

*Березовский Илья Игоревич
студент 2 курса магистратуры,
факультета систем управления
Томский государственный университет систем управления и
радиоэлектроники,
Россия, г. Томск
e-mail: iliaberez@mail.ru*

*Научный руководитель: Аксёнов С. В.,
кандидат технических наук,
доцент кафедры автоматизации и обработки информации
Томский государственный университет систем управления и
радиоэлектроники,
Россия, г. Томск*

ОБЗОР СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ СЕГМЕНТАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Аннотация: *В рамках статьи производится обзор существующих свёрточных нейросетевых структур, с целью нахождения наиболее подходящей для дальнейшего использования в решении задачи сегментации долей лёгких по изображениям компьютерной томографии грудной клетки.*

Обзор архитектур нейронных сетей производится с приведением схем моделей и их описанием.

Ключевые слова: компьютерное зрение, глубокое обучение, SegNet, U-net, PSPNet, свёрточные нейронные сети, сегментация медицинских изображений.

*Berezovskiy Ilya Igorevich
2nd year master student,
faculty of control system
Tomsk State University of Control System and Radioelectronics,
Russia, Tomsk*

*Scientific adviser: Aksenov S.V.,
candidate of technical sciences,
associate professor of the Department of automation and information processing
Tomsk State University of Control System and Radioelectronics,
Russia, Tomsk*

Abstract: *The article reviews the existing convolutional neural network structures in order to find the most suitable one for further use in solving the problem of segmentation of lung lobes from chest CT images.*

Neural network architectures are reviewed with model diagrams and their description.

Key words: Computer Vision, Deep Learning, SegNet, U-net, PSPNet, Convolutional Neural Networks, Medical Image Segmentation.

Введение

На сегодняшний день технологии компьютерного зрения глубоко вошли в нашу жизнь. Они применяются в большом количестве сфер деятельности человека, назовём наиболее крупные из них – это розничная торговля, промышленность и медицина. В связи с ростом вычислительных мощностей и появлением больших баз различных изображений, стало возможным обучать глубокие нейронные сети – нейронные сети с большим числом скрытых слоёв, и насчитывающие огромное количество нейронов. В задаче сегментации изображений особого успеха добились энкодер-декодер нейронные сети. Проанализировав различные источники по сегментации медицинских изображений, было решено выбрать наиболее современные и эффективные архитектуры, и провести их обзор в данной статье.

Целью данной статьи является обзор свёрточных нейронных сетей для сегментации медицинских изображений, а также проведение эксперимента с целью выбора наиболее подходящей для задачи сегментации долей лёгких.

1 Задача сегментации

Задача сегментации медицинских изображений представляет собой – процесс выделения на изображении интересующей области (области интереса) и определение её границ. Сегментация медицинских изображений позволяет составлять рекомендации заинтересованным лицам о расположении медицинских приборов, о расположении патологий при оперативных вмешательствах, а также позволяет отследить динамику распространения патологий.

2 Свёрточные нейронные сети

Особого успеха в сегментации изображений достигли свёрточные нейронные сети, поэтому было решено рассматривать именно их, и начать стоит с рассмотрения самой первой свёрточной нейронной сети.

Свёрточная нейронная сеть – это нейронная сеть, в которой присутствуют слои свёртки (convolutional layer). Обычно в свёрточных нейронных сетях также присутствует слой субдискретизации (pooling layer) и полносвязный слой (fully connected layer). Основы современной архитектуры свёрточных нейронных сетей были заложены в одной из первой широко известной свёрточной нейронной сети — LeNet-5 [2] Яна ЛеКуна, архитектура которой представлена на рисунке 1.

