# **Введение**

Морской грузопоток быстро растет. Большее количество судов увеличивает вероятность возникновения происшествий в море, таких как экологически разрушительные аварии, катастрофы на судах, пиратство, незаконный лов рыбы, незаконный оборот наркотиков и незаконные перевозки грузов. Это вынудило многие организации, от природоохранных учреждений до страховых компаний и национальных государственных органов, более внимательно следить за обстановкой в открытом море.

В результате чего была поставлена задача нахождения кораблей на спутниковых снимках. С задачей распознавания объектов на изображении прекрасно справляются свёрточные нейронные сети. Сверточная нейронная сеть - специальная архитектура искусственных нейронных сетей, предложенная Яном Лекуном в 1988 году и нацеленная на эффективное распознавание образов, входит в состав технологий глубокого обучения. Использует некоторые особенности зрительной коры, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Таким образом, идея свёрточных нейронных сетей заключается в чередовании свёрточных слоёв и субдискретизирующих. Структура сети — однонаправленная, принципиально многослойная. Для обучения используются стандартные методы, чаще всего метод обратного распространения ошибки. Функция активации нейронов (передаточная функция) — любая, по выбору исследователя.

Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу - ядро свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

Работа свёрточной нейронной сети обычно интерпретируется как переход от конкретных особенностей изображения к более абстрактным деталям, и далее к ещё более абстрактным деталям вплоть до выделения понятий высокого уровня. При этом сеть самонастраивается и вырабатывает сама необходимую иерархию абстрактных признаков (последовательности карт признаков), фильтруя маловажные детали и выделяя существенное.

Для актуальности темы можно добавить, что данный тип нейронной сети сейчас активно используется компанией Google, тот же Google – переводчик, в котором присутствует функция перевода текста с изображения. Та же компания ввела в свой поисковый сервис функцию поиска изображений, аналогичных нашему. Ну и если перейти к более серьезным вещам, то сверточные сети используются для автопилотов в современных автомобилях, например, как в бортовом компьютере автомобилей компании Tesla.

Помимо абстрактных примеров, есть и конкретные приложения сверточных нейронных сетей в медицине. В статье [[1]](#footnote-1) применяется сверточная нейронная сеть для локализации пневмоторакса в легких на рентгеновских снимках, если он присутствует. Реализовано с помощью архитектуры нейронной сети U-Net[[2]](#footnote-2), которая на данный момент стала стандартной для решения задач сегментации изображений. Аналогично можно использовать U-Net и для обработки изображений для нахождения раковых опухолей, что и применяется в работе[[3]](#footnote-3).

# **Постановка задачи**

## **Задача классификации**

Как было описано выше, у человечества появилась потребность в отслеживании движений морских судов в море для тех или иных целей. Для успешного отслеживания судов было принято использовать спутниковые снимки, но ведь вряд ли найдется человек, который будет скрупулезно рассматривать спутниковые снимки в поисках корабля на изображении, а если даже и будет, то компания не станет тратить свои ресурсы для оплаты труда такого работника. Из этой ситуации существует только одно рациональное решение – доверить монотонную работу автоматике. Вычислительной машине подается изображение, а далее следует ответ, присутствует ли морское судно на снимке.

Формализуя, задача будет классификации выглядеть следующим образом. Дана выборка изображений , которую мы разделим на обучающую подвыборку и тестовую подвыборку так, что и . Так же мы делим множество правильных ответов на и так, что и . Итак, есть выборка изображений и выборка правильных ответов . Пусть – случайная величина, представляющая собой случайное изображение из . И пусть – случайная величина, представляющая собой случайный правильный ответ из . Тогда определим случайную величину c распределением , которое является совместным распределением объектов и их классов. Тогда размеченная выборка – это элементы из распределения . Определим, что все элементы независимо и одинаково распределены. Тогда задача классификации будет сведена к задаче нахождения и заданном наборе элементов .

