения глобальной карты S_g из частичного декодера.

- *Context module.* Чтобы получить контекстную информацию из высокоуровневых признаков, применяется модуль CFP (Channel-wise Feature Pyramid) со степенью растяжения (dilation rate) $d = 8$. После контекстного модуля можно получить многомасштабные высокоуровневые признаки $\{f_3^{'}, f_4^{'}, f_5^{'}\}$.

- *Axial reverse attention.* Данный модуль состоит из двух частей: маршрут по оси (axial attention route) и обратный маршрут (reverse attention route). Глобальная карта S_g может поймать только приблизительное расположение тканей без структурных деталей, поэтому структурированный регион тканей постепенно добывается стиранием переднего плана объекта с помощью операции reverse attention: $R_i = 1 - Sigmoid(S_i)$. По другому маршруту применяется axial attention. Здесь сеть может извлечь глобальные зависимости и локальное представление совершая вычисления по горизонтальной и вертикальной оси.



Архитектура CaraNet, состоящая из трех контекстных модулей (CFP) и модулей axial reverse attention (A-RA). ’S’ - сигмоида.

Данные

- BraTS 2018 - опухоли ГМ

- Kvasir-SEG, CVC-ColonDB, CVC-ClinicDB, CVC- 300 and ETIS-LaribPolypDB - полипы

Результаты

По сегментации полипов на основе пяти датасетов, предложенная модель CaraNet не только превосходит сравниваемые модели по общей производительности, но и на примерах с полипами малых размеров.

	Methods	mean Dice	mean IoU	F_β^w	S_α	E_ϕ^{max}	MAE
Kvasir	UNet	0.818	0.746	0.794	0.858	0.893	0.055
	UNet++	0.821	0.743	0.808	0.862	0.910	0.048
	ResUNet-mod	0.791	n/a	n/a	n/a	n/a	n/a
	ResUNet++	0.813	0.793	n/a	n/a	n/a	n/a
	SFA	0.723	0.611	0.670	0.782	0.849	0.075
	PraNet	0.898	0.840	0.885	0.915	0.948	0.030
	CaraNet	**0.918**	**0.865**	**0.909**	**0.929**	**0.968**	**0.023**
CVC-ClinicDB	UNet	0.823	0.755	0.811	0.889	0.954	0.019
	UNet++	0.794	0.729	0.785	0.873	0.931	0.022
	ResUNet-mod	0.779	n/a	n/a	n/a	n/a	n/a
	ResUNet++	0.796	0.796	n/a	n/a	n/a	n/a
	SFA	0.700	0.607	0.647	0.793	0.885	0.042
	PraNet	0.899	0.849	0.896	0.936	0.979	0.009
	CaraNet	**0.936**	**0.887**	**0.931**	**0.954**	**0.991**	**0.007**
Colon DB	UNet	0.512	0.444	0.498	0.712	0.776	0.061
	UNet++	0.483	0.410	0.467	0.691	0.760	0.064
	SFA	0.469	0.347	0.379	0.634	0.765	0.094
	PraNet	0.709	0.640	0.696	0.819	0.869	0.045
	CaraNet	**0.773**	**0.689**	**0.729**	**0.853**	**0.902**	**0.042**
ETIS	UNet	0.398	0.335	0.366	0.684	0.740	0.036
	UNet++	0.401	0.344	0.390	0.683	0.776	0.035
	SFA	0.297	0.217	0.231	0.557	0.633	0.109
	PraNet	0.628	0.567	0.600	0.794	0.841	0.031
	CaraNet	**0.747**	**0.672**	**0.709**	**0.868**	**0.894**	**0.017**
CVC-T	UNet	0.710	0.627	0.684	0.843	0.876	0.022
	UNet++	0.707	0.624	0.687	0.839	0.898	0.018
	SFA	0.467	0.329	0.341	0.640	0.817	0.065
	PraNet	0.871	0.797	0.843	0.925	0.972	0.010
	CaraNet	**0.903**	**0.838**	**0.887**	**0.940**	**0.989**	**0.007**

Количественные результаты на Kvasir, CVC-ClinicDB, CVC-ColonDB, ETIS и CVC-T

Для дальнейшей оценки эффективности сегментации малых объектов с помощью CaraNet был проведен еще один эксперимент, уже с участием опухолей ГМ из датасета BraTS 2018. CaraNet была сравнена с PraNet и показала лучший результат особенно в случаях с очень мылми объектами.

Methods	mean Dice	mean IoU	F_β^w	S_α	E_ϕ^{max}	MAE
CaraNet (Ours)	**0.631**	**0.507**	**0.629**	**0.786**	**0.927**	0.003
PraNet (MICCAI'20)	0.619	0.494	0.606	0.776	0.920	0.003

Количественные результаты на датасете BraTS 2018

Заключение

Была предложена новая нейросетевая модель CaraNet, состоящая из комбинации моделей Axial Reverse Attention и Channel-wise Feature Pyramid, которая показала лучшие результаты сегментации малых объектов на медицинских изображениях по сравнению с уже существующими моделями UNet, UNet++, ResUNet-mod, ResUNet++, SFA, PraNet, что может внести большой вклад при постановке диагноза и выборе дальнейшей тактики лечения.