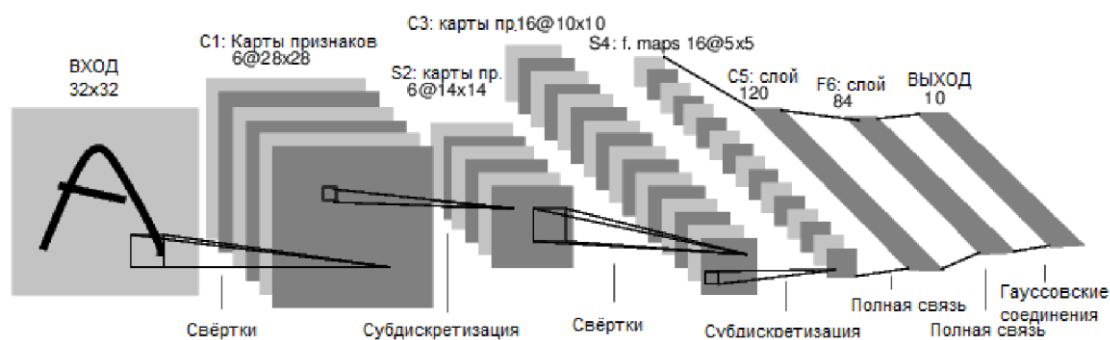


Рисунок 1. Архитектура LeNet - 5

В свёрточных нейронных сетях слои свёртки и субдискретизации состоят из нескольких «уровней» нейронов, называемых картами признаков (feature maps), или каналами (channels). Каждый нейрон такого слоя соединён с небольшим участком предыдущего слоя, называемым рецептивным полем. В случае изображения, карта признаков является двумерным массивом нейронов, или просто матрицей. Другие измерения могут быть использованы, если на вход принимается другой вид данных, например, аудио данные (одномерный массив) или объёмные данные (трёхмерный массив). В слое свёртки каждой карте признаков соответствует одно ядро свёртки, также называемое фильтром. Каждый нейрон в качестве своего выходного значения осуществляет операцию свёртки или взаимной корреляции со своим рецептивным слоем. Стоит заметить, что эти две операции в контексте обучения свёрточных нейронных сетей взаимозаменяемы, вследствие чего во многих программных реализациях

операция «свёртки» на самом деле является операцией взаимной корреляции. Так как ядро свёртки для каждой карты признаков одно, это позволяет нейронной сети научиться выделять признаки вне зависимости от их расположения во входном изображении и также приводит к значительному уменьшению числа параметров. Согласно устоявшейся нотации, говорят, что слой свёртки использует фильтр $W \times H$, если каждый фильтр в этом слое имеет размерность $W \times H \times C$, где C — число каналов в предыдущем слое. Слой субдискретизации осуществляет уплотнение карт признаков предыдущего слоя и не изменяет количество карт. Каждая карта признаков слоя соединена с соответствующей картой признаков предыдущего слоя, каждый нейрон выполняет «сжатие» своего рецептивного поля посредством какой-либо функции. Наиболее популярными видами этого слоя являются Max Pooling (из рецептивного слоя выбирается максимальное значение), Average Pooling (выбирается среднее значение) и L2 Pooling (выбирается норма L2). С помощью слоя субдискретизации достигается устойчивость к небольшим сдвигам входного изображения, а также уменьшается размерность последующих слоёв. Полносвязный слой — обычный скрытый слой многослойного перцептрона, соединённый со всеми нейронами предыдущего слоя. Таким образом, на вход свёрточной нейронной сети подаётся изображения, а на выходе, в зависимости от функции активации выходного нейрона и функции оптимизации, класс к которому принадлежит изображение или результат сегментации.

3 Свёрточные нейронные сети для сегментации изображений

3.1 SegNet

Модель SegNet [1] является автокодировщиком, основанным на свёрточной нейронной сети. Схема данной модели представлена на рисунке 2. Данная сеть состоит из блоков, в каждом из них находятся слои свёртки и Max Pooling слои или Up Sampling слои, а также ReLU слои активации и слои Batch Normalization. Архитектура данной нейронной сети полностью симметрична. На выходе данной нейронной сети находится слой Soft Max, который преобразует

каждый пиксель выходной матрицы изображения в целое число, показывающее к какому классу принадлежит каждый пиксель.

Главное отличие SegNet от других автокодировщиков состоит в том, что его Up Sampling слои декодера соединены с соответствующими Max Pooling слоями декодера, поэтому Up Sampling слои не обучаются, а получают нужную информацию о том, как повысить размерность и как восстановить сжатую топологию от соответствующих MaxPooling слоев.

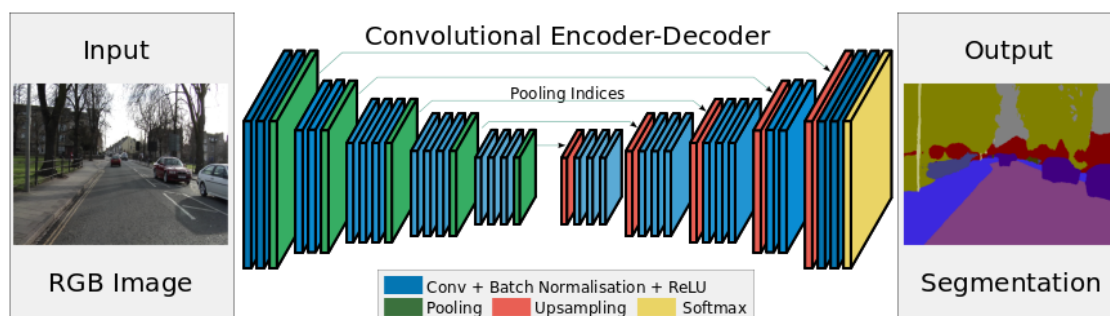


Рисунок 2. Архитектура SegNet

Основными достоинствами данной нейронной сети является высокая производительность. Недостатками сети являются невысокое разрешение карт сегментации на выходе и невысокая точность сегментации по современным меркам.