С помощью обучающей выборки и правильных ответов будем находить распределение , а уже на тестовой выборке и наборе правильных ответов для нее, будем смотреть, как хорошо сверточная нейронная сеть может распознавать тестовые изображения, которые никогда не видела, натренированная на обучающей выборки.

## **Задача семантической сегментации**

Казалось бы, автоматизированная система для нахождения изображений с кораблями – это уже полезное нововведение. Осталось прикрепить к изображениям географические метки и уже будет получена окрестность, внутри которой будет находиться морское судно. Но разве нет способа получить точные координаты? Конечно есть, но для этого нужно определить точное расположение корабля на изображении.

Данная задача является задачей семантической сегментации изображения. Суть задачи в нашем случае в том, что нужно выделить одним цветом каждый пиксель изображения, где находится корабль и другим цветом выделить каждый пиксель того, что не является кораблем. Формальная постановка задачи будет аналогичная с постановкой нашей задачи для классификации, только лишь с тем различием, что каждый элемент не будет правильным ответом, есть ли корабль на изображении , а будет являться матрицей, содержащая в себе правильный ответ для каждого пикселя элемента .

## **Исходные данные**

Исходные данные представляет из себя набор изображений , где пример на рисунке 1.

Рисунок 1

Далее предоставляется таблица, содержащее имя изображения и перечисляются пиксели, где находится корабль, если он есть на изображении, пример на рисунке 2.

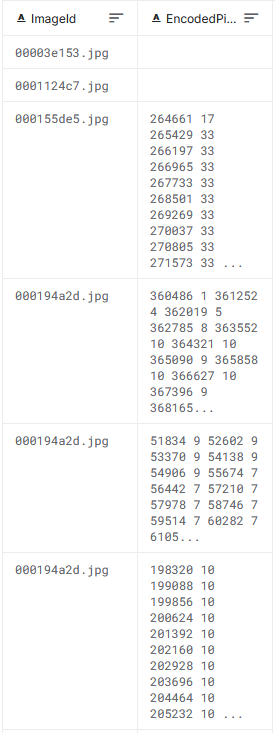


Рисунок 2

На основе этих данных будет производится обучение сверточной нейронной сети. Перед обучением производится операция перехода от изображения к его матричной форме. Пример матричной формы первого столбца для одного изображения в таблице 1.

|  |  |  |  |
| --- | --- | --- | --- |
| (w = 0, i) | Red | Green | Blue |
| i = 0 | 126 | 141 | 146 |
| i = 1 | 126 | 141 | 146 |
| i = 2 | 126 | 141 | 146 |
| ……. | | | |
| i = h - 3 | 117 | 134 | 138 |
| i = h - 2 | 115 | 133 | 137 |
| i = h - 1 | 116 | 134 | 138 |

Табл.1 – пример матричного представления изображения для обучения модели.

Для бинарной классификации, столбец пикселей корабля преобразуется в вектор – строку размерности два, пример для одного изображения на рисунке 3.

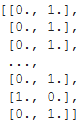


Рисунок 3

Для семантической сегментации столбец пикселей корабля заменяется на полную карту изображения, хранящая бинарное значение для каждого пикселя, являющееся признаком принадлежности к первому или второму классу. Пример карты для одного изображения на рисунке 4.



Рисунок 4

## **Выбор метрики**

Итак, определив наши задачи нужно понять, как же оценивать полученные результаты. Наиболее интуитивной для задачи бинарной классификации является такая метрика точность (accuracy). Чтобы определить ее и последующие метрики, введем карту обозначений, приведенную в таблице 1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Прогноз | |
| + | - |
| Правильный ответ | + | TP (True – positives) | FN (False – negatives) |
| - | FP (False – positives) | TN (True – negatives) |

Табл.2 – карта обозначений.

Определим обозначения, в нашем случае, как TP (True-positives) – число раз, когда корабль действительно опознан на изображении, FP (False – positives) – когда корабль ошибочно опознан и FN (False – negatives) – когда корабль присутствие корабля ошибочно отрицается.

Тогда точность можно определить, как:

.

