

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

Национальный исследовательский ядерный университет «МИФИ»



Институт интеллектуальных кибернетических  
систем

Кафедра №22 «Кибернетика»

Направление подготовки 09.03.04 Программная инженерия

## Пояснительная записка

к учебно-исследовательской работе студента на тему:

Разработка средств анализа эмоциональной окраски данных с использованием нейронных сетей

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
Национальный исследовательский ядерный университет «МИФИ»



Институт интеллектуальных кибернетических систем

КАФЕДРА КИБЕРНЕТИКИ

## Задание на УИР

Студенту гр. Б18-504  
(группа)

Борисова Анастасия Александровна  
(фио)

### ТЕМА УИР

Разработка средств анализа эмоциональной окраски данных с использованием нейронных сетей

Б18-504

Группа

Борисова А.А.

№ п/п	Содержание работы	Форма отчетности	Срок ис- полнения	Отметка о выполнении Дата, подпись
1.	<b>Аналитическая часть</b>			
1.1.	Изучение проблемы распознавания эмоций и обзор публикаций на тему чтения эмоциональной окраски данных	Пункт ПЗ	01.03.2021	
1.2.	Анализ подходов к обработке звука	Пункт ПЗ	02.03.2021	
1.3.	Анализ источников аннотированных данных	Пункт ПЗ	03.04.2021	
1.4.	Анализ публикаций на тему распознавания эмоциональной окраски сообщений разных типов	Пункт ПЗ	03.04.2021	
1.5.	<i>Оформление расширенного содержания пояснительной записки (РСПЗ)</i>	Текст РСПЗ	03.04.2021	

2.	<b>Теоретическая часть</b>			
2.1.	Проектирование архитектуры системы	Пункт ПЗ	09.04.2021	
2.2.	Выбор базы данных для обучения системы	Пункт ПЗ	15.04.2021	
3.	<b>Инженерная часть</b>			
3.1.	Проектирование архитектуры системы	Пункт ПЗ	17.04.2021	
3.2.	Выбор архитектуры, разработка и обучение нейронных сетей	Схема, Текст ПЗ	20.04.2021	
3.3.	Сравнение решений на основе различных нейронных сетей	Пункт ПЗ	22.04.2021	
3.4.	Выбор методов и инструментальных средств для реализации интерфейса системы	Пункт ПЗ	25.04.2021	
4.	<b>Технологическая и практическая часть</b>			
4.1.	Разработка системы распознавания эмоций в сообщениях разных типов	Исходный код	31.04.2021	
4.2.	Разработка тестовых примеров и тестирование системы	Исполняемые файлы, исходные тексты тестов и тестовых примеров	05.04.2021	
4.3.	Реализация интерфейса системы	Исходный код, Текст ПЗ	10.04.2021	
4.4.	Тестирование прототипа программного модуля системы распознавания эмоций в сообщениях	Исходный код	15.05.2021	
5.	<i>Оформление пояснительной записки (ПЗ) и иллюстративного материала для доклада.</i>	Текст ПЗ, презентация	17.05.2021	

## ЛИТЕРАТУРА

1.	Статья “Введение в задачу распознавания эмоций” <a href="https://habr.com/ru/company/speechpro/blog/418151/">https://habr.com/ru/company/speechpro/blog/418151/</a>
2.	Сергунов Дмитрий Игоревич, Артемова Анна Александровна, Гришунов Степан Сергеевич “Система распознавания эмоций по голосу на основе сверточной нейронной сети”
3.	Тимофеева Ольга Павловна, Неимущев Сергей Андреевич, Неимущева Любовь Ивановна, Тихонов Иван Александрович Распознавание эмоций по изображению лица на основе глубоких нейронных сетей // Труды НГТУ им. Р. Е. Алексеева. 2020. №1 (128).
4.	Ivanov I.A. Multi-objective based feature selection and neural networks ensemble method for solving emotion recognition problem
5.	Speech Emotion Recognition with Convolutional Neural Network by Reza Chu Перевод: <a href="https://habr.com/ru/company/Voximplant/blog/461435/">https://habr.com/ru/company/Voximplant/blog/461435/</a>
6.	Электронный ресурс <a href="https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1">https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1</a>
7.	Головко Владимир Адамович, Голенков Владимир Васильевич, Ивашенко Валерьян Петрович, Таберко Валерий Васильевич, Шаток Дмитрий Сергеевич, Крощенко Александр Александрович, Ковалёв Михаил Владимирович Интеграция искусственных нейронных сетей с базами знаний // Онтология проектирования. 2018.
8.	Теория базовых эмоций Пола Экмана  Ekman, P. (1992a). An Argument for Basic Emotions. Cognition and Emotion, 6(3/4), 169-200.

9.	Ekman, P. (1992b). Are There Basic Emotions? Psychological Review, 99(3), 550-553.

Дата выдачи задания:

Руководитель

Тихомирова Д.В.

(ФИО)

« 31 » марта 2021г.

Студент

Борисова А.А.

(ФИО)

## Реферат

Пояснительная записка содержит 28 страницы, 5 графиков, 2 таблицы, 12 приложений, 15 использованных источников.

Основные ключевые слова, описывающие тематику работы: распознавание эмоций, системы машинного зрения

В первом разделе проводится анализ рынка систем распознавания эмоций, а также анализ развития компьютерного зрения. Также рассматриваются существующие аннотированные базы данных эмоций.

Во втором разделе описываются базы изображений, изучаются проблемы распознавания эмоций, а также формулируются достоинства и недостатки систем машинного зрения.

В третьем разделе изучаются методы распознавания лиц и решаются их задачи на основе нейронных сетей.

В четвертом разделе приводятся результаты программной реализации системы.

## Оглавление

Введение .....	5
1.1 Изучение проблемы распознавания эмоций и обзор публикаций на тему чтения эмоциональной окраски данных .....	6
Анализ источников аннотированных данных .....	6
Анализ публикаций на тему распознавания эмоциональной окраски сообщений разных типов .....	8
Структура нейронных сетей .....	8
Нейросетевые алгоритмы .....	10
Свёрточная нейронная сеть .....	13
Устаревшие решения для компьютерного зрения .....	13
Проектирование нейронной сети .....	22
Заключение .....	25
Список литературы .....	26

## Введение

Нейронные сети - это искусственные, многослойные высокопараллельные логические структуры, составленные из формальных нейронов. Начало теории нейронных сетей и нейрокомпьютеров положила работа американских нейрофизиологов У. Мак-Каллока и У. Питтса «Логическое исчисление идей, относящихся к нервной деятельности» (1943), в которой они предложили математическую модель биологического нейрона. Среди основополагающих работ следует выделить модель Д. Хэбба, который в 1949 г. предложил закон обучения, явившийся стартовой точкой для алгоритмов обучения искусственных нейронных сетей. На дальнейшее развитие теории нейронной сети существенное влияние оказала монография американского нейрофизиолога Ф. Розенблатта «Принципы нейродинамики», в которой он подробно описал схему перцептрона (устройства, моделирующего процесс восприятия информации человеческим мозгом). Его идеи получили развитие в научных работах многих авторов. В 1985–86 гг. теория нейронных сетей получила «технологический импульс», вызванный возможностью моделирования нейронных сетей на появившихся в то время доступных и высокопроизводительных персональных компьютерах. Теория нейронной сети продолжает достаточно активно развиваться в начале 21 века. По оценкам специалистов, в ближайшее время ожидается значительный технологический рост в области проектирования нейронных сетей и нейрокомпьютеров. За последние годы уже открыто немало новых возможностей нейронных сетей, а работы в данной области вносят существенный вклад в промышленность, науку и технологии, имеют большое экономическое значение.

### ***Раздел 1. Предмет исследования.***

Предметом исследования являются средства анализа эмоций и нейронные сети.