9. Direct PET Image Reconstruction Incorporating Deep Image Prior and a Forward Projection Model

Ссылка

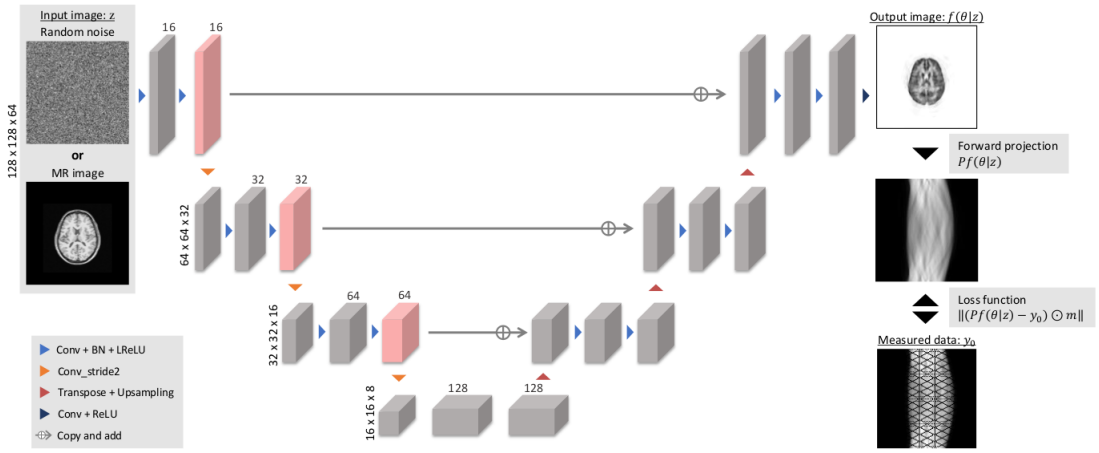
https://arxiv.org/abs/2109.00768

Введение

Для того, чтобы снизить уроень радиации, поглощаемый пациентом при обследовании, проводят ПЭТ-визуализацию с низкими дозами облучения, что в свою очередь негативно влияет на зашумленность изображения. Существуют различные методы по постобработке ПЭТ-изображений, такие как шумоподавление и восстановление для улучшения качества снимков при распознавании малых образований и количественном анализе. В данной работе предлагается метод прямого восстановление ПЭТ-изображения, включающий в себя DIP фреймворк. Алгоритм включает в себя модель прямой проеккции в функции потерь, чтобы достичь прямой реконструкции ПЭТ-изображения из синограм без учителя.

Основная идея

Модель прямой проекции ПЭТ может быть выражена так, что проектируемые данные $y \in \mathbb{R}^{Mx1}$ связаны с пространственным распределением радиоактивного индикатора $x \in \mathbb{R}^{Nx1}$ посредством афинного преобразования $y = Px$, где $P \in \mathbb{R}^{MxN}$ - матрица проекции, которая отражает вклад каждого вокселя в каждую линию ответа (line of response, LOR). Реконструируемое изображение x вычисляется с помощью DIP фреймворка следующим образом: $x = f(\Theta|z)$, где f- это CNN, Θ - веса CNN, а z - предыдущий входной вектор CNN. В данной работе для вычисления реконструированного изображения напрямую предлагается модель прямой проекции P, втсроенная в функцию потерь. Была использована модель, основанная на 3D UNet и адаптированная под текущую задачую



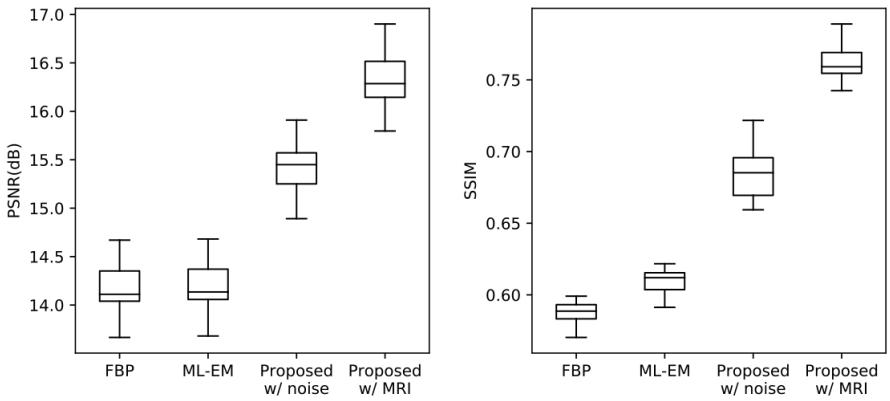
Общий вид предложенной модели прямой реконструкции ПЭТ.

Данные

BrainWeb и созданные данные по методу Монте-Карло.

Результаты

Был проведен сравнительный анализ предложенного метода с методом filtered back projection (FBP) с использованием фильтра Ханна и Гаусса, а также с методом ML-EM. Показано, что описанный в статье метод используя случайный шум и МРТ-изображение точно реконструирует ПЭТ-изображение без сокрытия или потерь информации в сравнении с методами FBP и ML-EM. Основное ограничение данного исследования состоит в том, что оценивался только датасет ПЭТ изображений, полученных с радиоактивной меткой ФДГ.



Количественные результаты реконструированных изображений по метрике PSNR(слева) и SSIM(справа) относительно различных алгоритмов. Линия внутри прямоугольника представляет медиану, верхние и нижние линии прямоугольника - 75-й и 25-й перцентили соответственно. Верхние и нижние „антенны“представляют максимум и минимум соответственно.

Заключение

В сравнении с традиционными алгоритмами FBP и ML-EM, предложенный метод показал лучший результат по метрикам PSNR и SSIM на данных, смоделированных с использованием ФДГ.

