



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА»

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра математики

**Проектная конференции в рамках МФК
по искусственному интеллекту
«Сегментация биомедицинских изображений при помощи
свёрточных нейронных сетей»**

Выполнил

студент 402 группы

Нурхабинов Т. Т.

Научный руководитель

д.ф.-м.н. Попов В. Ю.

Москва

2021

Содержание

<i>1.1 Цели работы</i>	2
2 Основная часть	3
<i>2.1 Компьютерное зрение</i>	3
<i>2.2. Постановка задачи</i>	3
<i>2.3 Свёрточные нейронные сети</i>	4
<i>2.4 Архитектуры свёрточных нейронных сетей</i>	5
<i>2.5 Результаты обучения свёрточных нейронных сетей</i>	7
<i>2.6 Выводы</i>	9
3 Заключение	9
Список литературы	9

1 Введение

Задачи обнаружения и выделения объектов на изображении являются фундаментальными проблемами компьютерного зрения, они возникают во многих областях практической деятельности человека, например при автоматическом управлении транспортными средствами[10], обнаружении опухолевых поражений в медицине[11], обнаружении аномалий и предотвращении опасных ситуаций на производственных площадках.

Одним из лучших подходов к решению задач компьютерного зрения является использование методов глубокого обучения, в частности свёрточных нейронных сетей. Настоящая работа посвящена задаче семантической сегментации изображения, то есть его разделение на группы пикселей, относящиеся к определенным классам. Для решения данной задачи используются свёрточные нейронные сети. Примерами для обучения и оценки точности выступают снимки ПЭТ/КТ с размеченными участками опухолевых поражений.

1.1 Цели работы

- Изучение нейросетевых методов решения задач компьютерного зрения,

- Обучение сверточных нейронных сетей для семантической сегментации участков опухолевых поражений на снимках ПЭТ/КТ,
- Сравнение архитектур сверточных нейронных сетей.

2 Основная часть

2.1 Компьютерное зрение

Все задачи компьютерного зрения сводятся к анализу изображения или видеопотока (представляющего из себя набор сменяющихся изображений).

В данной работе изображения являются цветными (RGB). Массив яркостей пикселей такого изображения – трехмерный массив размера $H \times W \times 3$. Множество всех таких изображений составляет евклидово пространство $\mathbb{R}^{H \times W \times 3}$.

Среди задач компьютерного зрения по обнаружению объектов можно выделить следующие: семантическая сегментация (semantic segmentation), многоклассовая сегментация (instance segmentation), классификация и локализация (object recognition), детектирование нескольких объектов (Object detection).

2.2. Постановка задачи

Искусственная нейронная сеть (ИНС) – функция многих переменных, параметры которой изменяются в процессе обучения.

Задача сегментации изображения сводится к классификации его пикселей. Пусть неизвестная целевая функция $y^*: X \rightarrow Y$, относящая пиксели к классам, известна для ограниченного набора данных - обучающей выборки $D = \{(x_i, y_i)\}_{i=1}^N$.

Вводится функция потерь $\mathcal{L}(y_i, f(x_i, \theta))$, характеризующая величину отклонения ответа $f(x_i, \theta)$, зависящего от параметров ИНС, от правильного ответа $y_i = y^*(x_i)$ на произвольном объекте $x_i \in X$. Тогда эмпирический риск [14] – функционал качества, характеризующий среднюю ошибку на обучающей выборке:

$$Q(\{f(x_i, \theta)\}_{i=1}^N, \{y_i\}_{i=1}^N) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(f(x_i, \theta), y_i), \quad (1)$$

$$\theta = \operatorname{argmin}_{\theta} Q(\{f(x_i, \theta)\}_{i=1}^N, \{y_i\}_{i=1}^N). \quad (2)$$

В задаче бинарной классификации $y \in \{0, 1\}$. Тогда введём распределение данных таким образом: $p(y = 1) = y, p(y = 0) = 1 - y$.