## Изучение проблемы распознавания эмоций и обзор публикаций на тему чтения эмоциональной окраски данных

Рынок систем детекции и распознавания эмоций (EDRS) активно развивается. По оценкам ряда экспертов, он продемонстрирует среднегодовой рост в 27,4% и достигнет планки 29,1 млрд долларов к 2022 году. Такие цифры вполне оправданны, поскольку программное обеспечение для распознавания эмоций уже позволяет определять состояние пользователя в произвольный момент времени при помощи веб-камеры или специализированного оборудования, параллельно анализируя поведенческие паттерны, физиологические параметры и изменения настроения пользователя.

Потенциальными областями применения искусственных нейронных сетей являются те, где человеческий интеллект малоэффективен, а традиционные вычисления трудоёмки или физически неадекватны (т. е. не отражают или плохо отражают реальные физические процессы и объекты). Актуальность применения нейронных сетей (т. е. нейрокомпьютеров) многократно возрастает, когда появляется необходимость решения плохо формализованных задач. Основные области применения нейронных сетей: автоматизация процесса классификации, автоматизация прогнозирования, автоматизация процесса распознавания, автоматизация процесса принятия решений; управление, кодирование и декодирование информации; аппроксимация зависимостей и др.

### Анализ источников аннотированных данных

Взгляд Пола Экмана на термин «базовые эмоции» включает две характеристики: (1) эмоции дискретны (могут быть принципиально отличны друг от друга по выражению лица, физиологии и причинам их вызывающие); (2) и эмоции, возникли во время адаптации, как филогенетической, так и онтогенетической (Ekman & Cordaro, 2011). Третья характеристика, часто обозначаемая фразой «базовые эмоции», состоит в том, что другие не-базовые эмоции представляют собой смеси или сочетания базовых эмоций, но Экман (1992а, сноска на стр.170) отказывается обсуждать эту третью характеристику и указывает, что нынешние исследователи, изучающие базовые эмоции не используют этот термин таким образом.

Перечень критериев, которые свойственны (почти) всем базовым эмоциям, согласно точки зрения Пола Экмана:

1. Отличительные универсальные сигналы.
2. Отличительная физиология.
3. Автоматическая оценка.
4. Отличительные универсалии в предшествующих событиях
5. Присутствие у других приматов
6. Возможность быстрого начала
7. Может быть недлительной
8. Незванное появление
9. Отличительные мысли, воспоминания и образы
10. Отличительный субъективный опыт
11. Фильтрации информации, допуская информацию поддерживающую эмоцию
12. Объект эмоции не ограничен

### 13. Эмоция может быть осуществлена как конструктивно, так и разрушительно.

Этот список помогает различать эмоции от других эмоционально насыщенных аффективных явлений (настроения, эмоциональные характеристики, эмоциональные расстройства и т.д.; Ekman, 1999). Более того, если эмоция соответствует критериям из списка, тогда это не просто эмоция, но и «базовая» эмоция. По определению, если эмоция соответствует всем (или почти всем критериям Ekman & Cordaro, 2011) — это базовая эмоция. Другими словами, в его теории любая эмоция является базовой, или это не эмоция вообще (Ekman & Cordaro, 2011).

#### Эмоциональные семейства

Еще один важный момент в его взгляде это то, что «эмоции — это не одно аффективное или психологическое состояние, а семья родственных состояний» (pp364, Ekman & Cordaro, 2011). Перечень критериев, перечисленных выше, распространяется на всю семью и позволяет также различать семьи. Термин «тема», используемый для описания совокупности характеристик уникальных для семьи, которые разделяют все члены этой семьи. Темы не только общая черта эмоций, но и их основной элемент, продукт эволюции, присущий всем эмоциям, его филогенетическая часть. Вариации относятся к различиям членов семьи. Вариации отвечают за индивидуальные различия, результат социального опыта или обучения (Ekman, 1992a, 1997; Ekman & Cordaro, 2011). Экман описал свой взгляд на примере гнева, где вариации интенсивности могут быть между досадой и яростью (Ekman, 1997). В данном примере, гнев — семья, которая характеризуется набором общих свойств — темой, а досада и ярость — вариации семьи гнева.



# Анализ публикаций на тему распознавания эмоциональной окраски сообщений разных типов

## Структура нейронных сетей

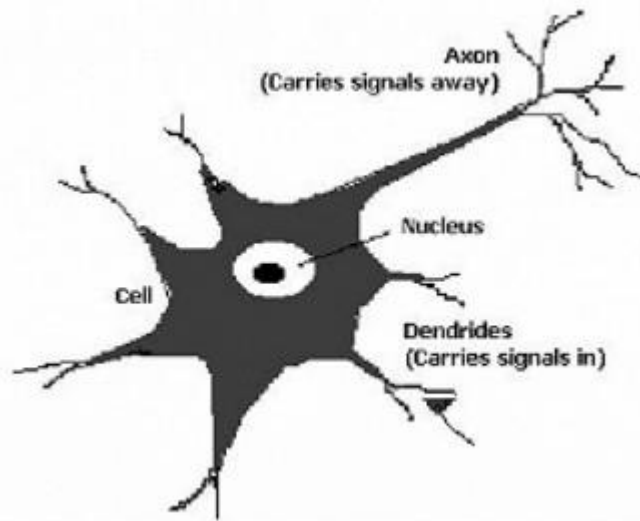


Рис. Биологический (или естественный) нейрон

Нервная система и мозг человека состоят из нейронов, соединённых между собой нервными волокнами. Нервные волокна способны передавать электрические импульсы между нейронами. Все процессы передачи раздражений от нашей кожи, ушей и глаз к мозгу, процессы мышления и управления действиями – всё это реализовано в живом организме как передача электрических импульсов между нейронами.

Биологический нейрон (Cell) имеет ядро (Nucleus), а также отростки нервных волокон двух типов (рис. 1) – дендриты (Dendrites), по которым принимаются импульсы (Carries signals in), и единственный аксон (Axon), по которому нейрон может передавать импульс (Carries signals away). Аксон контактирует с дендритами других нейронов через специальные образования – синапсы (Synapses), которые влияют на силу передаваемого импульса. Структура, состоящая из совокупности большого количества таких нейронов, получила название биологической (или естественной) нейронной сети.

Появление формального нейрона во многом обусловлено изучением биологических нейронов. Формальный нейрон (далее – нейрон) является основой любой искусственной нейронной сети. Нейроны представляют собой относительно простые, однотипные элементы, имитирующие работу нейронов мозга. Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены и заторможены. Искусственный нейрон, так же как и его естественный прототип, имеет группу синапсов (входов), которые соединены с выходами других нейронов, а также аксон – выходную связь данного нейрона, откуда сигнал возбуждения или торможения поступает на синапсы других нейронов.



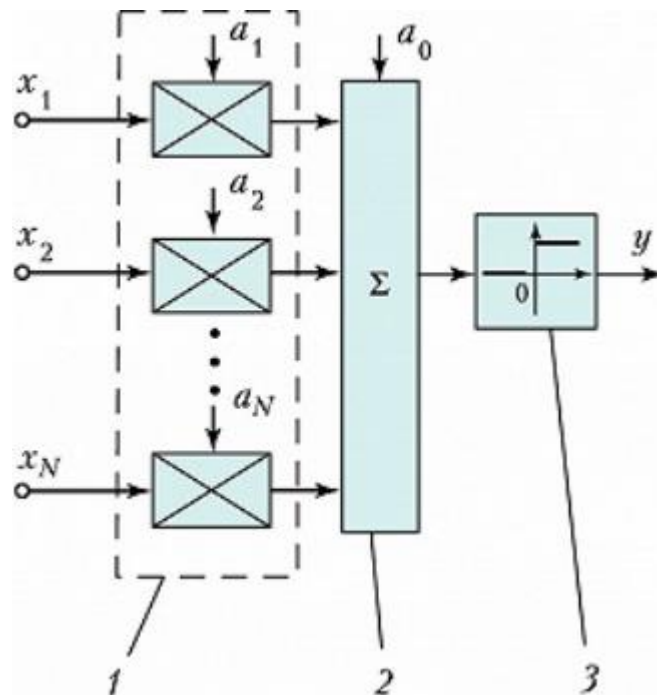


Рис. 1. Функциональная схема нейрона: 1 – блок входа; 2 – сумматор; 3 – блок нелинейного преобразования.