Как можно заметить по выражению точности, данная метрика не позволит корректно оценить качество прогноза, если в выборка есть перевес классов ту или иную сторону. Например, если выборка из 1000 элементов будет иметь 990 экземпляров с отрицательным ответом и 10 с положительным, то данная метрика не позволит оценить, как хорошо мы прогнозируем истинно положительный ответ, ведь точность будет представлять собой долю правильных ответов в 99%. Получается, что доля положительных ответов не представляет из себя никакой информации о качестве прогнозирования. Однако, если есть возможность сбалансировать выборку, то данная метрика будет очень информативной.

Но все же, если нет возможности сбалансировать выборку, лучше перейти к другим метрикам, таким как точность (precision) и полнота (recall):

,

.

Точность (precision) показывает, какая доля положительных прогнозов является истинно положительными. Полнота (recall) же показывает, какая доля положительных ответов была определена.

На основании двух вышеописанных метрик можно построить график зависимости точности (precision) от полноты (recall), называемый PR – кривой. Данный график будет давать представление о качестве прогнозов и для несбалансированной выборки. Далее находим площадь под этой кривой и получаем количественное значение, представляющее собой меру качества прогнозов в сравнении с истинными ответами.

Далее нужно определить метрику для задачи семантической сегментации. Так как нам априори известно, что на изображениях будет преимущественно преобладать площадь морской глади над площадью морского судна, то требуется мера, которая будет противостоять данному, достаточно внушительному, перевесу. С данной проблемой отлично справится метрика, называемая коэффициентом Соренсена – Дайса (DICE коэффициент). Определяется она как:

,

где А – прогнозируемое множество, а B – множество истинных ответов. Но для нашего, бинарного, случая метрика принимает более удобный вид:

.

Данная метрика демонстрирует успешную борьбу с преобладанием одного класса над другим на одном изображении, что доказывается в очередной медицинской статье по семантической сегментации[[4]](#footnote-4).

## **Описание экспериментальной установки**

Сверточная нейронная сеть реализована на высокоуровневом языке программирования Python, с использованием встроенных библиотек: keras, scipy, numpy, tensorflow. Данные программные библиотеки позволяют строить нейронные сети, проверять те или иные гипотезы, не затрачивая время на программную реализацию самого механизма проектирований модели и на реализацию обучения модели. Запуск программы совершен на бесплатной облачной платформе Google Colab, которая позволяет совершать неподъемные для домашней машины вычисления.

Несмотря на преимущества облачных вычислений, бесплатная платформа имеет ограничения по объему файлового хранилища и времени работы программы. Время работы не должно превышать восьми часов, а объем файлового пространства не больше 15 Гб. Итак, фактор времени является наиболее весомым, так как время работы программы тратится не только на обучение модели сверточной нейронной сети, но и на предобработку данных. Поэтому решено задать количество изображений для исследования гипотез равное 20000, из которых 16000 непосредственно для обучения модели, а 4000 для проверки качества прогнозов на изображениях, которые модель никогда не обрабатывала. Выборка данного объема сбалансирована для бинарного классификатора и все элементы подобраны независимо друг от друга, что дает уверенность, что при исследовании, результаты для тех или иных гипотез будут релевантные относительно друг друга.

Обучение модели будет проходить на протяжении 50 эпох с объемом пакета одновременной обработки в 256 изображений.

## **Построение базового решения**

Для исследования результатов работы сверточной нейронной сети требуется определить асимптоты. Так как для решения задачи семантической сегментации морских судов нужно полностью освоить упрощенную задачу в лице бинарной классификации изображений, следует задать граничное значение метрик, результаты гипотез ниже которых сразу будут отклонены.

Для задания минимального значения метрик будем использовать генератор случайного подбрасывания монеты, где результатом будет равновероятный исход орел или решка, интерпретируемые как нуль и единица. Проведем сравнение результатов генерации и истинных ответов на прежде описанной выборке в 20000 изображений и найдем для этого такие метрики, как: точность (accuracy), точность (precision), полнота (recall).

Результаты эксперимента приведены на рисунке 5.