Классификатор пытается оценить вероятность положительного ответа:

$$q(y) = p(y|D, x). \quad (4)$$

Введём *перекрёстную энтропию* (cross-entropy):

$$H(p, q) = \mathbb{E}_p[-\log q] = -\sum_y p(y) \log q(y). \quad (5)$$

Для бинарной классификации на наборе данных $D = \{(x_i, y_i)\}_{i=1}^N$ минимизируется следующий функционал:

$$Q(\theta) = H(p_{data}, q(\theta)) = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i(\theta) + (1 - y_i) \log(1 - \hat{y}_i(\theta))], \quad (6)$$

где $\hat{y}_i(\theta)$ – оценка вероятности ответа 1, полученная классификатором.

Можно использовать и другие функционалы качества, например:

$$Q(\theta) = MSE = \frac{1}{N} \sum_{i=1}^N (f(x_i, \theta) - y_i)^2. \quad (7)$$

2.3 Свёрточные нейронные сети

В задачах компьютерного зрения обычно используются свёрточные нейронные сети (CNN). Прежде, чем перейти к описанию архитектур ИНС, используемых в работе, рассмотрим основные элементы свёрточных нейронных сетей.

Слоем свёрточной сети будем называть массив чисел, представленный в виде трёхмерной матрицы размером $H \times W \times C$, где H, W, C – высота, ширина и число каналов соответственно.

Основным структурным блоком свёрточной нейронной сети является *свёрточный слой*, получаемый в результате применения операции трёхмерной свёртки с ядром, веса которого являются обучаемыми параметрами (рис.1). Результат действия свёртки выражается формулой:

$$Y_{ij} = \sum_{m=1}^h \sum_{l=1}^w \sum_{k=1}^C w_{cmlk} X_{(i+m-1), (j+l-1), k}, \quad (8)$$

где $\{w_{cmlk}, i = 1, \dots, h, j = 1, \dots, w, k = 1, \dots, C\}$ – ядро с номером c . Введём обозначение $Y = Conv_{h,w}^{H,W,C,K}$ для такого свёрточного оператора, тогда формула (1) примет вид:

$$Y = Conv_{h,w}^{H,W,C,K} X. \quad (9)$$

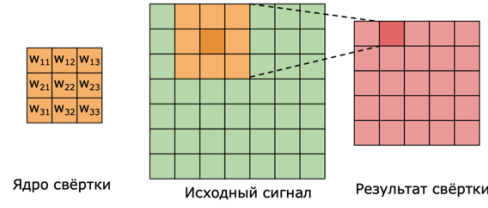


Рис. 1: Действие свёртки

Свёртка – линейный оператор.

Пулинг-слой – слой понижающий размерность сигнала (обычно не содержит обучаемых параметров). Одним из наиболее распространённых видов пулинга является *max-пулинг*. Каждый канал изображения разбивается на блоки одинакового размера $h \times w$, далее из максимальных значений в каждом блоке формируется изменённый канал изображения (рис. 2).

$$Y_{ijk} = \max\{X_{(i+m-1),(j+l-1),k}, m = 1, \dots, h, l = 1, \dots, w\}. \quad (10)$$

Введём обозначение $P_{h,w}^{H,W,C}$ для *max-пулинга* с окном размером $h \times w$, действующего в пространстве $\mathbb{R}^{H \times W \times 3}$, перепишем формулу (3) в виде:

$$Y = P_{h,w}^{H,W,C} X. \quad (11)$$

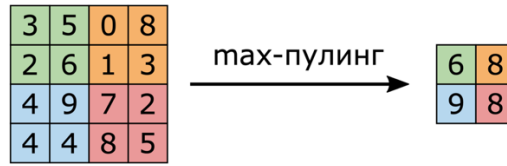


Рис. 2: Действие *max-пулинга* с размером окна 2×2

Функция активации ReLU, определённая в $\mathbb{R}^{H \times W \times C}$ - $ReLU^{H \times W \times C}$. Её действие на $X \in \mathbb{R}^{H \times W \times C}$ выражается в виде:

$$(ReLU^{H,W,C}(X))_{ijk} = \max\{0, X_{ijk}\}. \quad (12)$$

Ещё одна функция, используемая в свёрточных нейронных сетях – *softmax*^d, определённая на d -мерном пространстве и результат её действия на $z \in \mathbb{R}^d$ выражается формулой:

$$softmax^d(z)_j = \frac{e^{z_j}}{\sum_{i=1}^d e^{z_i}}, j = 1, \dots, d. \quad (13)$$

2.4 Архитектуры свёрточных нейронных сетей

На данном этапе работы обучены следующие архитектуры CNN: U-Net [5] (рис. 3), LinkNet[6] (рис. 4), PSPNet[7] (рис. 5) и модифицированная U-Net.