Выбор структуры нейронной сети осуществляется в соответствии с особенностями и сложностью задачи. Теоретически число слоёв и число нейронов в каждом слое нейронной сети может быть произвольным, однако фактически оно ограничено ресурсами компьютера или специализированной микросхемы, на которых обычно реализуется нейронная сеть. При этом если в качестве активационной функции для всех нейронов сети используется функция единичного скачка, нейронная сеть называется многослойным персептроном.

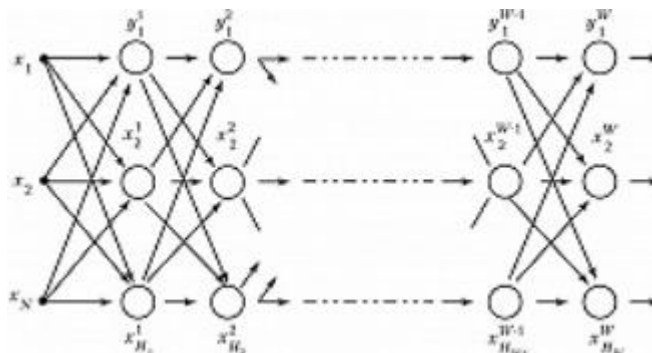


Рис. 3. Граф многослойной нейронной сети с последовательными связями:  $x_i$  –  $i$ -й входной сигнал ( $i=1...N$ );  $H_1$  – число нейронов 1-го слоя;  $W$  – число слоёв;  $x_{H_W}^W$  – выходной...

На рис. 3 показана общая схема многослойной нейронной сети с последовательными связями. Высокий параллелизм обработки достигается путём объединения большого числа формальных нейронов в слои и соединения определённым образом различных нейронов между собой.

В общем случае в эту структуру могут быть введены перекрёстные и обратные связи с настраиваемыми весовыми коэффициентами (рис. 4).

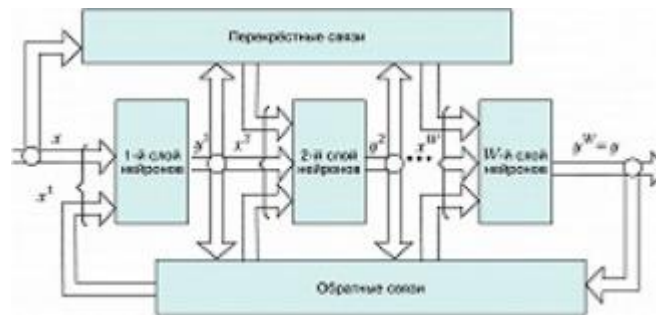


Рис. 3. Многослойная нейронная сеть.

Нейронные сети являются сложными нелинейными системами с огромным числом степеней свободы. Принцип, по которому они обрабатывают информацию, отличается от принципа, используемого в компьютерах на основе процессоров с фон-неймановской архитектурой – с логическим базисом И, ИЛИ, НЕ (см. Дж. фон Нейман, Вычислительная машина). Вместо классического программирования (как в традиционных вычислительных системах) применяется обучение нейронной сети, которое сводится, как правило, к настройке весовых коэффициентов с целью оптимизации заданного критерия качества функционирования нейронной сети.

## Нейросетевые алгоритмы

Нейросетевым алгоритмом решения задач называется вычислительная процедура, полностью или по большей части реализованная в виде нейронной сети той или иной структуры (например, многослойная нейронная сеть с последовательными или перекрёстными связями между слоями формальных нейронов) с соответствующим алгоритмом настройки весовых коэффициентов. Основой разработки нейросетевого алгоритма является системный подход, при котором процесс решения задачи представляется как функционирование во времени некоторой динамической системы. Для её построения необходимо определить: объект, выступающий в роли входного сигнала нейронной сети; объект, выступающий в роли выходного сигнала нейронной сети (например, непосредственно решение или некоторая его характеристика); желаемый (требуемый) выходной сигнал нейронной сети; структуру нейронной сети (число слоёв, связи между слоями, объекты, служащие весовыми коэффициентами); функцию ошибки системы (характеризующую отклонение желаемого выходного сигнала нейронной сети от реального выходного сигнала); критерий качества системы и функционал её оптимизации, зависящий от ошибки; значение весовых коэффициентов (например, определяемых аналитически непосредственно из постановки задачи, с помощью некоторых численных методов или процедуры настройки весовых коэффициентов нейронной сети).

Количество и тип формальных нейронов в слоях, а также число слоёв нейронов выбираются исходя из специфики решаемых задач и требуемого качества решения. Нейронная сеть в процессе настройки на решение конкретной задачи рассматривается как многомерная нелинейная система, которая в итерационном режиме целенаправленно ищет оптимум некоторого функционала, количественно определяющего качество решения поставленной задачи. Для нейронных сетей, как многомерных нелинейных объектов управления, формируются алгоритмы настройки множества весовых коэффициентов. Основные этапы исследования нейронной сети и построения алгоритмов настройки (адаптации) их весовых коэффициентов включают: исследование

характеристик входного сигнала для различных режимов работы нейронной сети (входным сигналом нейронной сети является, как правило, входная обрабатываемая информация и указание так называемого «учителя» нейронной сети); выбор критериев оптимизации (при вероятностной модели внешнего мира такими критериями могут быть минимум средней функции риска, максимум апостериорной вероятности, в частности при наличии ограничений на отдельные составляющие средней функции риска); разработку алгоритма поиска экстремумов функционалов оптимизации (например, для реализации алгоритмов поиска локальных и глобального экстремумов); построение алгоритмов адаптации коэффициентов нейронной сети; анализ надёжности и методов диагностики нейронной сети и др.

Необходимо отметить, что введение обратных связей и, как следствие, разработка алгоритмов настройки их коэффициентов в 1960–80 годы имели чисто теоретический смысл, т. к. не было практических задач, адекватных таким структурам. Лишь в конце 1980-х – начале 1990-х годов стали появляться такие задачи и простейшие структуры с настраиваемыми обратными связями для их решения (так называемые рекуррентные нейронные сети). Разработчики в области нейросетевых технологий занимались не только созданием алгоритмов настройки многослойных нейронных сетей и нейросетевыми алгоритмами решения различных задач, но и наиболее эффективными (на текущий момент развития технологии электроники) аппаратными эмуляторами (особые программы, которые предназначены для запуска одной системы в оболочке другой) нейросетевых алгоритмов. В 1960-е годы, до появления микропроцессора, наиболее эффективными эмуляторами нейронных сетей были аналоговые реализации разомкнутых нейронных сетей с разработанными алгоритмами настройки на универсальных ЭВМ (иногда системы на адаптивных элементах с аналоговой памятью). Такой уровень развития электроники делал актуальным введение перекрёстных связей в структуры нейронных сетей. Это приводило к значительному уменьшению числа нейронов в нейронной сети при сохранении качества решения задачи (например, дискриминантной способности при решении задач распознавания образов). Исследования 1960–70-х годов в области оптимизации структур нейронных сетей с перекрёстными связями наверняка найдут развитие при реализации мемристорных нейронных систем [мемристор (memristor, от memory – память, и resistor – электрическое сопротивление), пассивный элемент в микроэлектронике, способный изменять своё сопротивление в зависимости от протекавшего через него заряда], с учётом их специфики в части аналого-цифровой обработки информации и весьма значительного количества настраиваемых коэффициентов. Специфические требования прикладных задач определяли некоторые особенности структур нейронных сетей с помощью алгоритмов настройки: континуум (от лат. continuum – непрерывное, сплошное) числа классов, когда указание «учителя» системы формируется в виде непрерывного значения функции в некотором диапазоне изменения; континуум решений многослойной нейронной сети, формируемый выбором континуальной функции активации нейрона последнего слоя; континуум числа признаков, формируемый переходом в пространстве признаков от представления выходного сигнала в виде NN-мерного вектора вещественных чисел к вещественной функции в некотором диапазоне изменения аргумента; континуум числа признаков, как следствие, требует специфической программной и аппаратной реализации нейронной сети; вариант континуума признаков входного пространства был реализован в задаче распознавания периодических сигналов без преобразования их с помощью аналого-цифрового преобразователя (АЦП) на входе системы, и реализацией аналого-цифровой многослойной нейронной сети;

континуум числа нейронов в слое; реализация многослойных нейронных сетей с континуумом классов и решений проводится выбором соответствующих видов функций активации нейронов последнего слоя.