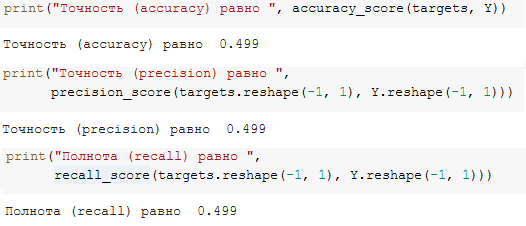


Рисунок 5

Итак, если округлить, то нижний порог для результатов гипотез составляет 0.5 для точности (accuracy), 0.5 для точности (precision), 0.5 для полноты (recall). Гипотезы с результатами ниже данного порога отклоняются.

## **Решение с использованием человеческих ресурсов**

Для обоснования использования сверточной нейронной сети на практике нужно проверить, будет ли это решение более оптимальным, нежели человек сам будет помечать изображения с кораблями.

Для нахождения метрик качества определения класса изображения человеком создадим некий тестовый стенд. Для этого нам понадобится стационарный компьютер, клавиатура, монитор и программа, реализующая вывод изображения на экран и ввод ответа в бинарном виде.

Проведя данное исследование с использованием 100 изображений, за 3 минуты были получены ответы и следующие метрики качества ответов в сравнении с истинными. Результаты на рисунке 6.

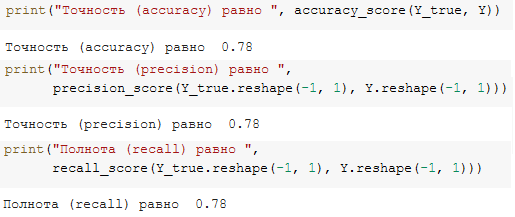


Рисунок 6

Итак, метрики качества ответов человека на 100 изображениях за 3 минуты составляет 0.78 для точности (accuracy), 0.78 для точности (precision), 0.78 для полноты (recall). Гипотезы с результатами выше данного порога будут считаться оптимальными.

# **Исследование гипотез**

Для исследований мы не будем составлять свою архитектуру сети, а позаимствуем уже готовую, под названием «VGG16». «VGG16» — модель сверточной нейронной сети, предложенная K. Simonyan и A. Zisserman из Оксфордского университета в статье “Very Deep Convolutional Networks for Large-Scale Image Recognition”. Модель достигает точности 92.7% в задаче распознавания объектов на 14 миллионов изображений, принадлежащих к 1000 классам. Для нашей модели мы сократим размеры входных данных из-за технических ограничений и для решения нашей задачи бинарной классификации мы сократим количество классов с тысячи до двух классов. Структура данной модели изображена на рисунке 7:

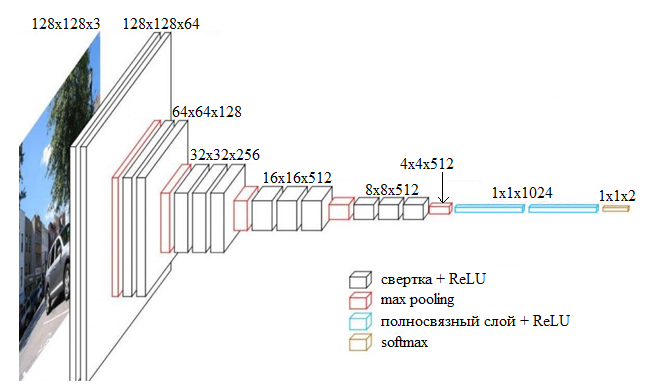


Рисунок 7

Так как используется изображение в формате Red-Green-Blue, каждый из трех каналов цвета обрабатывается отдельно, то сеть принимает три двумерной матрицы 128х128 c интенсивностями цвета. Если обратить внимание на легенду структуры на рисунке 6, можно заметить, что помимо обозначения самого слоя мы также обозначаем и функцию активации, определенную на слое.