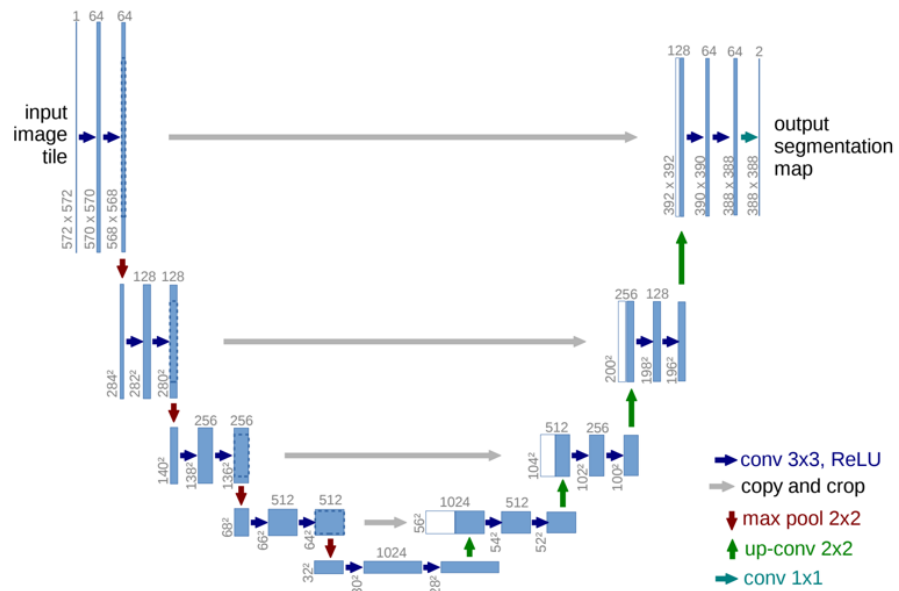


Рис. 3: Аproxимектура U-Net

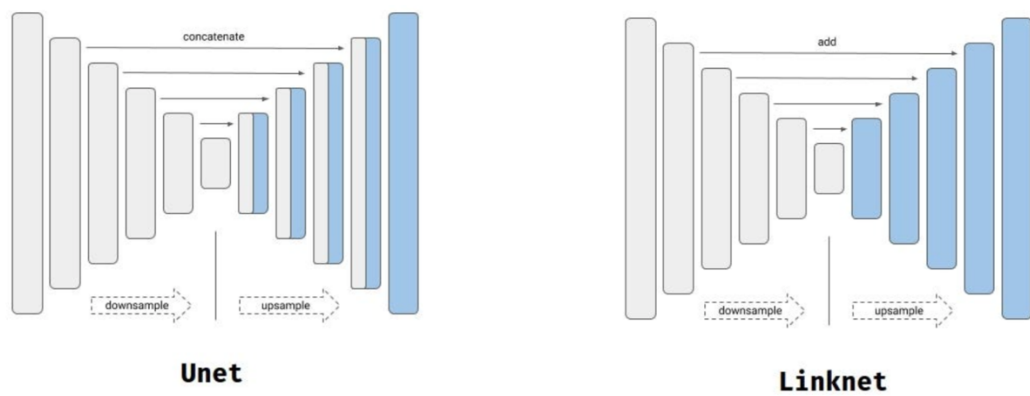


Рис.4: Аproxимектура LinkNet

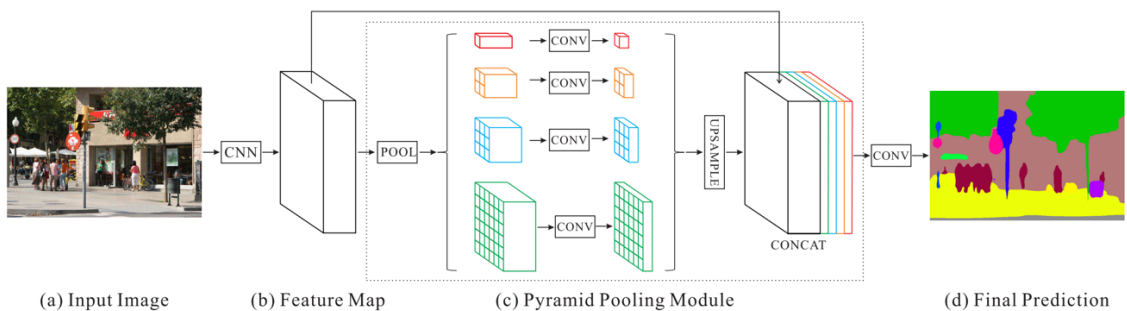


Рис. 5: Аproxимектура PSPNet

При решении задачи сегментации используют свёрточную часть из задачи классификации для получения карт признаков[6], которые затем увеличивают до исходных размеров (upsampling), с последующим обучением по картам сегментации.