В таблице показан систематизированный набор вариантов алгоритмов настройки многослойных нейронных сетей в пространстве «Входной сигнал – пространство решений». Представлено множество вариантов характеристик входных и выходных сигналов нейронных сетей, для которых справедливы алгоритмы настройки коэффициентов, разработанных российской научной школой в 1960–70 годах. Сигнал на вход нейронной сети описывается количеством классов (градаций) образов, представляющих указания «учителя». Выходной сигнал нейронной сети представляет собой количественное описание пространства решений. В таблице дана классификация вариантов функционирования нейронных сетей для различных видов входного сигнала (2 класса, КК классов, континуум классов) и различных вариантов количественного описания пространства решений (2 решения, КрКр решений, континуум решений). Цифрами 1, 7, 8 представлены конкретные варианты функционирования нейронных сетей.

Таблица. Набор вариантов алгоритмов настройки

Пространство(число) решений	Входной сигнал			
	2 класса		КК классов	Континуум классов
2	1		7	8
КрКр	Кр=3Кр=3	3 а	К<КрК<Кр	9
			К=КрК=Кр	2
	Кр=constКр=const	3 б	К>КрК>Кр	4
Континуум	5		6	11

Основными преимуществами нейронных сетей как логического базиса алгоритмов решения сложных задач являются: инвариантность (неизменность, независимость) методов синтеза нейронных сетей от размерности пространства признаков; возможность выбора структуры нейронных сетей в значительном диапазоне параметров в зависимости от сложности и специфики решаемой задачи с целью достижения требуемого качества решения; адекватность текущим и перспективным технологиям микроэлектроники; отказоустойчивость в смысле его небольшого, а не катастрофического изменения качества решения задачи в зависимости от числа вышедших из строя элементов.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Изучить архитектуры построения нейронных сетей.
2. Проанализировать применяемые функции активации нейронов.
3. Реализовать разработанный алгоритм на языке python с одной из 3 библиотек.
5. Провести анализ
6. Оформить результаты работы в виде пояснительной записки.

## Свёрточная нейронная сеть

С появлением глубокого обучения “компьютерное зрение” перешло на новую ступень развития. На смену разрозненным значениям пикселей и ограниченному количеству созданных вручную признаков пришли способы сделать машинное распознавание деталей изображения более простым и понятным — это привело к смене парадигмы в этой области. Сегодня в привычных нам вещах из сфер производства и торговли используется множество самых современных приложений для компьютерного зрения. Недавний прорыв в сфере глубокого обучения в компьютерном зрении привнёс колоссальные изменения в нашу повседневную жизнь. Вы могли даже не заметить, как именно в каких-то вещах используется компьютерное зрение. Вот несколько любопытных примеров: автопилот в автомобилях Tesla, разблокировка с помощью Face ID, Animoji и продвинутый функционал камеры в iPhone, эффект боке в режиме портретной съёмки, фильтры в мессенджерах Snapchat и Facebook и т. д.

Основная идея компьютерного зрения начинается с очень простой задачи — определить, что изображено на картинке. Оказывается, эту задачу чрезвычайно сложно решить, хотя мы, люди, легко справляемся с ней.

### Устаревшие решения для компьютерного зрения

С появлением машинного обучения проблемы компьютерного зрения решались относительно успешно. Прежде всего, в этом помогали созданные вручную признаки и традиционные алгоритмы машинного обучения, такие как метод опорных векторов (SVM). Созданные вручную признаки — это параметры изображений, извлекаемые с помощью множества других алгоритмов. Типичный пример — поиск контуров и углов. Простой алгоритм контурного детектора ищет области резкого изменения насыщенности изображения, то есть большую разницу в значениях соседних пикселей. Несколько таких вот простых и пара более сложных признаков выделялись с помощью комбинации алгоритмов и далее передавались алгоритму контролируемого машинного обучения.

Такой подход работает, однако результаты не особо впечатляют. Прежде всего, чтобы создать признаки самостоятельно, придётся приложить немало усилий, скажу больше — это требует серьёзного уровня предметных знаний. К тому же признаки сильно отличаются от случая к случаю. К примеру,

то, что создано для диагностики переломов на рентгеновских снимках, вполне может не подойти для распознавания имени на почтовой посылке.

В цифровом формате изображения представлены в виде 3D-матрицы из значений пикселей (длины, ширины и цветовых каналов RGB). Извлекать информацию из этой 3D-матрицы не так уж просто.

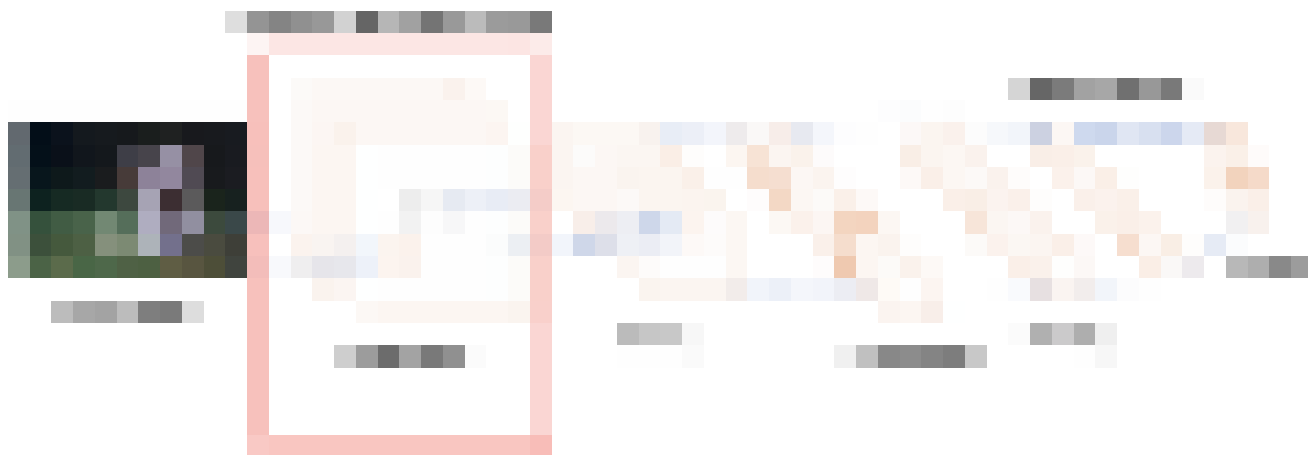
“Свёртка” — операция из области обработки сигнала. В глубоком обучении это перемножение матрицы изображения (собственно матрица) и ядра/фильтра (ещё одна матрица меньшего размера) путём прохождения через длину и ширину. На анимации ниже демонстрируется свёртка фильтра/ядра размером  $3 \times 3$  и изображения размером  $5 \times 5$ . Результат свёртки — изображение меньшего размера ( $3 \times 3$ ).

Глубокое обучение

Это один аспект компонента. Другой его аспект — пулинг. Слой пулинга помогает сократить пространственное представление изображения, чтобы уменьшить количество параметров и объём вычислений в сети. Это простая операция: надо только задать максимальное значение определённому размеру ядра. Ниже дан простой пример пулинга: он проводится с использованием ядра размером  $10 \times 10$  на выходе свёртки (другой матрицы) размером  $20 \times 20$ . В итоге получается матрица размером  $2 \times 2$ .