10. Virtual PET Images from CT Data Using Deep Convolutional Networks: Initial Results

Ссылка

`https://arxiv.org/abs/1707.09585`

Введение

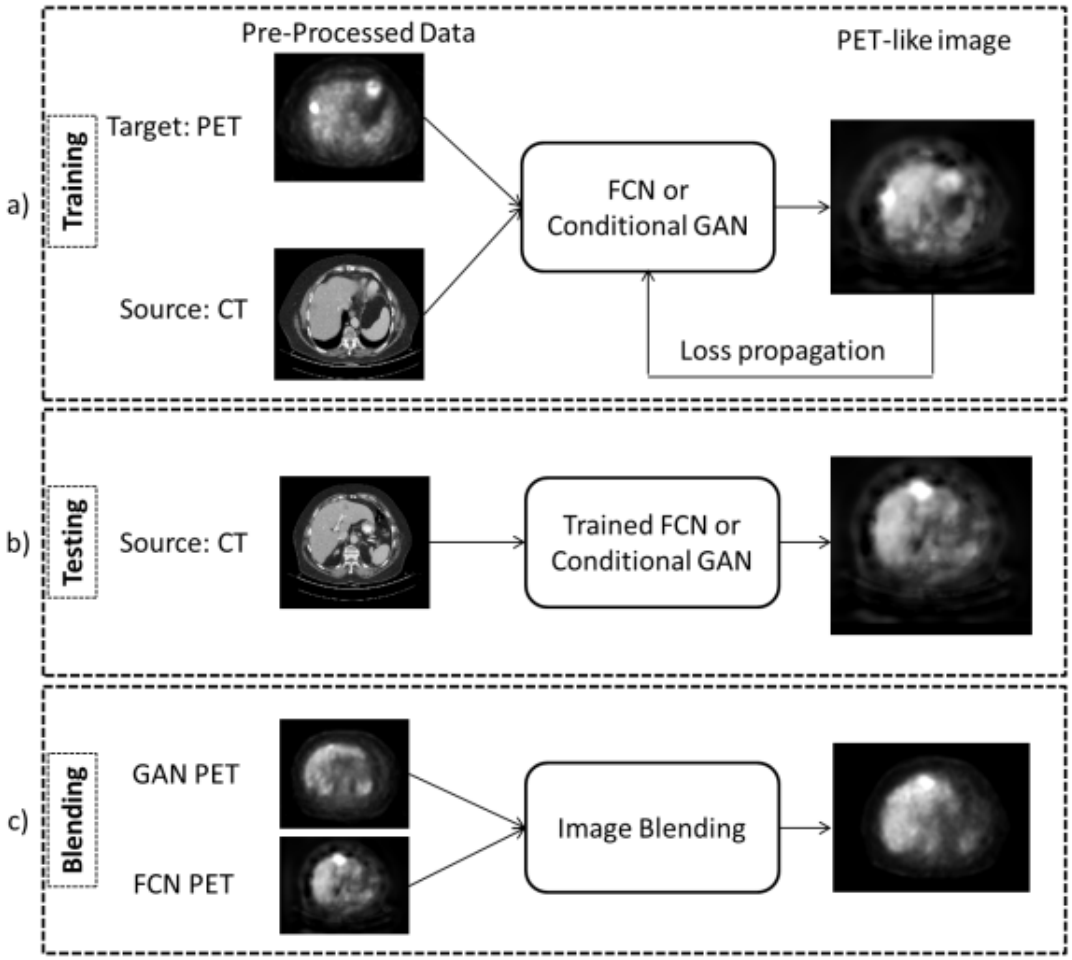
Несмотря на то, что ПЭТ-исследования имеет большое количество положительных сторон, у него так же есть и недостатки - радиоактивный компонент опасен для беременных и кормящих женщин. Также, ПЭТ - сравнительно новый метод, который все еще является дорогостоящим для среднестатистического человека, также возможность получить ПЭТ обследование есть не во всех медицинских центрах. Сложность получения ПЭТ-изображений для последующего лечения послужила возникновению идеи поиска альтернативы - менее дорогостоящего, быстрого и легко в применении ПЭТ-подобного изображения. В данной работе исследуется модуль для создания виртуальных ПЭТ-изображений на основе информации из КТ-изображений.

Основная идея

Фреймворк включает в себя три модуля:

- Тренировочный модуль, который также включает в себя предобработку данных;
- Тестовый модуль, на вход которому подается КТ-изображение для предсказания ПЭТ-подобного изображения на выходе;
- Модуль смешения (blending module), который соединяет выходы FCN и GAN.

FCN и GAN участвуют как в тренировке, так и в тестировании.



Предложенная система по созданию виртуальных ПЭТ-изображений.

Так как GAN обучается созданию реалистичных ПЭТ-изображений, результат его работы был намного ближе к реальным ПЭТ-изображениям, чем у FCN, который воспроизвел размытые изображения. Однако, FCN показал лучший результат на злокачественных образованиях, чем GAN. Авторы использовали достоинства каждого метода, чтобы создать смешанное изображение, которое соединяет в себе реалистичность от GAN и более точный ответ о злокачественности от FCN. Сперва создается маска из выходного изображения FCN, которое содержит регионы с повышенным SUV (>2.5). По этой маске берется часть изображения из FCN, а остальная достраивается из изображения от GAN.

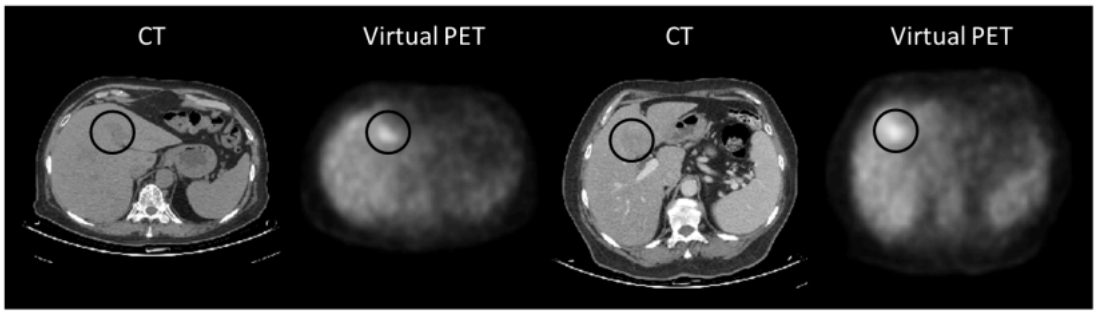
Данные

Датасет включает в себя ПЭТ изображения печени (с опухолями и без) и соответствующие им КТ изображения из Медицинского центр имени Хаима

Шибы (Израиль).

Результаты

Сгенерированные ПЭТ-изображения были визуаьлно оценены радиологом и сравнены с реальными ПЭТ-изображениями для распознавания опухолей печени. Распознанный регион считается опухолью, если он имеет значени $SUV_{max} > 2.5$. Для оценки были вычислены значения TPR и FPR. Система успешно распознала 24 из 26 опухолей (TPR 92.3%), с только двумя ложноположительными ответами среди всех 8 сканов (FPR 0.25).



Ложноположительные результаты выделены черным кругом.

Заключение

Была разработана система создания виртуальных ПЭТ-изображений по КТ изображениям с использованием FCN и GAN, которая показала сравнительно хорошие результаты. Работа интересная, однако, сложно представить ее использование в реальной жизни и степень востребованности и доверия к методу.

11. Is it Time to Replace CNNs with Transformers for Medical Images?

Ссылка

https://arxiv.org/abs/2108.09038

Введение

В течение последних лет сверточные нейронные сети (СНС) являлись лидирующим методом в автоматической медицинской диагностике. Однако, недавно появившиеся vision transformers (ViT), являются достойной альтернативной для СНС, достигая схожих уровней производительности, обладая некоторыми интересными свойствами, которые могут быть полезными в задачах распознавания медицинских изображений. В данной работе исследуется возможность замены сверточных нейронных сетей трансформерами ViT в задачах медицинской автоматизации и какие плюсы это принесет.