2.5 Результаты обучения свёрточных нейронных сетей

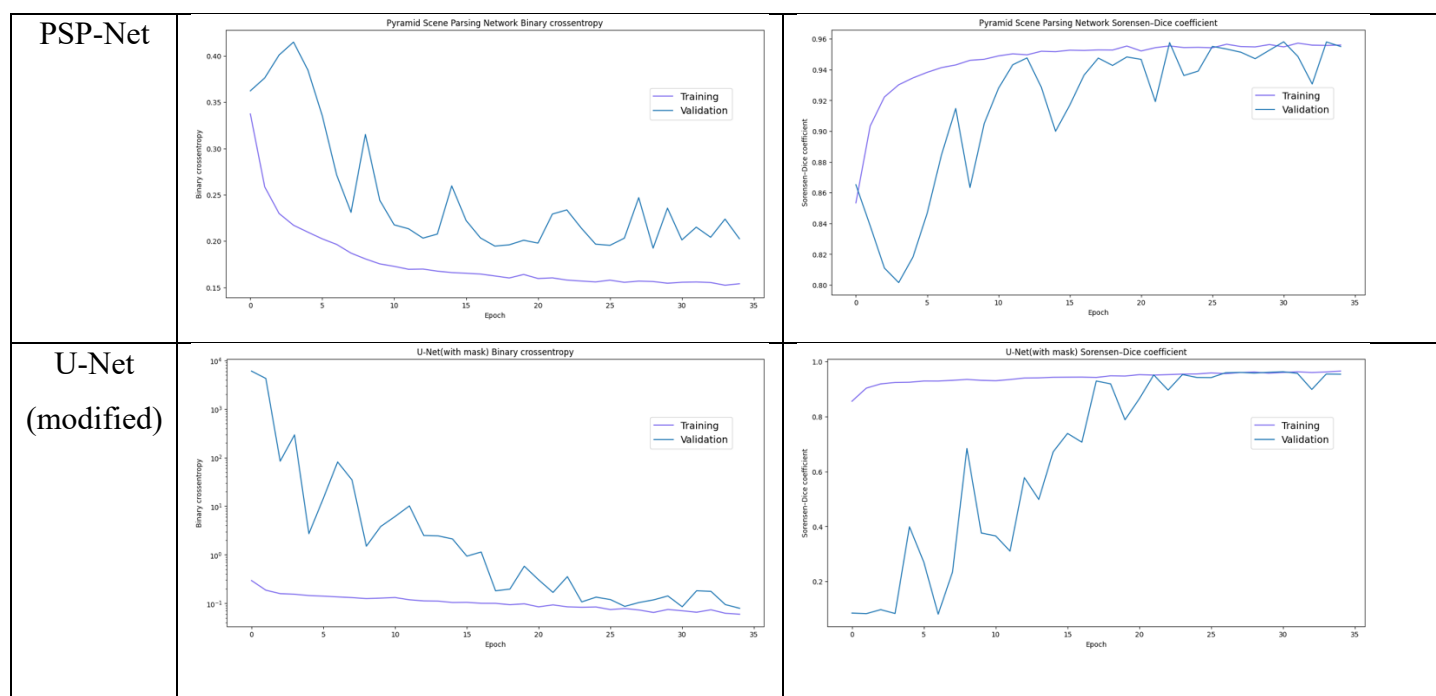
Минимизация функции потерь $\mathcal{L}(\theta)$ осуществлялась методом стохастического градиентного спуска на тренировочной выборке данных [9] на языке программирования «Python» средствами библиотеки «TensorFlow»[12]. В процессе обучения измерялся коэффициент Соренсена-Дайса(14) на валидационной выборке (таб. 1).