3.2 UNet

Модель UNet [2] является самым известным представителем энкодер-декодер моделей, для сегментации медицинских изображений. Архитектура данной нейронной сети представлена на рисунке 3. Данная сеть состоит из двух частей, а именно сужающейся (энкодер) и расширяющейся (декодер).

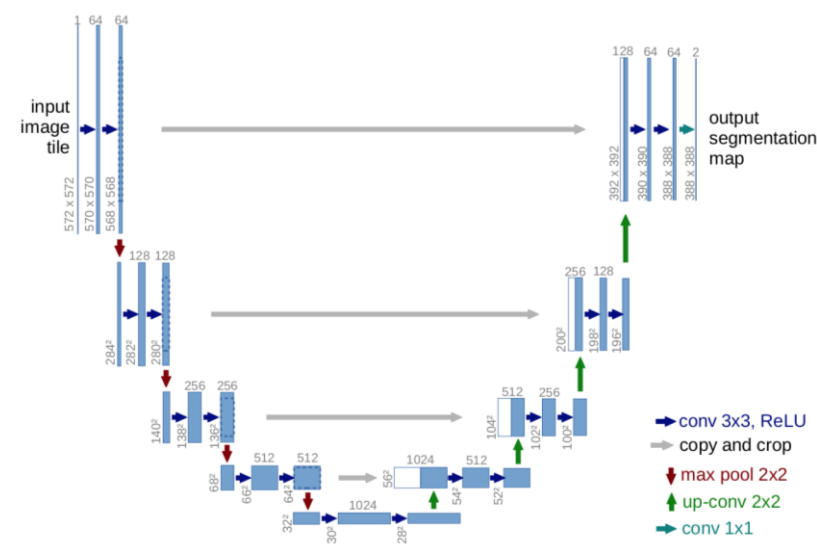


Рисунок 3. Архитектура UNet

Первая представляет из себя типичную архитектуру свёрточную классификационной сети. Состоит из повторяющихся применений двух сверток 3x3, за которыми следуют слой активации ReLU и слой Max Pooling 2x2 с шагом 2. На каждом шаге повышается количество каналов вдвое. Расширяющаяся часть состоит из слоёв так называемой обратной свертки DeConv 2x2, уменьшающей количество каналов, затем конкатенация с соответствующим образом, обрезанным карту признаков от соответствующей части сужающейся части и две свертки 3x3 и слой активации ReLU. Обрезка необходима из-за потери пограничных пикселей в каждой свертке. На последнем уровне свертка 1x1 используется для сопоставления каждому 64-компонентного вектору признаков класса.

Основные достоинства данной нейронной сети являются высокая точность сегментации медицинских изображений, возможность обучения модели на небольшом объёме данных.

3.3 PSPNet

На рисунке 4 изображена архитектура PSPNet. На (a) расположено входное изображение. На (b) ResNet используется с расширенной сетевой стратегией для извлечения функций. Размер карты признаков здесь составляет 1/8 от входного изображения. На (c) среднее объединение субрегионов выполняется для каждой карты объектов.