Начнем с функции активации ReLU (англ. Rectified linear unit) или иначе линейный выпрямитель. Функция определяется как

и имеет график, изображенный на рисунке 8:

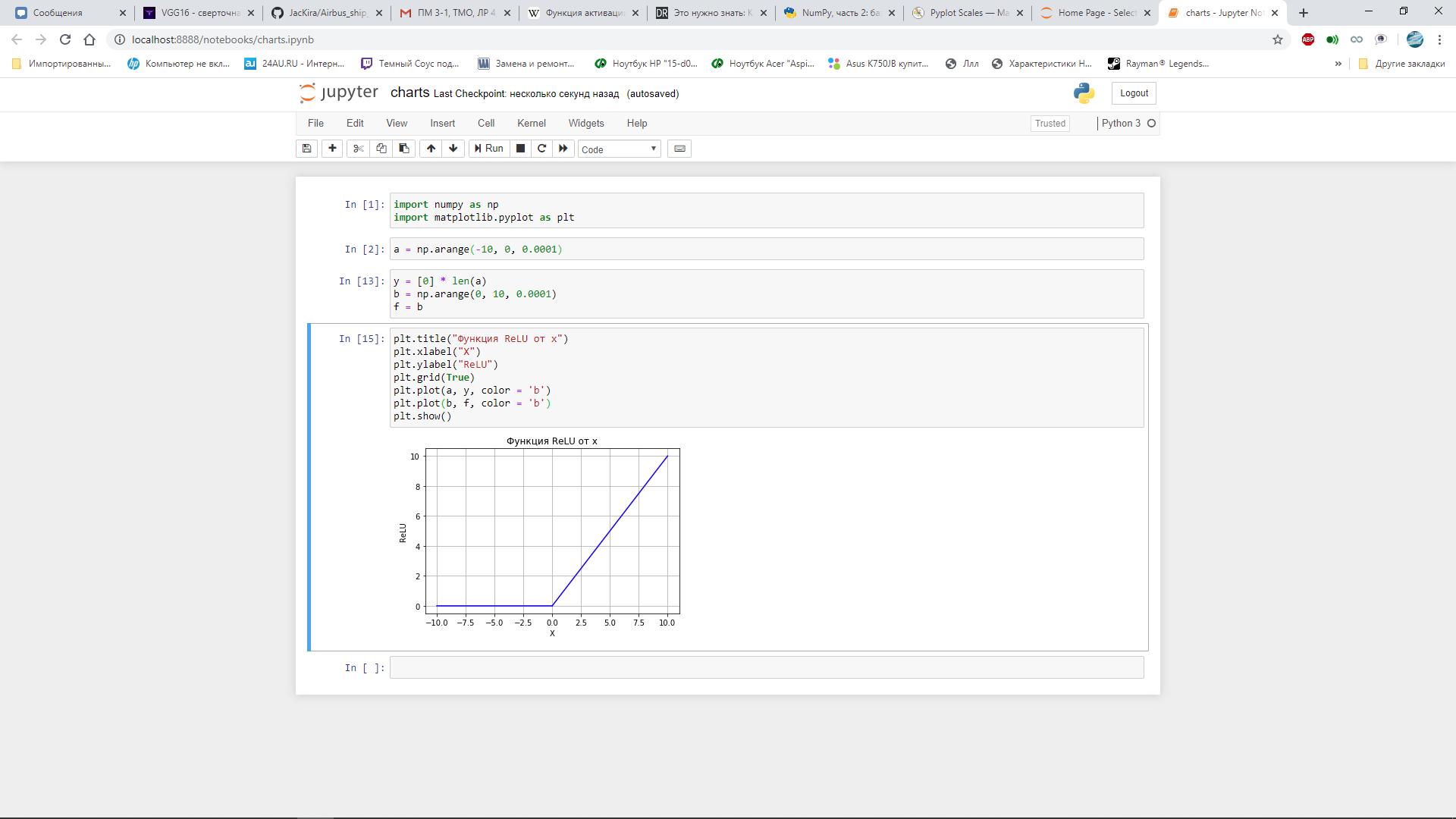


Рисунок 8

Сущность функции предельно понятна, если функция активации принимает отрицательное значение, она возвращает ноль и данное значение не искажает данные для дальнейшей обработки, если иначе, то значение следует в дальнейшие слои. Преимущество этой функции активации в том, что вычисление ReLU реализовано с помощью простого порогового преобразования матрицы активаций в нуле. Также к преимуществу можно отнести то, что ReLU не подвержен насыщению. Применение ReLU существенно повышает скорость сходимости стохастического градиентного спуска по сравнению с гиперболическим тангенсом. Считается, что это обусловлено линейным характером и отсутствием насыщения данной функции. К сожалению, ReLU не всегда достаточно надежны и в процессе обучения могут выходить из строя («умирать»). Слишком большой градиент, проходящий через ReLU, приводит к такому обновлению весов, что данный нейрон никогда больше не активируется. Если это произойдет, то, начиная с данного момента, градиент, проходящий через этот нейрон, всегда будет равен нулю. Соответственно, данный нейрон будет необратимо выведен из строя. Например, при слишком большой скорости обучения может оказаться, что до 40% нейронов с ReLU «мертвы» или иначе говоря, больше не когда не активируются. Эта проблема решается посредством подбора приемлемой скорости обучения.

Далее разберем функцию активации softmax.Softmax – это обобщение логистической функции для многомерного случая. Функция преобразует ***z*** – вектор ***K***{\displaystyle K} – {\displaystyle z}размерности в – вектор {\displaystyle \sigma }той же размерности, где каждая {\displaystyle \sigma \_{i}} – координата полученного вектора представлена вещественным числом в интервале [0,1] и сумма координат равна 1.

Координаты {\displaystyle \sigma \_{i}} {\displaystyle \sigma \_{i}} вычисляются следующим образом:

, где

, где

x – вектор – столбец признаков объекта размерности , а M – количество признаков объектов. – транспонированная матрица весовых коэффициентов признаков, имеющая размерность , а – вектор – столбец с пороговыми значениями размерностью , где K – количество классов объектов, в нашем случае K = 2.