Основная идея

Для того, чтобы получить ответ на поставленный вопрос, авторы провели серию экспериментов, сравнивая ViT и СНС при одинаковых условиях, минимально изменяя гиперпараметры. Чтобы обеспечить чистоту эксперимента, были выбраны ResNet50, как представитель СНС и DeiT-S, как представитель ViT, так как они сравнимы по количеству параметров, затратам памяти и вычислительным мощностям. Инициализация СНС проводилась по трем стратегиям: (1) инициализация весов случайными значениями; (2) трансферное обучение (transfer learning) с с использованиям весов, предобученных на ImageNet; (3) self-supervised предобучение на целевом датасете, после инициализации как в пункте (2). Каждый эксперимент повторялся пять раз и выбирались результаты с самой высокой точностью на валидационном множестве.

Данные

APTOS 2019, ISIC 2019, CBIS-DDSM

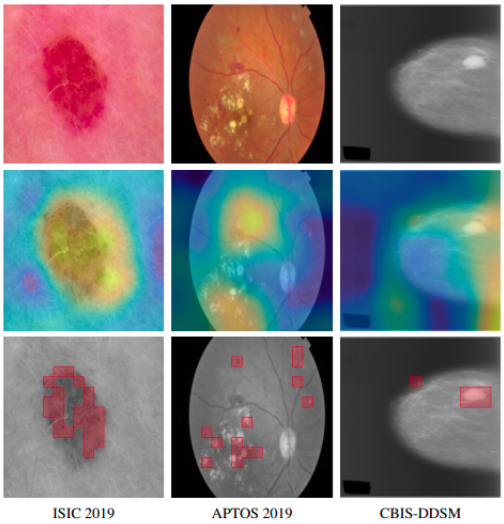
Результаты

При случайной инициализации весов CHC превосходит ViT. Такая закономерность выявлена при обучении на всех трех датасетах. Однако, при использовании весов, предобученных на ImageNet, разрыв между производительностью CHC и ViT в данной задаче сходит почти на нет. Таким образом, можно заключить:

- ViT проигрывает CHC при случайной инициализации весов и обучении с нуля;
- Трансферное обучение устраняет разрыв в производительности между ViT и CHC;
- Наилучший результат получен при подходе self-supervised+ pre-training+ fine-tuning, при котором ViT слегка превосходит CHC.

Initialization	**Model**	**APTOS2019**, $\kappa \uparrow$	**ISIC2019**, Recall $\uparrow$	**DDSM**, ROC-AUC $\uparrow$
Random	ResNet50	0.849 ± 0.022	0.662 ± 0.018	0.917 ± 0.005
	DeiT-S	0.687 ± 0.017	0.579 ± 0.028	0.908 ± 0.015
ImageNet (supervised)	ResNet50	0.893 ± 0.004	0.810 ± 0.008	0.953 ± 0.008
	DeiT-S	0.896 ± 0.005	0.844 ± 0.021	0.947 ± 0.011
ImageNet (supervised) + Self-supervised with DINO [4]	ResNet50	0.894 ± 0.008	0.833 ± 0.007	0.955 ± 0.002
	DeiT-S	0.896 ± 0.010	0.853 ± 0.009	0.956 ± 0.002

Сравнение результатов предсказания CHC и ViT в разрезе различных стратегий инициализации весов на медицинских изображениях.



Сравнение карт значимости (saliency maps) изображений из трех датасетов. В каждой колонке представлены оригинальное изображение, визуализация ResNet50 Grad-CAM saliency map и карты внимания (attention map) DEIT-S.

Заключение

В данной работе проводится анализ возможности замены сверточных нейронных сетей трансформерами (ViT) в задачах распознавания медицинских изображений. Показано, что ViT по качеству сравнима с СНС и может быть использована как альтернативный уже существующим метод.

12. ViT-V-Net: Vision Transformer for Unsupervised Volumetric Medical Image Registration

Ссылка

https://arxiv.org/abs/2104.06468

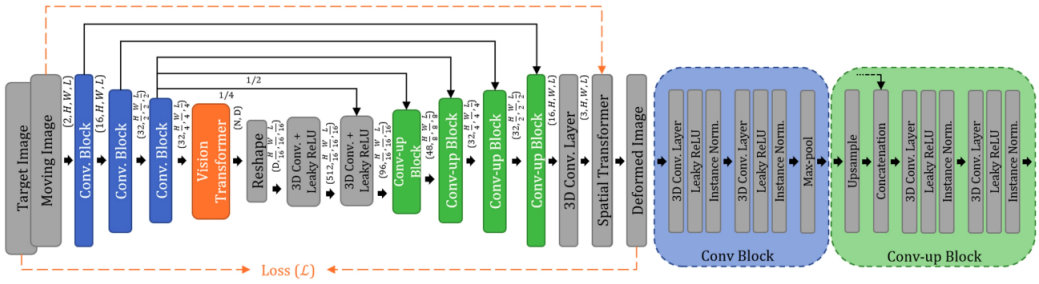
Введение

Несмотря на хорошую производительность, сверточные нейронные сети в общем случае имеют ограничения в моделировании явных пространственных отношений на большом расстоянии (например, отношения между двумя вокселями, которые находятся далеко друг от друга), присутствующих в изображении из-за локальности операции свертки. Для преодоления этого ограничения были предложены различные решения, такие как U-Net, atrous convolution и self-attention. Недавно возрос интерес в проектировании архитектуры, основанной на самовнимании (self-attention), которая хорошо проявила себя в обработке естественного языка. Была предложена ViT - архитектура, целиком и полностью основанная на self-attention. В данной работе исследуется применение ViT в объемных медицинских изображениях. Авторы предлагают ViT-V-Net, которая воплощает в себе гибридную архитектуру „сверточная нейронная сеть-трансформер“(ConvNet-Transformer) для применения self-supervised метода в исследовании трехмерных медицинских изображений.

Основная идея

В предложенном методе ViT была применена к высокоуровневым признакам изображений, что требовало от сети выявить зависимости между точками, находящимися на дальнем расстоянии. Наивное применение ViT к полномасштабным изображениям приводит к увеличению вычислительной сложности. Поэтому, изображения сначала были закодированы с помощью нескольких сверточных слоев и слоев max-pooling для получения объектов, содержащих высокоуровневые признаки. Далее, в ViT, высокоуровневые признаки делятся на патчи, а затем патчи отображаются в скрытое пространство

с помощью обучаемой линейной проекции (например, patch embedding). Затем, результирующие патчи подаются в энкодер трансформера, а полученный выход декодируется V-Net подобным декодером.