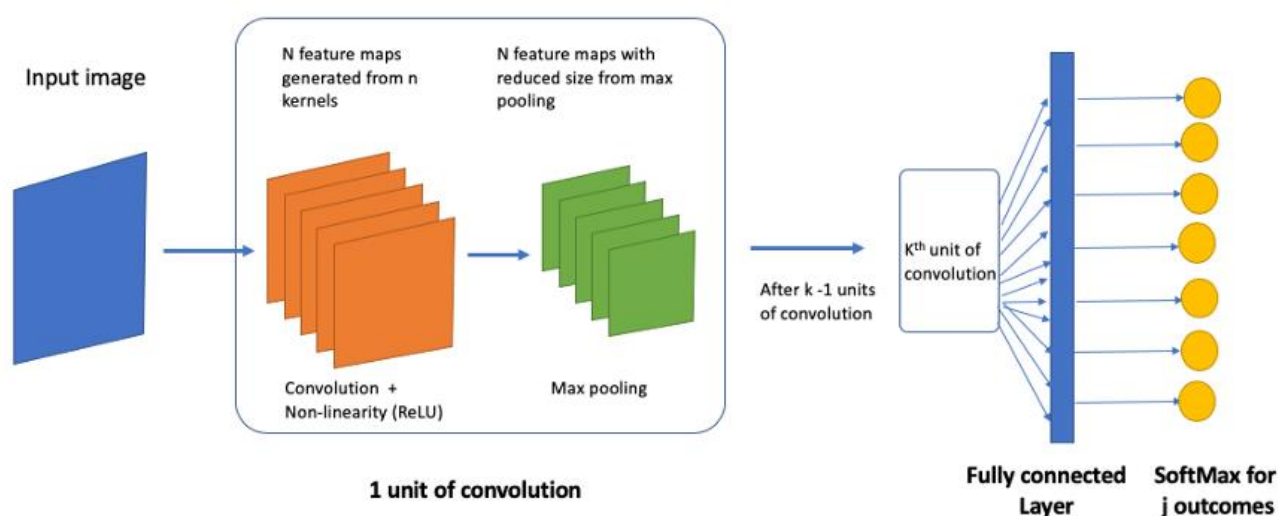
Используя комбинацию слоёв свёртки и слоёв пулинга (с определением максимального значения), мы получаем основной структурный элемент СНС. Свёртка и пулинг уменьшают исходные размеры изображения на входе в зависимости от размеров ядра и пулинга. Применяя свёртку с одним ядром, получаем карту признаков. В СНС обычно применяется несколько ядер на одну свёртку. На рисунке ниже показаны карты признаков, извлечённых из  $n$  ядер при свёртке.



Многократное повторение этого процесса приводит к углублению свёрточных нейронных сетей. Каждый слой извлекает признаки из предыдущего. Иерархическая организация слоёв способствует последовательному изучению признаков: от контуров к более сложным признакам, созданным из простых, и далее к высокоуровневым признакам, которые уже содержат достаточно информации для составления нейросетью точного прогноза.

Последний свёрточный слой связан с полносвязным слоем, который используется для применения подходящей функции-активатора для прогнозирования выхода: для бинарных выходов используется сигмоидная, а для небинарных — многопеременная функция.

Вся описанная архитектура в упрощённом виде показана ниже.



Вопрос, как определить нужный фильтр, отпадает сам собой из-за процесса обучения нейронной сети.



## Функция активации

Функция активации определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения.

Рассмотрим нейрон:

$$Y = \sum (weight * input) + bias$$

Теперь значение  $Y$  может быть любым в диапазоне от -бесконечности до +бесконечности. В действительности нейрон не знает границу, после которой следует активация. Ответим на вопрос, как мы решаем, должен ли нейрон быть активирован (мы рассматриваем паттерн активации, так как можем провести аналогию с биологией. Именно таким образом работает мозг).

Для этой цели существует активационная функция. Она проверяет произведенное нейроном значение  $Y$  на предмет того, должны ли внешние связи рассматривать этот нейрон как активированный, или его можно игнорировать.

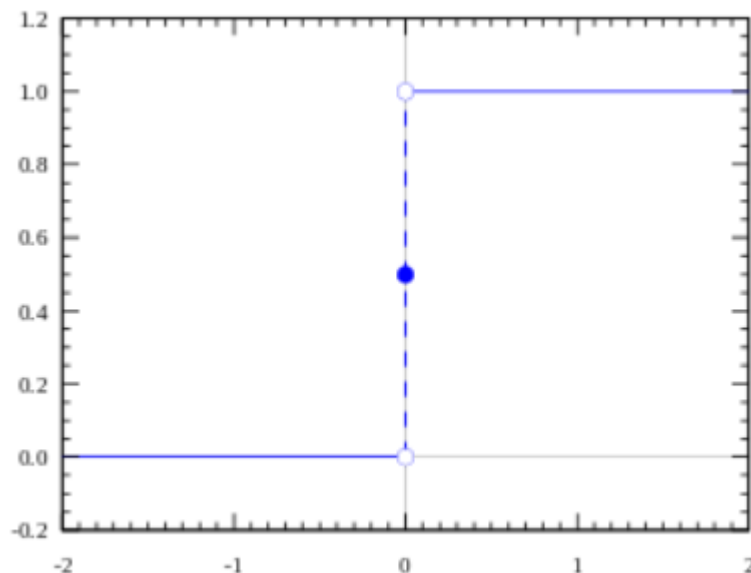
### Ступенчатая функция активации

Если значение  $Y$  больше некоторого порогового значения, считаем нейрон активированным. В противном случае говорим, что нейрон неактивен. Такая схема должна сработать, но сначала давайте её формализуем.

Функция  $A$  = активирована, если  $Y > \text{граница}$ , иначе нет.

Другой способ:  $A = 1$ , если  $Y > \text{граница}$ , иначе  $A = 0$ .

Такая функция называется ступенчатой.



Функция принимает значение 1 (активирована), когда  $Y > 0$  (граница), и значение 0 (не активирована) в противном случае. Это простой способ, однако в нём есть недостатки.



Представим, что мы создаем бинарный классификатор — модель, которая должна говорить “да” или “нет” (активирован или нет). Ступенчатая функция сделает это за вас — она в точности выводит 1 или 0.

Теперь представим случай, когда требуется большее количество нейронов для классификации многих классов: класс1, класс2, класс3 и так далее. И все нейроны из функции активации выведут 1. В таком случае появляются вопросы о том, какой класс должен в итоге получиться для заданного объекта.

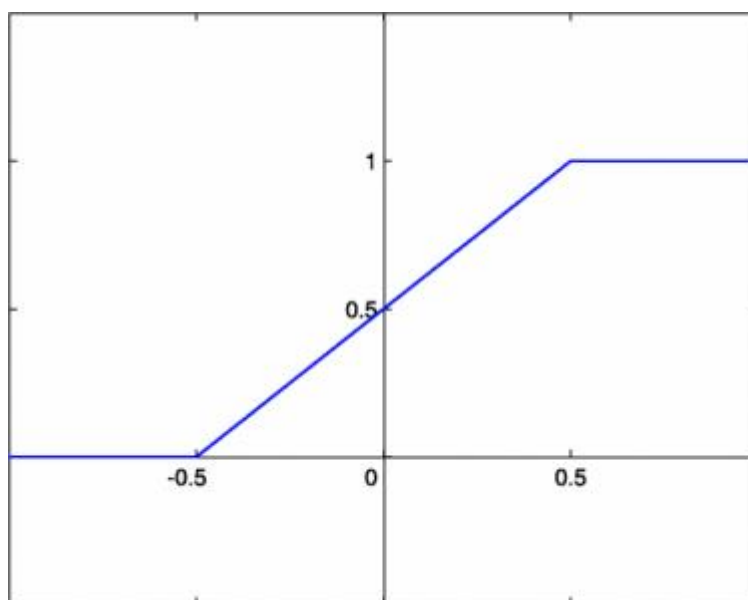
Если активированы несколько нейронов, можно найти нейрон с наибольшим значением активационной функции.

Но в таком случае, как и ранее, если более одного нейрона говорят “активирован на 100%”, проблема по-прежнему остается.