Рассмотрим пример работы функции softmax на выходном слое сети для нашей задачи. Пусть i = 1 – отсутствие корабля на изображении, а i = 2 – присутствие. Допустим, что нашли вектора z по формуле

вида

*.*

Тогда имеем

и получаем

,

выбирая максимальное значение получаем, что с вероятностью в 99% корабль присутствует на фотографии, то есть результатом обработки изображения сетью будет 1 – корабль присутствует на фотографии.

Теперь перейдем к функции потерь. Так как на выходном слое используется softmax, для определения потерь с ней используют функцию перекрестной энтропии, которая имеет вид:

.

Свойства данной функции:

1. Минимум функции достигается при

Так как мы будем обучать нашу сеть не по одной фотографии за раз, а сразу на нескольких, то функция потерь для всей выборки находится как:

.

## **Опорная гипотеза**

Так как задача машинного обучения – это прежде всего задача оптимизации, то и результаты обучения во многом зависят от методов нахождения оптимального решения.

За опорное решение возьмем результаты прошлых исследований, где использовался метод оптимизации RMSProp.

RMSProp (от англ. Root Mean Square Propagation) среднеквадратичное распространение — это метод, в котором скорость обучения настраивается для каждого параметра. Идея заключается в делении скорости обучения для весов на сгруппированные средние значения градиентов для этого веса. Таким образом, первое сгруппированное среднее вычисляется в терминах среднеквадратичного. Как и все градиентные методы, этот имеет общий вид:

,

где – матрица весов нейронной сети, а – направление для минимизации целевой функции, а h – шаг или в нашем случае это скорость обучения. Вся суть метода в том, как мы находим наше направление.

Возьмем нашу скорость h и скорость затухания (параметр сглаживания для экстраполяции). Определим наше начальное значение весов как нули. И определим малую константу чтобы избежать деления на нуль в будущем. Задаем параметр для агрегирования градиента r = 0. И пока условие остановки не выполнено, повторяем следующие шаги:

1. Выбираем часть экземпляров количества m

из нашей выборки и соответствующие им

1. Находим градиент

,

Где– целевая функция, а – функция активации.

