



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

---

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

# РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

## К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

### НА ТЕМУ:

*«Методы реализации нейронных сетей»*

Студент ИУ7-51Б  
(Группа)

\_\_\_\_\_  
(Подпись, дата) Д.И. Костев  
(И.О.Фамилия)

Руководитель

\_\_\_\_\_  
(Подпись, дата) В.М. Градов  
(И.О.Фамилия)

2021 г.

**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)**

---

УТВЕРЖДАЮ

Заведующий кафедрой ИУ-7

И. В. Рудаков

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ г.

## З А Д А Н И Е

### на выполнение научно-исследовательской работы

по теме \_\_\_\_\_ «Методы реализации нейронных сетей»

Студент группы \_\_\_\_\_ ИУ7-51Б

Костев Дмитрий Игоревич

(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

\_\_\_\_\_ учебная

Источник тематики (кафедра, предприятие, НИР) \_\_\_\_\_ НИР

График выполнения НИР: 25% к 4 нед., 50% к 7 нед., 75% к 11 нед., 100% к 14 нед.

**Техническое задание: классифицировать методы реализации алгоритмов нейронных сетей, предложить наиболее подходящий для анализа изображений, а именно определение каких либо заданных статических объектов.**

#### ***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на 15-25 листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

Презентация на 8-10 слайдах.

Дата выдачи задания «03» сентября 2021 г.

**Руководитель НИР**

\_\_\_\_\_  
(Подпись, дата)

В.М. Градов

(И.О.Фамилия)

**Студент**

\_\_\_\_\_  
(Подпись, дата)

Д.И. Костев

(И.О.Фамилия)

**Примечание:** Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	4
1 Анализ предметной области .....	6
1.1 Основные определения .....	6
1.2 Выводы.....	7
2 Классификация существующих решений.....	8
2.1 Однослойные нейронные сети.....	8
2.2 Многослойные нейронные сети .....	9
2.3 Нейросети прямого распределения.....	10
2.4 Сети с обратными связями.....	13
2.5 Классификация сетей.....	15
2.6 Оценка методов.....	16
2.6 Выводы.....	17
ЗАКЛЮЧЕНИЕ .....	18
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	19

## ВВЕДЕНИЕ

В наше время мы всё чаще слышим понятие нейронная сеть, и это не просто так. Технология является одной из самых перспективных в мире и активно внедряется во все сферы деятельности.

Задача таких искусственных нейронных сетей — выполнять такие когнитивные функции, как решение проблем и машинное обучение. Нейросетевые алгоритмы успешно применяются для решения сложных практических задач, традиционно считающихся интеллектуальными: распознавание лиц (и другие задачи распознавания изображений и объектов на изображении), управление беспилотными летательными аппаратами, медицинская диагностика заболеваний.

Конечно, технологии и методы искусственного интеллекта делают основной упор на ситуации, обладающие одной или несколькими следующими особенностями:

- алгоритм решения неизвестен или не может быть использован из-за ограниченности ресурсов компьютера;
- задача не может быть определена в числовой форме;
- цели задачи не могут быть выражены в терминах точно определенной целевой функции-критерия.

Однако, поскольку "знания — это формализованная информация, которую используют в процессе логического вывода", то можно сказать, что нейросеть берёт факты (фактические знания о мире, представленные в виде обучающей выборки) и в процессе обучения формирует правила — знания, описывающие найденный нейросетью способ решения. Эти правила принятия решения можно затем извлечь из нейронной сети и записать в одном из традиционных для классических экспертных систем формализмов представления знаний (например, в виде набора продукционных правил логического вывода).

Целью данной работы является классификация существующих видов нейронных сетей и выбор наилучшей для анализа изображений.

Задачи, которые необходимо решить для достижения поставленной цели:

- 1) изучить существующие виды реализации нейронных сетей;
- 2) выбрать реализацию, предположительно наиболее эффективно решающую задачу анализа изображений, а именно определение каких-либо заданных статических объектов.

# 1 Анализ предметной области

## 1.1 Основные определения

Нейронная сеть — это последовательность нейронов, соединенных между собой синапсами. Структура нейронной сети пришла в мир программирования прямоком из биологии. Благодаря такой структуре машина обретает способность анализировать и даже запоминать различную информацию. Нейронные сети также способны не только анализировать входящую информацию, но и воспроизводить ее из своей памяти. Другими словами, нейросеть это машинная интерпретация мозга человека, в котором находятся миллионы нейронов, передающих информацию в виде электрических импульсов.

Широкий круг задач, решаемый *нейронными сетями* (НС), не позволяет в настоящее время создавать универсальные, мощные сети, вынуждая разрабатывать специализированные НС, функционирующие по различным алгоритмам. Модели НС могут быть программного и аппаратного исполнения. В дальнейшем речь пойдет в основном о первом типе.