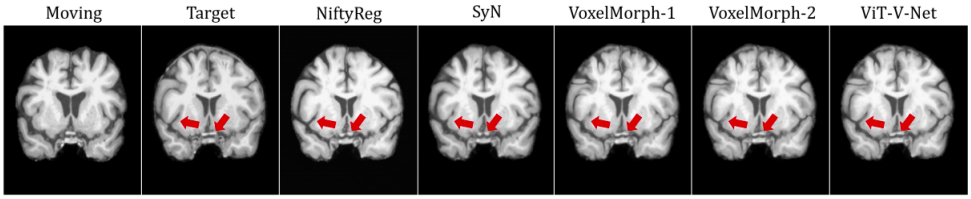
Архитектура ViT-V-Net.

Данные

МРТ-изображения в модальности T1-ВИ.

Результаты

Был проведен сравнительный анализ предложенного метода с методами Symmetric Normalization (SyN), NiftyReg и VoxelMorph-1 и -2 по мере Серенсена (Dice score). В результате ViT-V-Net более превзошла более, чем на 0.1 все рассматриваемые методы.



Результаты предсказания на корональном срезе МРТ.

	Affine only	NiftyReg	SyN	VoxelMoprh-1	VoxelMoprh-2	ViT-V-Net
Dice	0.569±0.171	0.713±0.134	0.688±0.140	0.707±0.137	0.711±0.135	**0.726±0.130**

Общие результаты производительности предложенного метода с другими по Dice score

Заключение

Предложенная архитектура, основанная на ViT достигла большей производительности, чем существующие методы и показала свою эффективность.

13. 3D Self-Supervised Methods for Medical Imaging

Ссылка

https://arxiv.org/abs/2006.03829

Введение

В данной работе предлагаются трехмерные варианты self-supervised методов, которые облегчают обучение нейронной сети на признаках по немаркированным трехмерным изображениям, что приводит к снижению затрат на экспертную аннотацию. Рассмотрены 5 алгоритмов и проведен сравнительный анализ на трехмернвх медицинских изображениях (МРТ, КТ). Выбор алгоритмов обусловен их успешным применением в двумерном случае и тем, что ни один из них не был расширен до трехмерного на момент выхода статьи.

Основная идея

Авторы предлагают 5 алгоритмов, которые целиком используют пространственную информацию 3D-изображения. В каждом методе используется энкодер g_{enc}, который может быть дообучен под различные задачи.

- **3D Contrastive Predictive Coding (3D-CPC)**
  Следуя идее, предложенной в двумерном случае, этот метод предсказывает скрытое пространство для следующих (смежных) образцов. Предложенный CPC определяет proxy-задачу, обрезая одинаковые по размеру и перекрывающиеся участки каждого сканирования. Далее, энкодер g_{enc} сопоставляет каждый входной патч $x_{i,j,k}$ его скрытому представлению $z_{i,j,k} = g_{enc}(x_{i,j,k})$. Затем, следующая модель, называемая контекстой сетью g_{cxt} суммирует скрытые вектора патчей контекста $x_{i,j,k}$ и составляет свой контекстный вектор $c_{i,j,k} = g_{cxt}(\{z_{u,v,w}\})$, где $\{z\}$ - это множество скрытых векторов. Наконец, так как $c_{i,j,k}$ захватывает высокоуровневый контент из контекста, который отвечает $x_{i,j,k}$, это позволяет предсказать скрытые представления следующих(смежных)

патчей $z_{i+l,j,k}$, где $l \geqslant 0$. Стоит отметить, что в предложенном 3D-CPC в качестве g_{enc} и g_{cxt} могут использоваться сети любой архитектуры.

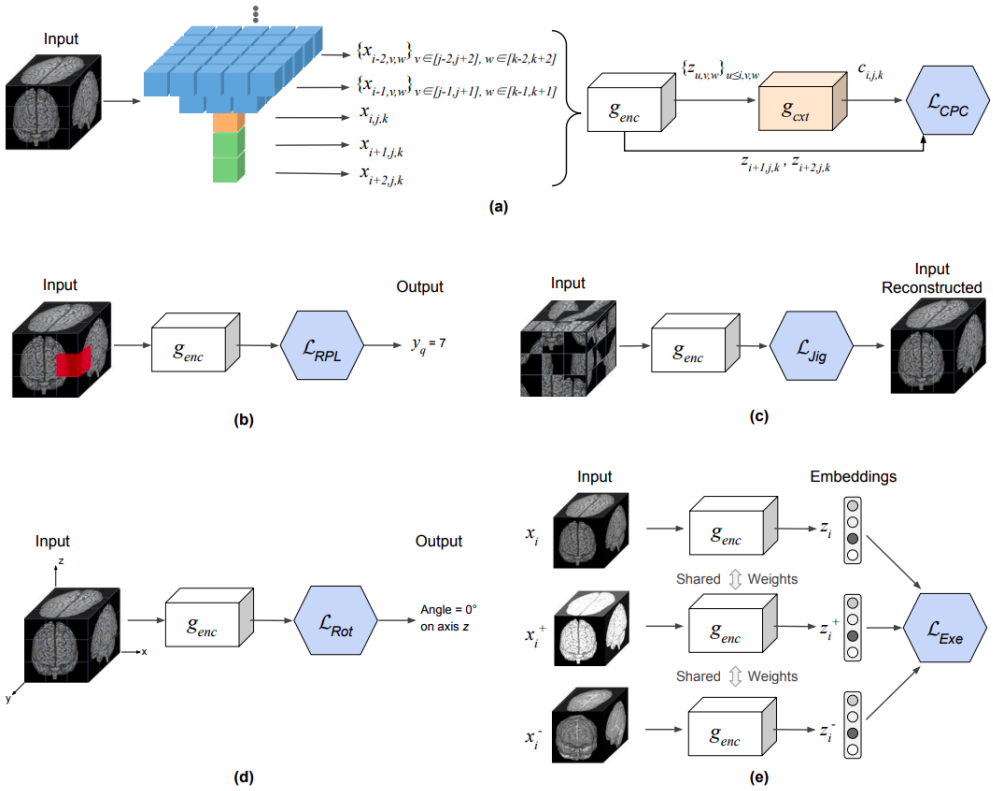
- **Relative 3D patch location (3D-RPL)**
  В этой задаче пространственный контекст в изображениях используется как богатый источник для семантического представления данных. В предложенной 3D версии из каждого входного 3D изображения выбирается сетка из N неперекрывающихся участков $\{x_i\}_{i\in\{1,\cdots N\}}$ случайного расположения. Далее, центральный патч x_c используется как ссылка, а очередной патч x_q выбирается из окружающих $N-1$ патчей. Далее, расположение x_q относительно x_c выбирается как положительная метка y_q. Таким образом, задача сводится к $N-1$-классовой классификации, где расположения оставшихся патчей используются как негативные метки.