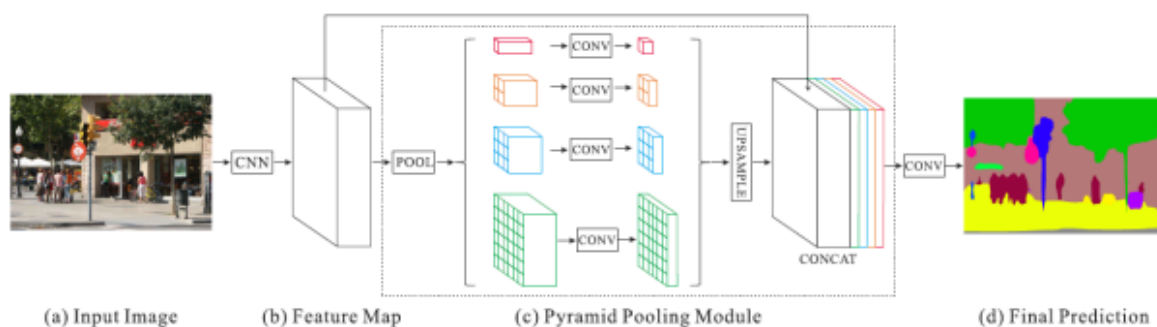


Рисунок 4. Архитектура PSPNet

Красный — самый грубый уровень, который выполняет глобальное среднее объединение для каждой карты объектов, чтобы сгенерировать один выходной лоток. Оранжевый — это второй уровень, который делит карту объектов на 2×2 субрегиона, а затем выполняет среднее объединение для каждого субрегиона. Синий — это третий уровень, который делит карту объектов на 3×3 субрегиона, а затем выполняет среднее объединение для каждого субрегиона. Зеленый — это лучший уровень, который делит карту объектов на 6×6 субрегионов, а затем выполняет объединение для каждого субрегиона. Затем выполняется свертка 1×1 для каждой объединенной карты признаков, чтобы уменьшить представление контекста до $1/N$ от исходного (черного), если размер уровня пирамиды равен N . В этом примере $N = 4$, потому что всего есть 4 уровня (красный, оранжевый, синий и зеленый). Если количество входных карт объектов равно 2048, то выходная карта объектов будет $(1/4) \times 2048 = 512$, т.е. 512 номеров выходных карт объектов. Билинейная интерполяция выполняется для повышения дискретизации каждой карты объектов низкого размера, чтобы иметь тот же размер, что и исходная карта признаков (черная). Все различные уровни карт признаков с повышенной дискретизацией объединяются с исходной картой признаков (черный цвет). Эти карты функций слиты как глобальный приоритет. Это конец” пирамидального” пулинг модуля в (с).

4 Проведение эксперимента

Для оценки качества построенных модели используют следующие метрики качества: Ассигасу и IoU. Разберём данные метрики.

Ассигасу – доля правильных ответов. В нашем случае, отображает процент пикселей в изображении, которые были правильно классифицированы. Формула расчёта представлена на рисунке 5.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Рисунок 5. Accuracy

Mean Intersection over Union (mIoU) – измеряет общее количество между целевой и прогнозной масками, делённое на общее количество пикселей, присутствующих в обеих масках. Формула расчёта представлена на рисунке 6.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


Рисунок 6. IoU

Для оценки эффективности работы архитектур нейронных сетей, был выбран датасет – LIDC-IDRI, который содержит в себе 463 файла, формата dcm компьютерной томографии грудной клетки людей, заболевших COVID-19, файлы данного формата содержат в себе огромное количество изображений срезов грудной клетки. И ещё один датасет Chest X-ray Images состоящий из 5856 изображений грудной клетки в jpg формате. Сравнение архитектур проводится на разных датасетах, для того, чтобы выявить особенности моделей, связанные с размером предоставляемых данных. Результаты эксперимента отображены в таблицах 1-2.

Таблица 1.

Результаты для набора данных LIDC-IDRI

Архитектура	Accuracy	IoU
SegNet	0.95	0.9
UNet	0.91	0.89
PSPNet	0.89	0.83

Таблица 2.**Результаты для набора данных Chest X-ray Images**

Архитектура	Accuracy	IoU
SegNet	0.85	0.82
UNet	0.95	0.92
PSPNet	0.87	0.84

Исходя из результатов экспериментов, можно сделать вывод, что с датасетом LIDC-IDRI, нейронная сеть, построенная на архитектуре SegNet, справилась с заданием лучше остальных моделей. Во второй таблице, результаты нейронной сети, построенной на архитектуре UNet, лучше остальных. Можно подвести итог, что нейронная сеть, построенная на архитектуре SegNet, лучше справляется с большими наборами данных, чем остальные, а модель, построенная на архитектуре UNet, способна выдать хорошую точность даже на не больших объёмах данных.

Заключение

В данной работе был проведён обзор самых актуальных архитектур нейронных сетей, для сегментации медицинских изображений. Они позволили сильно улучшить точность сегментации и достичь результатов, не достигнутых классическими методами компьютерного зрения. Для решения задачи сегментации лёгких на доли, было решено выбрать архитектуру нейронной сети UNet, так как автор ограничен количеством изображений в датасете. Таким образом все поставленные цели были выполнены.

Список литературы:

1. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Электронный ресурс] // Режим доступа URL: <https://arxiv.org/pdf/1511.00561.pdf> (дата обращения 23.11.2021 г.).
2. U-Net: Convolutional Networks for Biomedical Image Segmentation [Электронный ресурс] // Режим доступа URL: <https://arxiv.org/pdf/1505.04597.pdf> (дата обращения 23.11.2021 г.).