Несмотря на существенные различия, отдельные типы НС обладают несколькими общими чертами. Основу каждой НС составляют относительно простые, в большинстве случаев — однотипные, элементы (ячейки), имитирующие работу нейронов мозга. Далее под нейроном будет подразумеваться искусственный нейрон, то есть ячейка НС.

Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены или заторможены. Он обладает группой синапсов — односторонних входных связей, соединенных с выходами других нейронов, а также имеет аксон — выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов. Каждый синапс

характеризуется величиной синаптической связи или ее весом  $w_i$ , который по физическому смыслу эквивалентен электрической проводимости.

Нейронным сетям присущ принцип параллельной обработки сигналов, который достигается путем объединения большого числа нейронов в так называемые слои и соединения определенным образом нейронов различных слоев, а также, в некоторых конфигурациях, и нейронов одного слоя между собой, причем обработка взаимодействия всех нейронов ведется послойно.

### **1.3 Выводы**

В данной части были введены основные определения для работы с алгоритмами нейронных сетей.

## 2 Классификация существующих решений

Как правило, в большинстве нейронных сетей есть так называемый входной слой, который выполняет только одну задачу — распределение входных сигналов остальным нейронам. Нейроны этого слоя не производят никаких вычислений. В остальном нейронные сети делятся на основные категории, представленные ниже [1].

### 2.1 Однослойные нейронные сети

Однослойная нейронная сеть (англ. Single-layer neural network) — сеть, в которой сигналы от входного слоя сразу подаются на выходной слой, который и преобразует сигнал и сразу же выдает ответ. Как видно из схемы, представленной на рис. 1 однослойной нейронной сети, представленной справа, сигналы  $x_1, x_2, \dots, x_n$  поступают на входной слой (который не считается за слой нейронной сети), а затем сигналы распределяются на выходной слой обычных нейронов. На каждом ребре от нейрона входного слоя к нейрону выходного слоя написано число — вес соответствующей связи.

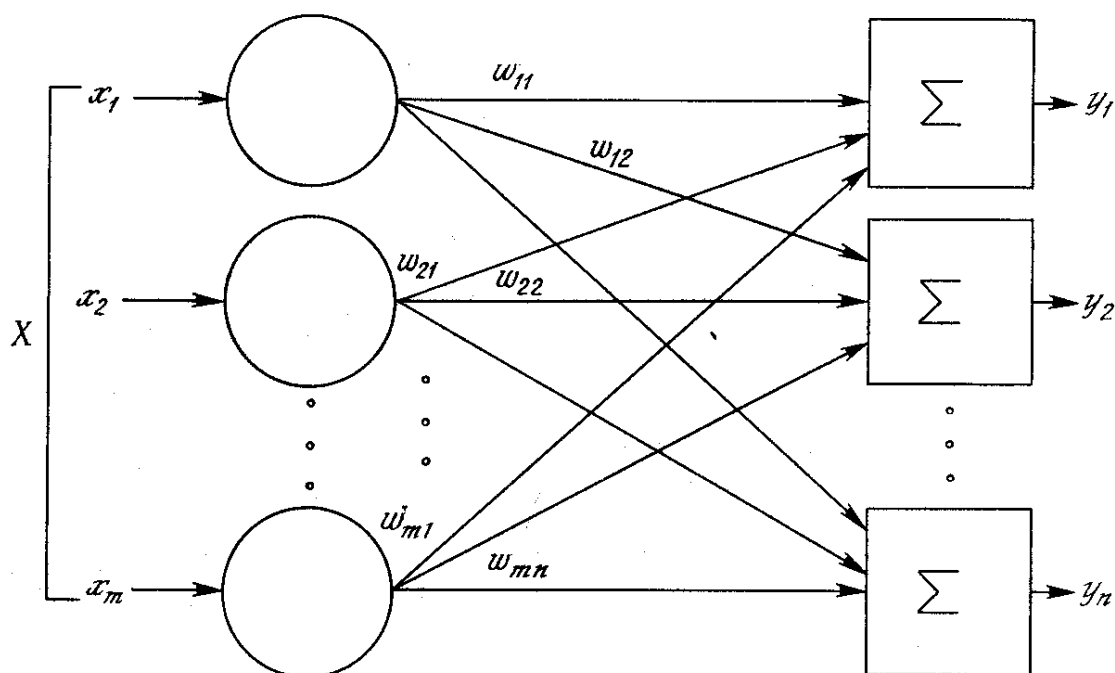


Рис. 1. Схема однослойной нейронной сети



## 2.2 Многослойные нейронные сети

Многослойная нейронная сеть (англ. Multilayer neural network) — нейронная сеть, состоящая из входного, выходного и расположенного(ых) между ними одного (нескольких) скрытых слоев нейронов.

Помимо входного и выходного слоев эти нейронные сети содержат промежуточные, скрытые слои. Такие сети обладают гораздо большими возможностями, чем однослойные нейронные сети, однако методы обучения нейронов скрытого слоя были разработаны относительно недавно.