- **3D Jigsaw puzzle Solving (3D-Jig)**
  Получение мозаичной сетки из входного изображения может рассматриваться как расширение вышеприведенной задачи RPL, основанной на патчах. Пазлы формируются путем выбора *nxnxn* сетки из 3D патчей, далее эти патчи перемешиваются следуя произвольной перестановке из множества предопределенных перестановок с индексом $y_p \in \{1,\ldots,P\}$, где P - размерность множества перестановок, выбранного из $n^3!$ всевозможных перестановок. Таким образом, задача сводится к P-классовой классификации - модель тренируется просто запомнить индекс p примененной перестановки.

- **3D Rotation prediction (3D-Rot)**
  В данной задаче модель должна предсказать угол, на который повернуто изображение. Входное изображение поворачивается случайным образом на угол $r \in \{1,\ldots R\}$. Поворот изображения на угол в 0^o вдоль трех осей произведет три идентичных версии исходного изображения, поэтому рассматриваются только 10 возможных поворотов из 12. В таких условиях задача сводится к 10-классовой классификации.

- **3D Exemplar networks (3D-Exe)**
  Для получения supervised-меток метод опирается на аугментацию изображений. Здесь для тренировочного набора данных определяется множество трансформаций изображения, а новый суррогатный класс со-

здается с помощью трансформации тренировочного примера. Задача является обычной задачей классификации с кросс-энтропийной фукнцией потерь. Однако, с увеличением датасета и количства классов задача становится более вычислительно сложной, поэтому в предложенной 3D версии внедрен механизм, который опирается на тройную функцию потерь.



(a) - 3D-CPC; (b) - 3D-RPL; (c) - 3D-Jig; (d) - 3D-Rot; (e) - 3D-Exe.

Данные

BraTS 2018, 3D КТ сканы с опухолями поджелудочной железы, снимки из Diabetic Retinopathy 2019 Kaggle challenge.

Результаты

Предложенные методы были опробованы в различных медицинских задачах и показали следующие результаты:

1. Сегментация опухолей мозга

Все предложенные методы продолевают бейзлайны, так же, как и двумерные

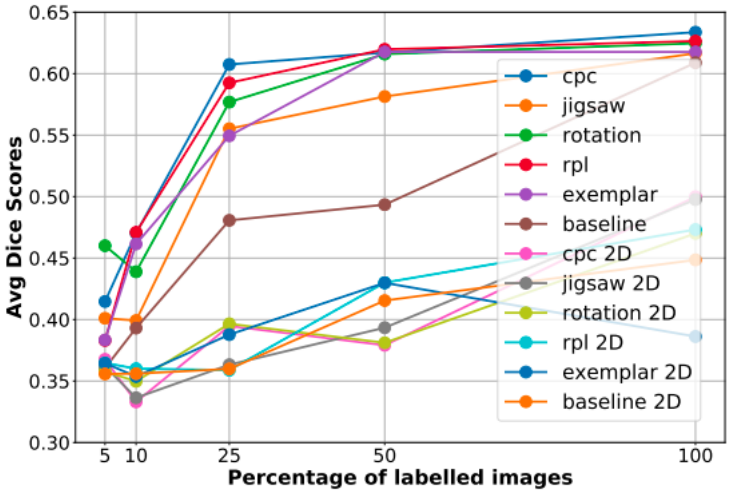
версии этих методов. Результаты, полученные в данной здаче показвают наличие обощающей способности у всех предложенных методов.

Model	ET	WT	TC
3D-From scratch	76.38	87.82	83.11
3D Supervised	78.88	90.11	84.92
2D-CPC	76.60	86.27	82.41
2D-RPL	77.53	87.91	82.56
2D-Jigsaw	76.12	86.28	83.26
2D-Rotation	76.60	88.78	82.41
2D-Exemplar	75.22	84.82	81.87
Popli *et al.* [66]	74.39	89.41	82.48
Baid *et al.* [67]	74.80	87.80	82.66
Chandra *et al.* [68]	74.06	87.19	79.89
Isensee *et al.* [65]	80.36	**90.80**	84.32
3D-CPC	80.83	89.88	85.11
3D-RPL	**81.28**	90.71	**86.12**
3D-Jigsaw	79.66	89.20	82.52
3D-Rotation	80.21	89.63	84.75
3D-Exemplar	79.46	**90.80**	83.87

Результаты сегментации BraTS

2. Сегментация опухолей поджелудочной железы

Результаты, полученные с помощью предложенных методов преодолевают бейзлайны для поставленной задачи. Также, предложенные методы показывают достаточно быструю сходимость.



Результаты сегментации опухолей поджелудочной железы. На меньшем количестве размеченных данных supervised baseline (коричневый) показывает низкую обобщающую способность по сравнению с предложенными методами. Также, 3D методы превосходят свои двумерные аналоги.

Заключение

В данной работе были продемонстрированы результаты применения предложенных алгоритмов в разрезе эффективности обработки данных и более быстрой сходимости. Полученные результаты являются конкурентноспособными, а разработанные методы могут применяться в дальнейших исследованиях.

14. Self-Supervised Learning for 3D Medical Image Analysis using 3D SimCLR and Monte Carlo Dropout

Ссылка

https://arxiv.org/abs/2109.14288

Введение

Self-supervised обучение проявило себя как мощный инструмент, позволяющий конструировать значимые представления из неразмеченных данных, что может быть использовано в задачах с малым количеством размеченных данных. В данной статье авторы представляют метод, который использует возможности SimCLR в сегментации 3D изображений. Также показано, что дополнительное включение неопределенности посредством Байесовского вывода в форме метода Монте-Карло значительно улучшает производительность в задаче сегментации.

Основная идея

Метод состоит из трех частей: сначала производится self-supervised обучение энкодера, далее энкодер дообучается под необходимую задачу сегментации с использованием размеченных данных, а затем применяется метод Монте-Карло во время предсказания и вычисляется Dice score на тестовом множестве.