1. Агрегируем квадраты градиента

где – это поэлементное произведение матриц градиента.

1. Вычисляем направление

,

где операция применяется к каждому элементу матрицы.

1. Находим новое значение

Условием остановки алгоритма может быть оптимальное значение целевой функции или ситуация, когда веса практически не меняются.

Эмпирически показано, что RMSprop - эффективный и практичный алгоритм оптимизации глубоких нейронных сетей. В настоящее время он считается одним из лучших методов оптимизации.[[5]](#footnote-5)

Проведя обучение модели на протяжении 50 эпох, с размером пакета одновременной обработки изображений в 256 элементов были получены следующие результаты, описанные ниже.

Как можно судить по графику показаний функции потерь на рисунке 9, к 30 – й эпохе модель содержит уже достаточно оптимальные значения параметров и дальше происходит уже топтание вокруг оптимума, что иногда приводит к аномальному росту значения ошибки.

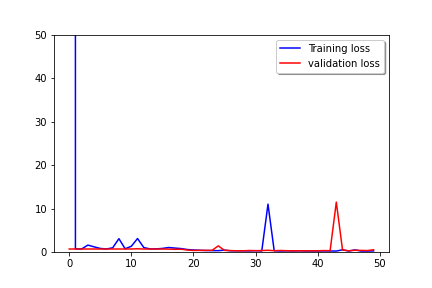


Рисунок 9

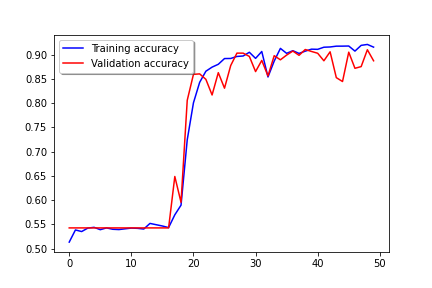
 Аналогичная картина наблюдается и с графиком метрик, в частности точности (accuracy), максимальное значение которой достигается еще до 30 – эпохи и дальше происходит уже брожение вокруг оптимума с последующими аномальными скачками. Но очень положительным моментом является то, что значение метрики как для тренировочной выборки, так и для тестовой подобны, что говорит об отсутствии переобучения модели. Также стоит отметить, что во время обучения модели точность достигла своей заложенной асимптоты.

Рисунок 10

Площадь под кривой PR так же достигает своего максимума к 30 – й эпохе, а график метрики на тренировочной выборке так же подобен графику метрики на тестовой выборке, что еще раз доказывает отсутствие переобучения.

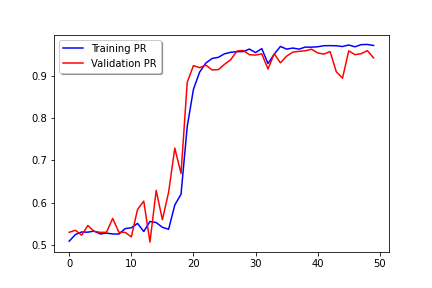


Рисунок 11

1. the 2st-unet for pneumothorax segmentation in chest x-rays using resnet34 as a backbone for u-net. arXiv:2009.02805v1 [eess.IV] 6 Sep 2020 [↑](#footnote-ref-1)
2. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 [cs.CV] 18 May 2015 [↑](#footnote-ref-2)
3. Colorectal Cancer Segmentation using AtrousConvolution and Residual Enhanced UNet. arXiv:2103.09289v1 [eess.IV] 16 Mar 2021 [↑](#footnote-ref-3)
4. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv:1606.04797v1 [cs.CV] 15 Jun 2016 [↑](#footnote-ref-4)
5. Глубокое обучение / Ян Гудфеллоу, Иошуа Бенджио, Аарон Курвилль // ДМК Пресс, 2018г., второе цветное издание, исправленное. [↑](#footnote-ref-5)