При существовании промежуточных значений на выходе нейрона, процесс обучения проходит более гладко и быстро, а вероятность появления нескольких полностью активированных нейронов во время тренировки снижается по сравнению со ступенчатой функцией активации.

Линейная функция активации

$$A = cx$$



Линейная функция представляет собой прямую линию и пропорциональна входу (то есть взвешенной сумме на этом нейроне).

Такой выбор активационной функции позволяет получать спектр значений, а не только бинарный ответ. Можно соединить несколько нейронов вместе и, если более одного нейрона активировано, решение принимается на основе применения операции max (или softmax).

Производная от  $A=cx$  по  $x$  равна  $c$ . Это означает, что градиент никак не связан с  $x$ . Градиент является постоянным вектором, а спуск производится по постоянному градиенту. Если производится ошибочное предсказание, то изменения, сделанные обратным распространением ошибки, тоже постоянны и не зависят от изменения на входе  $\delta(x)$ .

Существует и другая проблема. Рассмотрим связанные слои. Каждый слой активируется линейной функцией. Значение с этой функции идет в следующий слой в качестве входа, второй слой считает взвешенную сумму на своих входах и, в свою

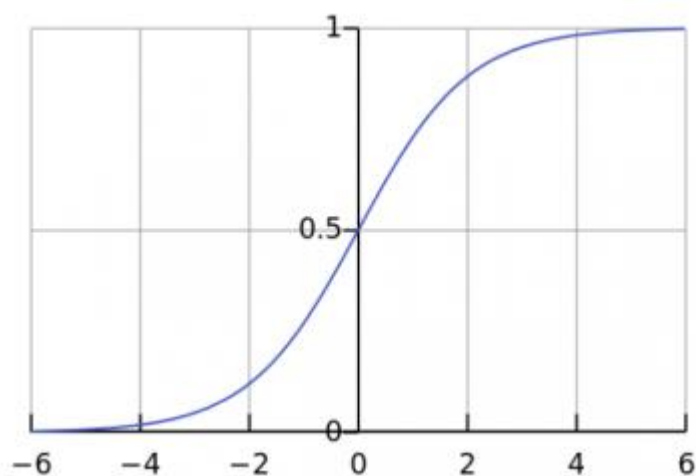
очередь, включает нейроны в зависимости от другой линейной активационной функции.

Не имеет значения, сколько слоев имеется, если все они по своей природе линейные - финальная функция активации в последнем слое будет просто линейной функцией от входов на первом слое.

Это означает, что два слоя (или N слоев) могут быть заменены одним слоем. Таким образом будет потеряна возможность делать наборы из слоев.

Сигмоида

$$A = \frac{1}{1+e^{-x}}$$



Сигмоида подобна ступенчатой функции. Рассмотрим её преимущества.

Во-первых, сигмоида — нелинейна по своей природе, а комбинация таких функций производит тоже нелинейную функцию.

Еще одно достоинство такой функции — она небинарна, что делает активацию аналоговой, в отличие от ступенчатой функции. Для сигмоиды также характерен гладкий градиент.

Сигмоида действительно выглядит подходящей функцией для задач классификации. Она стремится привести значения к одной из сторон кривой (например, к верхнему при  $x=2$  и нижнему при  $x=-2$ ). Такое поведение позволяет находить четкие границы при предсказании.

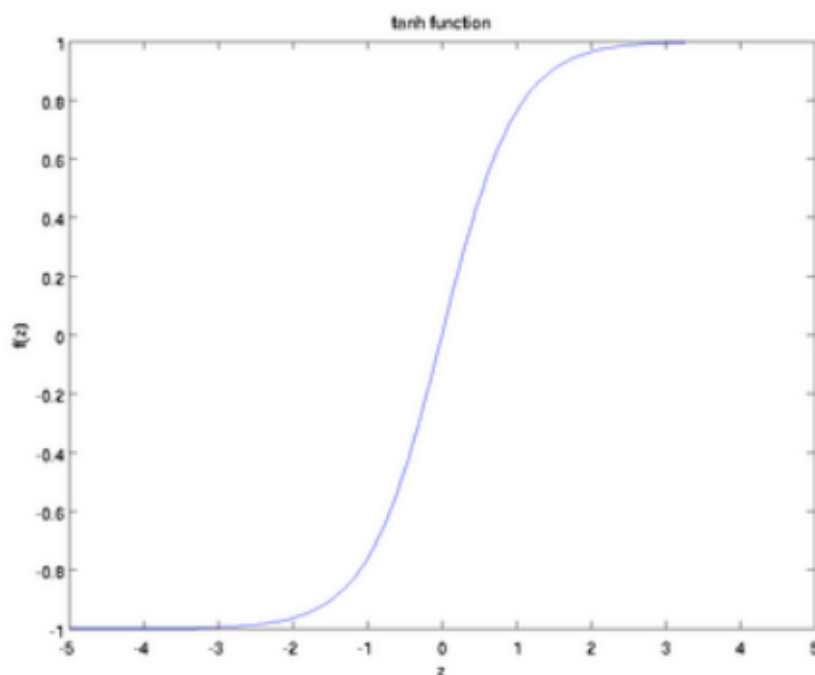
Другое преимущество сигмоиды над линейной функцией заключается в следующем. В первом случае имеем фиксированный диапазон значений функции —  $[0,1]$ , тогда как линейная функция изменяется в пределах  $(-\infty, \infty)$ . Такое свойство сигмоиды очень полезно, так как не приводит к ошибкам в случае больших значений активации.

Сегодня сигмоида является одной из самых частых активационных функций в нейросетях. Но и у неё есть недостатки, на которые стоит обратить внимание.

При приближении к концам сигмоиды значения  $Y$  имеют тенденцию слабо реагировать на изменения в  $X$ . Это означает, что градиент в таких областях принимает маленькие значения. А это, в свою очередь, приводит к проблемам с градиентом исчезновения. Существуют варианты работы над этими проблемами, а сигмоида всё ещё очень популярна для задач классификации.

## Гиперболический тангенс

Еще одна часто используемая активационная функция.



$$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$$

Гиперболический тангенс очень похож на сигмоиду. И действительно, это скорректированная сигмоидная функция.

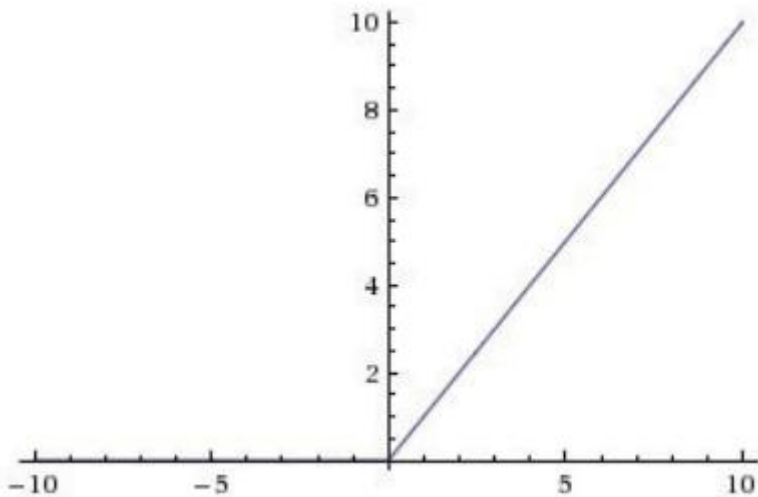
$$\tanh(x) = 2 \operatorname{sigmoid}(2x) - 1$$

Поэтому такая функция имеет те же характеристики, что и у сигмоиды, рассмотренной ранее. Её природа нелинейна, она хорошо подходит для комбинации слоёв, а диапазон значений функции  $(-1, 1)$ . Однако стоит отметить, что градиент тангенциальной функции больше, чем у сигмоиды, соответственно, производная круче. Также как и сигмоиде, гиперболическому тангенсу свойственная проблема исчезновения градиента.

## ReLU

$$A(x) = \max(0, x)$$

Пользуясь определением, становится понятно, что ReLU возвращает значение  $x$ , если  $x$  положительно, и  $0$  в противном случае. Схема работы приведена ниже.