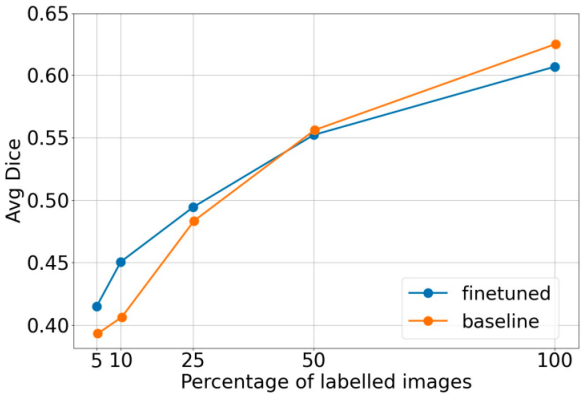
Если рассматривать метод более детально, то в первую очередь решается предварительная задача, которая обобщает SimCLR до трехмерных входов для исследования трехмерного пространственного контекста. Случайным образом 3D сканы разбиваются на батчи размером M, затем каждый скан делится на P равных неперекрывающихся 3D патча, в результате получая $N = M * P$ входных примеров. К входным данным применяется два случайных типа аугментации. Архитектура модели, решающая предзадачу следующая: энкодер (3D-CNN), за ним следует слой нелинейной проекции (Dense layer).

Для решения основной задачи использовался предобученный энкодер без слоя нелинейной проекции. Выходы энкодера подаются на вход декодеру (U-Net), который тренируется уже на размеченных данных.

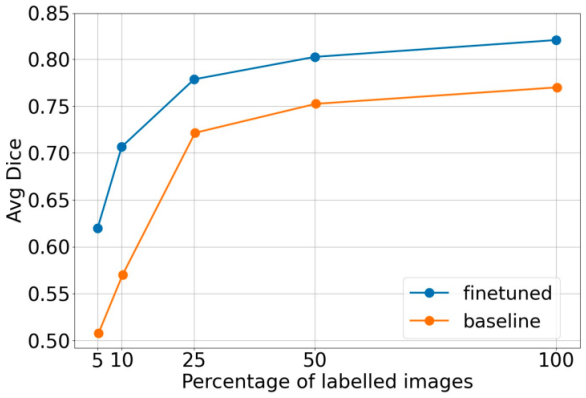
Данные

BraTS 2018, 3D КТ сканы с опухолями поджелудочной железы (ПЖЖ).

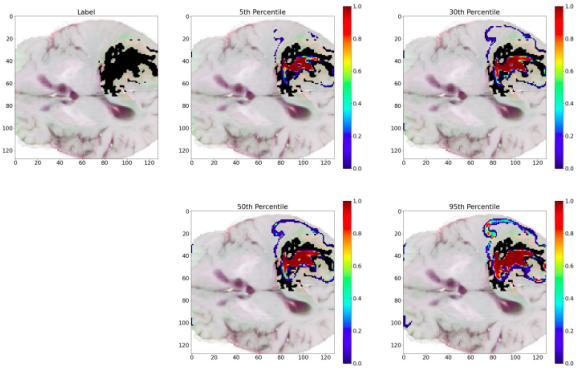
Результаты



Средний Dice коэффициент после дообучения модели на 5%, 10%, 25%, 50%, 100% снимков ПЖЖ. 3D SimCLR модель (синий) превосходит baseline (оранжевый), когда доступно менее, чем 25% данных.



По оси Y - средний Dice коэффициент после дообучения модели на 5%, 10%, 25%, 50%, 100% тренировочного множества (BraTS).Предложенная модель - синяя линия, baseline - оранжевая.



Тепловые карты различных перцентилей предсказаний классов опухоли для примера из датасета BraTS. Черные пиксели представляют опухоль целиком.

Заключение

Результаты экспериментов показывают потенциал предлагаемого метода 3D SimCLR с дополненной информацией из Байесовского вывода в разрезе эффективности обработки данных и повышения производительности.

15. PGL: Prior-Guided Local Self-supervised Learning for 3D Medical Image Segmentation

Ссылка

https://arxiv.org/abs/2011.12640

Введение

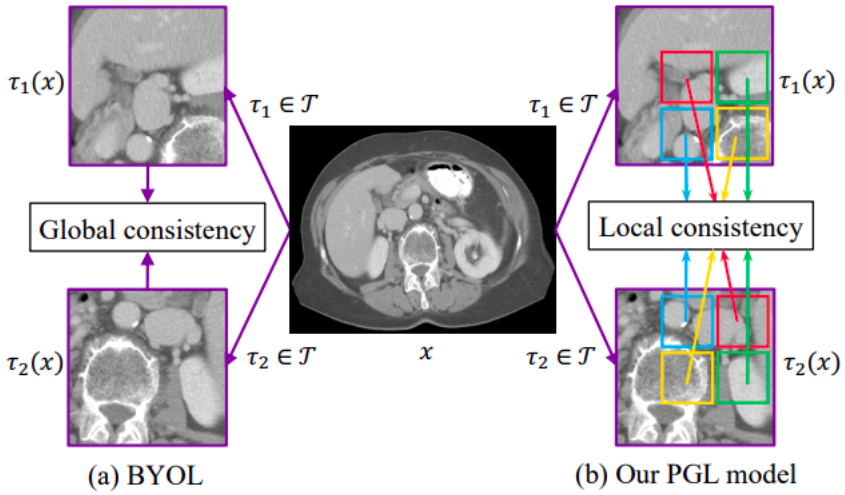
В данной работе предланается self-supervised модель Prior-Guided Local (PGL) для сегментации трехмерных медицинских изображений, которая использует изначально известное расположение между парой позитивных изображений, чтобы выявить местную зависимость признаков в одном и том же регионе.