$$DSC = \frac{2|\hat{Y} \cap Y|}{|\hat{Y}| + |Y|}, \quad (14)$$

где $\hat{Y} = \{\hat{y}_j\}_{j=1}^M$ – множество выходных значений модели, $Y = \{y_i\}_{i=1}^N$ – множество исходных размеченных значений. $|\hat{Y}|$, $|Y|$ – мощности множеств.

Таблица 1: Binary cross-entropy и Sorensen–Dice coefficient в зависимости от эпохи обучения для разных архитектур FCNN

	Binary cross-entropy	Sorensen–Dice coefficient
U-Net		
LinkNet		



В таблице 2 представлены минимальные значения функции потерь и максимальные значения коэффициента Соренсена-Дайса.

Таблица 2: Значения функции потерь и коэффициента Соренсена-Дайса

	Binary cross-entropy (or MSE for U-Net(MSE))	Sorensen–Dice coefficient
U-Net	0.1163	0.9665
LinkNet	0.1026	0.9430
PSP-Net	0.2025	0.9551
U-Net (modified)	0.1446	0.9594
U-Net (MSE)	0.0258	0.9464

Для обучения U-Net помимо $Q(\theta) = H(p_{\text{data}}, q(\theta))$ использовался функционал $Q(\theta) = \text{MSE}$, сравнения результатов представлено в таблице 2.

2.6 Выводы

Исходя из коэффициентов оценки точности, полученных в результате экспериментов, свёрточные нейронные сети действительно хорошо справляются с задачей сегментации.

В данной работе использование в качестве эмпирического риска бинарной кросс-энтропии показало лучшие результаты, чем использование среднеквадратической ошибки, что согласуется с рекомендациями статьи [4].

Реализованная модифицированная архитектура U-Net не показала преимуществ.

3 Заключение

В ходе работы описана математическая модель задачи сегментации изображения, изучены архитектуры свёрточных нейронных сетей, таких как U-Net, LinkNet, PSPNet. А также построена и обучена модифицированная FCNN U-Net. Выполнено сравнение результатов работы программных реализаций этих моделей.

Список литературы

1. *Neural Networks: A Comprehensive Foundation by Simon Haykin*
2. *Математические методы обучения по прецедентам (теория обучения машин) К. В. Воронцов*
3. *Глубокие нейронные сети. Погружение в мир нейронных сетей. Екатерина Архангельская, Артур Кадури, Сергей Николенко, 2018*
4. *Семантическая сегментация спутниковых снимков аэропортов с помощью свёрточных нейронных сетей. В.А. Горбачёв, И.А. Криворотов, А.О. Маркелов, Е.В. Котлярова*
5. *U-Net: Convolutional Networks for Biomedical Image Segmentation*
6. *LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation*
7. *Pyramid Scene Parsing Network Hengshuang Zhao Jianping Shi Xiaojuan Qi Xiaogang Wang Jiaya Jia The Chinese University of Hong Kong SenseTime Group Limited, 2017*

8. *VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION* Karen Simonyan * & Andrew Zisserman + Visual Geometry Group, Department of Engineering Science, University of Oxford 2015
9. *Dataset–Busy-with-GT*
<https://www.kaggle.com/datasets/anaselmasry/datasetbusiwithgt>
10. Paclik P., Novovicova J., Duin R. P. W. *Building Road-Sign Classifiers Using a Trainable Similarity Measure // IEEE Transactions on Intelligent Transportation Systems.* - 2006. - T. 7, - №3. - C. 309-321.
11. *Deep Learning for MR Angiography: Automated Detection of Cerebral Aneurysms* / D. Ueda // *Radiology.* - 2019. - T. 290, № 1. - C. 187-194.
12. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* / Martin Abadi [и др.]. - 2015.