На первый взгляд кажется, что ReLu имеет все те же проблемы, что и линейная функция, так как ReLu линейна в первом квадранте. Но на самом деле, ReLu нелинейна по своей природе, а комбинация ReLu также нелинейна. На самом деле, такая функция является хорошим аппроксиматором, так как любая функция может быть аппроксимирована комбинацией ReLu. Область допустимых значений ReLu —  $[0, \infty)$ , то есть активация может “взорваться”.

Следующий пункт — разреженность активации. Представим большую нейронную сеть с множеством нейронов. Использование сигмойды или гиперболического тангенса повлечёт за собой активацию всех нейронов аналоговым способом. Это означает, что почти все активации должны быть обработаны для описания выхода сети. В идеале лишь некоторые нейроны должны быть активированы, это сделало бы активации разреженными и эффективными.

ReLu позволяет это сделать. Представим сеть со случайно инициализированными весами (или нормализованными), в которой примерно 50% активаций равны 0 из-за характеристик ReLu. В такой сети включается меньшее количество нейронов (разреженная активация), а сама сеть становится легче. Но и у ReLu будут свои недостатки.

Из-за того, что часть ReLu представляет из себя горизонтальную линию (для отрицательных значений  $x$ ), градиент на этой части равен 0. Из-за равенства нулю градиента, веса не будут корректироваться во время спуска. Это означает, что пребывающие в таком состоянии нейроны не будут реагировать на изменения в ошибке/входных данных. Такое явление называется проблемой умирающего ReLu (Dying ReLu problem). Из-за этой проблемы некоторые нейроны просто выключатся и не будут отвечать, делая значительную часть нейросети пассивной. Однако существуют вариации ReLu, которые помогают эту проблему избежать. Например, имеет смысл заменить горизонтальную часть функции на линейную. Если выражение для линейной функции задается выражением  $y = 0.01x$  для области  $x < 0$ , линия слегка отклоняется от горизонтального положения. Существует и другие способы избежать нулевого градиента. Основная идея здесь — сделать градиент неравным нулю и постепенно восстанавливать его во время тренировки.

ReLu менее требовательно к вычислительным ресурсам, чем гиперболический тангенс или сигмоида, так как производит более простые математические операции. Поэтому имеет смысл использовать ReLu при создании глубоких нейронных сетей.

На основе проведенного анализа различных подходов к построению нейронных сетей и методов классификации данных были сделаны следующие выводы:

- Для классификации изображений выбрана архитектура свёрточной нейронной сети.
- Выбрана функция активации ReLu, которая будет эффективна и позволит облегчить обработку больших объёмов данных.
- Выбрана аннотированная база данных изображений эмоций с 7 основными типами.

## Проектирование нейронной сети

У ResNet есть простая идея: передать выходные данные двух последовательных сверточных слоев, а также обойти входные данные для следующих слоев. Обход после 2 слоев-это ключевая интуиция, так как обход одного слоя не дал особых улучшений. По 2 слоям можно рассматривать как небольшой классификатор или сеть внутри сети. Это также самый первый случай, когда была обучена сеть из более чем ста, даже 1000 слоев.

Основная задача данной нейронной сети - распознавание эмоций на лице. Распознавание лиц в принципе сложная задача из-за невозможность конкретно систематизировать данные. В типичных задачах распознавания лиц в общем можно выделить 6 проблем.

Одной из них является освещение. Мимика лица может быть нечётко выражена из-за затенения, разных источников света. Если набор данных не был собран в идентичных условиях - это добавит шанс ошибки.

Также появляется проблема положения лица на фотографии. Потребуется инвариантность на поворот головы, нахождения лица в определённом углу фотографии. Идеальным положением должен быть прямой взгляд в камеру, так как не все эмоции могут быть распознаны даже при угле поворота лица в 30 градусов. При сильном различии по этому пункту также повысится ошибка.

Мимические морщины могут быть спутаны с фоновыми помехами. К примеру, складка на фоновой ткани может быть принята за часть лица, и алгоритм начнёт сбивать, показывая перевес в финальных параметрах на совершенно ненужный класс.

Лица на изображениях должны быть масштабированы. Дело обстоит не в базовой подготовке изображения в виде масштабирования самого изображения, а именно в размерах лица. Алгоритм может как упускать важные детали мимики на маленьких образцах, так и придавать слишком большие значения случайным теням, выглядящим слишком объёмно.

Имеет место быть проблема заслонения части лица. Примеры: чёлка может заслонять брови, что усложнит распознавание эмоций из спектра злости или страха; борода или усы могут усложнить распознавание улыбки, затеняя её.

Также есть специфичные проблемы именно для распознавания мимики. Возраст людей может варьироваться, соответственно и размеры лица, старческие морщины подобные особенности будут накладывать отпечаток на распознавание.

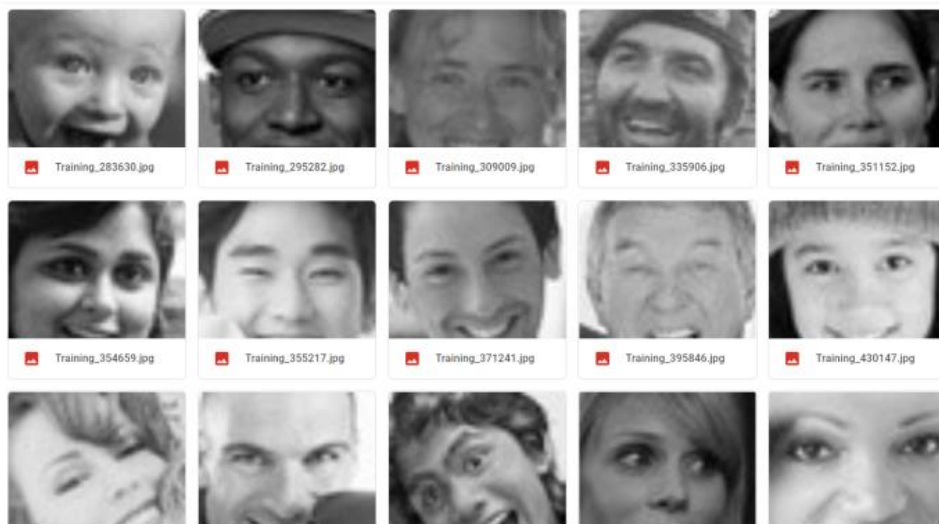
Люди бывают совершенно разные: сильно эмоциональные, чрезмерно реагирующие, всегда спокойные. Из-за этого в первую очередь появляется человеческая ошибка - сами люди не всегда могут прочесть эмоции по выражению лица, как и само понятие «эмоция» - вещь достаточно субъективная. Сами выражения могут быть смешанными, отражающими сразу несколько классов: счастье и удивление, отвращение и гнев. Эта проблема и появится в дальнейшем.

Часть данных из датасета представлена выражениями эмоций актёров. Наигранность поведения также вызовет ошибку, так как выражения лиц людей в повседневной жизни будут отличаться.