Основная идея

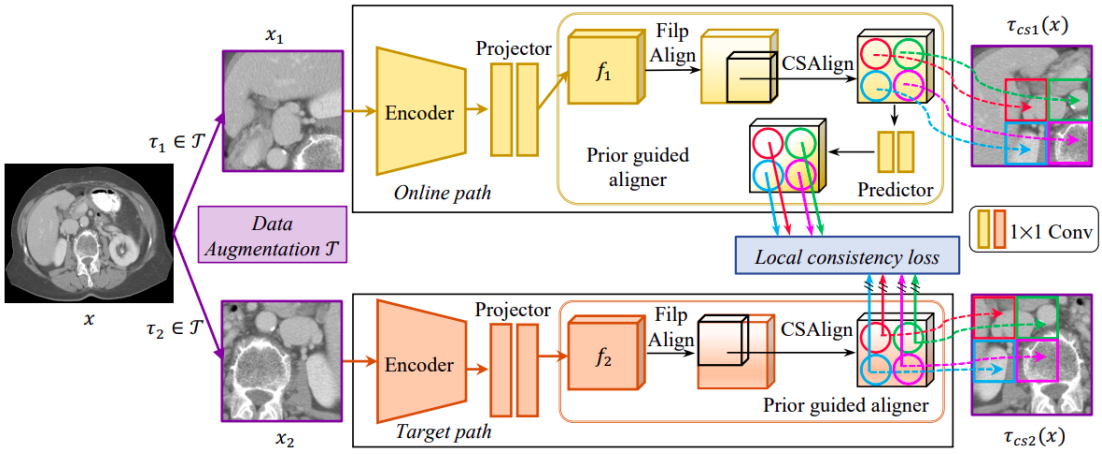
Предложенная модель состоит из модуля аугментации данных для генерации представлений изображения и модуля известного двойного пути (prior dual-path module) для извлечения признаков. Далее конструируется функция потерь местных зависимостей, для минимизации различий между каждой парой выявленных признаков. Таким образом, модель учится захватывать больше структурной информации и больше подходит для решения задачи сегментации, чем методы, основанные на выявлении глобальных зависимостей.

Отличие метода PGL от BYOL

Bootstrap Your Own Latent (BYOL) - метод, в котором сеть обучается онлайн на представлениях изображений для того, чтобы предсказать вид следующего представления. BYOL фокусируется на изучении глобальных зависимостей между парой представлений, в то время как PGL использует априорную информацию о взаимном расположении двух представлений, чтобы извлечь локальные зависимости в одинаковых регионах.



BYOL и PGL



Архитектура PGL

Данные

Liver,Spleen,KiTS, BCV из Medical Segmentation Decathlon (MSD) соревнования, RibFac датасет.

Результаты

Производительность baseline сети с использованием случайной инициализации или одной из трех стратегий предобучения: Models Genesis (MG), BYOL и PGL на датасете BCV:

Methods		Organs													Ave
		Sp	RK	LK	Gb	Es	Li	St	Aorta	IVC	PSV	Pa	RAG	LAG	
Dice ↑	Random Init	94.01	92.97	92.15	51.98	71.85	94.82	77.74	87.47	84.85	70.91	74.12	62.27	67.30	78.65
	MG [39]	94.92	93.03	91.87	59.80	71.28	95.27	80.88	87.92	85.34	71.95	75.88	63.70	67.77	79.97
	BYOL [7]	95.04	93.53	92.55	59.70	70.98	95.35	80.69	88.37	85.36	71.93	75.95	63.71	68.27	80.11
	PGL(Ours)	**95.46**	**93.54**	**92.62**	**59.91**	**72.59**	**96.14**	**81.99**	**89.20**	**86.49**	**72.50**	**77.00**	**63.85**	**69.75**	**80.85**
IoU ↑	Random Init	88.87	86.95	85.76	40.97	57.12	90.51	65.87	78.65	73.95	55.42	59.48	46.95	51.43	67.84
	MG [39]	90.44	87.06	85.19	**47.86**	56.52	91.28	69.98	79.07	74.68	56.74	61.60	48.31	51.8	69.27
	BYOL [7]	90.63	87.89	86.42	47.73	56.30	91.43	69.71	79.74	74.72	56.70	61.69	48.06	52.31	69.49
	PGL(Ours)	**91.35**	**87.93**	**86.50**	47.72	**58.19**	**92.63**	**71.84**	**80.90**	**76.38**	**57.37**	**63.00**	**48.32**	**54.16**	**70.48**
HD ↓	Random Init	38.31	2.06	2.54	51.75	8.83	3.64	48.28	26.92	6.12	16.73	14.66	5.22	3.82	17.61
	MG [39]	4.43	2.07	24.89	12.69	7.65	3.24	20.77	26.20	5.20	**8.61**	6.02	5.31	4.44	10.12
	BYOL [7]	3.46	1.92	**2.45**	25.96	20.41	3.11	22.61	17.45	5.20	16.36	5.94	**4.52**	4.46	10.30
	PGL(Ours)	**2.50**	**1.83**	2.47	**11.52**	**7.18**	**2.52**	**12.45**	**6.23**	**4.77**	13.85	**6.00**	4.75	**3.56**	**6.13**

Результаты предсказания моделей на датасете BCV. Ave - средний результат сегментации 13 органов.

		Random Init	Models Genesis [39]	BYOL [7]	**PGL (Ours)**
Organ	Dice ↑	95.73	96.00	96.29	**96.43**
	IoU ↑	91.94	92.45	92.92	**93.16**
	HD ↓	7.49	4.75	5.46	**4.72**
Tumor	Dice ↑	52.20	53.47	53.34	**55.66**
	IoU ↑	41.64	42.91	43.25	**44.95**
	HD ↓	29.69	29.87	32.82	**24.01**
Ave	Dice ↑	73.97	74.74	74.82	**76.05**
	IoU ↑	66.79	67.68	68.09	**69.06**
	HD ↓	18.59	17.31	19.14	**14.37**

Производительность PGL модели с различными пространственными трансформациями на датасете Liver. Ave - средний результат сегментации печени и опухоли печени.

		Random Init	Models Genesis [39]	BYOL [7]	**PGL (Ours)**
Organ	Dice↑	96.07	96.29	96.23	**96.80**
	IoU ↑	92.62	92.94	92.99	**93.83**
	HD ↓	1.95	1.84	2.1	**1.34**
Tumor	Dice ↑	67.07	68.35	**71.89**	71.77
	IoU ↑	56.63	58.07	62.12	**62.70**
	HD ↓	25.64	21.12	21.73	**17.59**
Ave	Dice ↑	81.57	82.32	84.06	**84.29**
	IoU ↑	74.63	75.51	77.56	**78.27**
	HD ↓	13.80	11.48	11.92	**9.47**

Производительность PGL модели с различными пространственными трансформациями на датасете KiTS. Ave - средний результат сегментации почек и опухолей почек.

Заключение

Были проведены масштабные эксперименты на четырех КТ датасетах, которые включали в себя 11 органов и два вида опухолей. Результаты показали, что использование PGL для инифиализации сети для сегментации позволяет сильно улучшить производительность сети, также показано превосходство предложенной модели PGL над моделью BYOL.