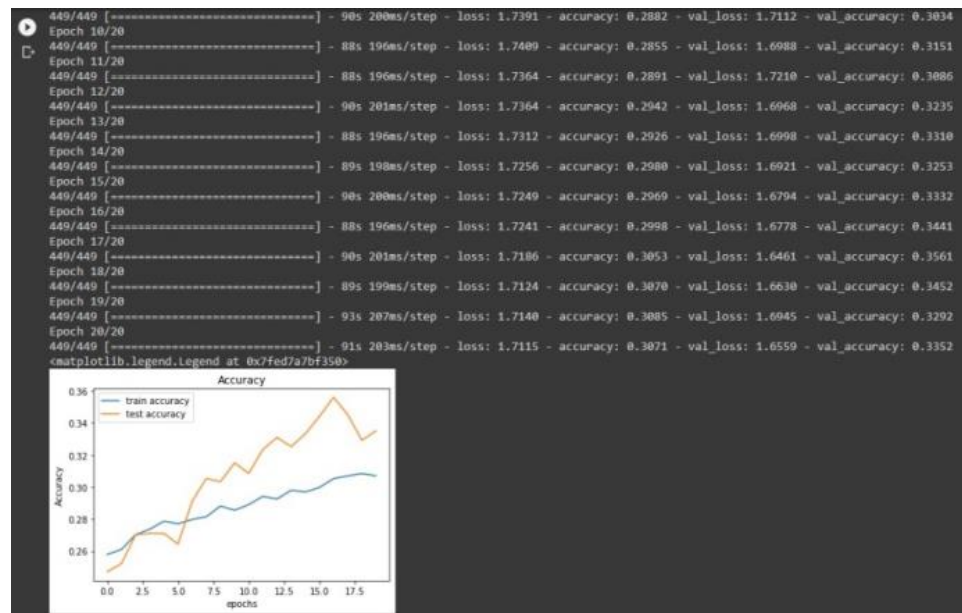
Таким образом, если эффективность современных алгоритмов распознавания лиц давно перевалила за 95 процентов, эмоции на лице - область проблемная. Один из выходов: распознавание мимики на видео данных. Это даст несколько кадров сразу

и, хотя это может частично повысить ошибку, если лицо меняло положение, позволит «рассмотреть» лицо тщательнее.

Дтасет состоит из чёрнобелых 28709 фотографий размером 48x48 пикселей, разделённых на 7 классов: злость, отвращение, страх, радость, нейтраль, грусть, удивление.

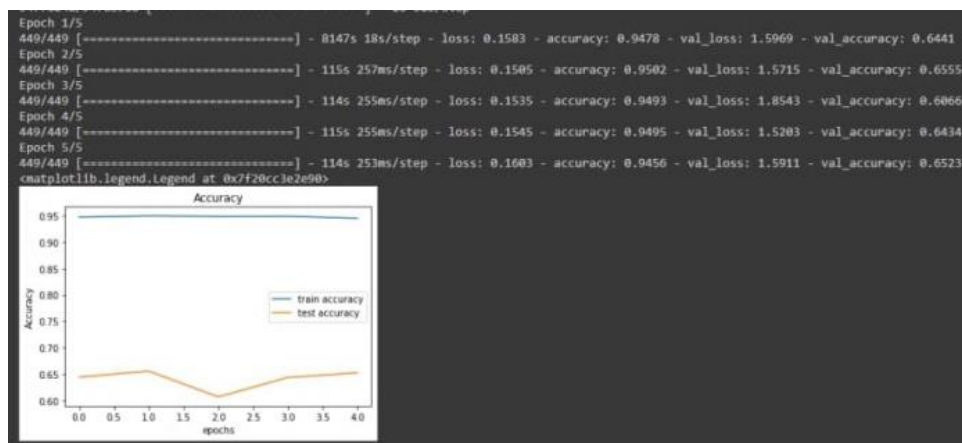


Первая попытка реализации трёхслойной свёрточной сети дала представленные результаты. По прохождению 20 эпох значение точности вышло на плато, дальнейшее обучение не повысило бы эффективности архитектуры.



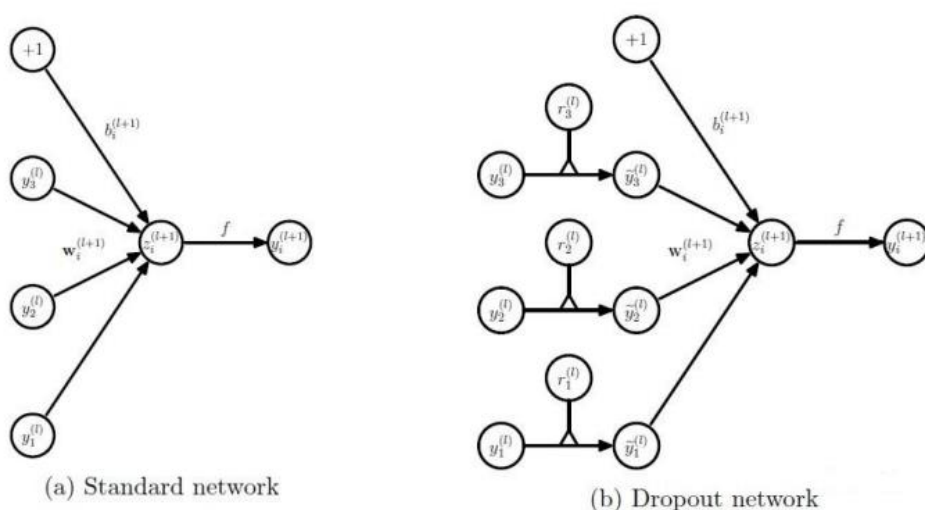
## ResNet

У ResNet есть простая идея: передать выходные данные двух последовательных свёрточных слоев, а также обойти входные данные для следующих слоев. Обход после 2 слоев-это ключевая интуиция, так как обход одного слоя не дал особых улучшений. По 2 слоям можно рассматривать как небольшой классификатор или сеть внутри сети. Это также самый первый случай, когда была обучена сеть из более чем ста, даже 1000 слоев.



Обучение на вид прошло отлично, с макисмальным значением в районе 95% точности. Однако при взгляде на данные тестирования видно, что значение точности упало до 65%. Характерная просадка в районе 2 эпохи говорит о переобучении нейросети на обучающем наборе. Это можно исправить посредством метода Dropout.

Это метод регуляризации искусственных нейронных сетей, предназначен для уменьшения переобучения сети за счет предотвращения сложных коадаптаций отдельных нейронов на тренировочных данных во время обучения. Термин «dropout» (выбивание, выбрасывание) характеризует исключение определённого процента (например 30%) случайных нейронов (находящихся как в скрытых, так и видимых слоях) на разных итерациях (эпохах) во время обучения нейронной сети. Вероятность выключения каждого нейрона одинакова.



На тестовом наборе эффективность выше 70% не была замечена. С другой стороны, возможна ли вообще высокая эффективность при сильном разбросе данных?





На первой картинке, взятой из датасета папки “нейтраль” выражение лица девушки можно интерпретировать, как счастье. Подобные обучающие материалы, как и 2, 4, 5, в большом количестве будут влиять на нейроны “не в ту сторону”, смешивая определение эмоциональных окрасок.

Картинки 3 и 6 (возможно, и 2) демонстрируют проблему плохого подбора данных с точки зрения систематизированности. Проблема поворота головы уже упоминалась выше, как и заслонённость части лица.

Эти проблемы можно исправить только сужением набора данных, подбора идеально идентичных изображений.

Также возможен вариант приписывания картинкам сразу двух эмоциональных окрасок, но это также может вызывать проблемы конкретного определения эмоции на изображении.

## Заключение

Эмоции субъективны и их сложно фиксировать. Необходимо чётко выделить, какие эмоции надо выделять. Нужно работать с разделёнными данными, так как даже такой объём пришлось прогонять с подпиской гугл колаба. Нехватка данных — это важный фактор успеха в SER, однако создать хороший датасет с эмоциями — это комплексная задача. Слишком большой разброс данных ведёт к малоэффективному обучению. Модель требует упрощения, так как начинает переобучаться.

## Список литературы

1. Ekman, P. (1999). Basic Emotions In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion Pp. 45-60.
2. Статья “Введение в задачу распознавания эмоций”  
<https://habr.com/ru/company/speechpro/blog/418151/>
3. Ekman, P. (1992b). Are There Basic Emotions? Psychological Review, 99(3), 550-553.
4. Тимофеева Ольга Павловна, Неимуцев Сергей Андреевич, Неимуцева Любовь Ивановна, Тихонов Иван Александрович Распознавание эмоций по изображению лица на основе глубоких нейронных сетей // Труды НГТУ им. Р. Е. Алексеева. 2020. №1 (128).
5. Speech Emotion Recognition with Convolutional Neural Network by Reza Chu Перевод: <https://habr.com/ru/company/Voximplant/blog/461435/>
6. Ivanov I.A. Multi-objective based feature selection and neural networks ensemble method for solving emotion recognition problem