Работу скрытых слоев нейронов можно сравнить с работой большого завода. Продукт (выходной сигнал) на заводе собирается по стадиям на станках. После каждого станка получается какой-то промежуточный результат. Скрытые слои тоже преобразуют входные сигналы в некоторые промежуточные результаты.

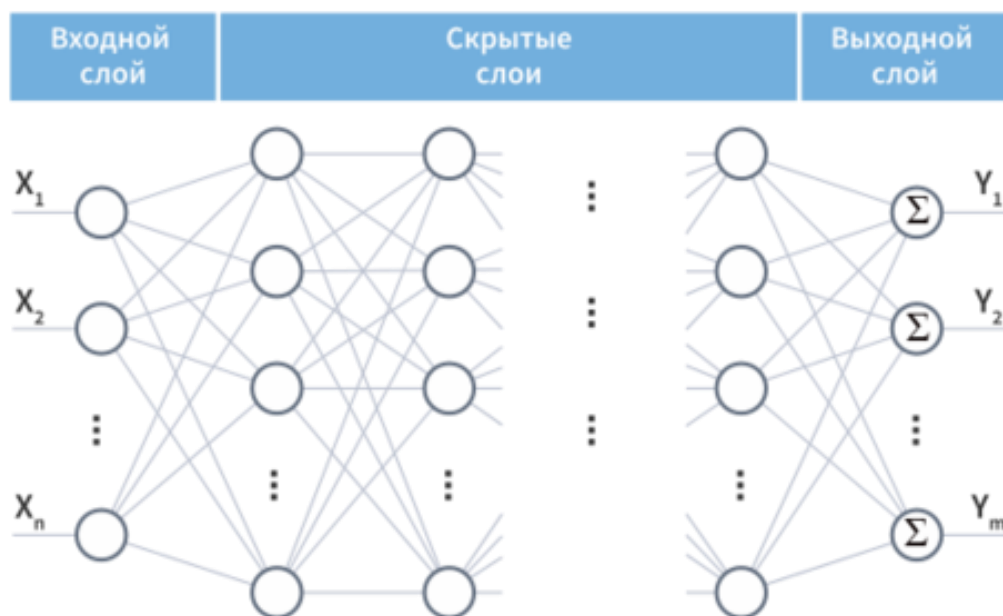


Рис. 2. Схема многослойной нейронной сети

## 2.3 Нейросети прямого распространения

Нейросети прямого распространения (англ. Fully Connected Feed-Forward Neural Networks, FNN) — это искусственные нейронные сети, в которых сигнал распространяется строго от входного слоя к выходному [2]. В обратном направлении сигнал не распространяется. Она является наиболее базовой из архитектур нейронных сетей. Она состоит из последовательно соединённых слоёв, среди которых выделяются входной слой (внизу на рис. 3), выходной слой (вверху) и скрытые слои.

Все сети, описанные выше, являлись сетями прямого распространения, как следует из определения. Такие сети широко используются и вполне успешно решают определенный класс задач: прогнозирование, кластеризация и распознавание.

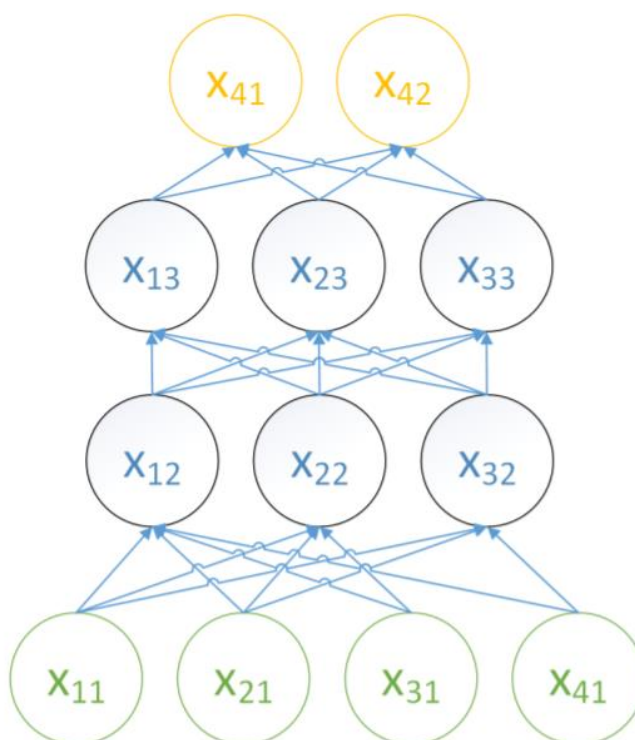


Рис. 3. Нейронная сеть прямого распространения

Каждый слой содержит в себе набор нейронов. Выход  $i$ -го нейрона  $j$ -го слоя ( $j > 1$ , если входной слой считать первым) определяется следующим образом:

$$x_{ij} = f_j \left( \sum_k w_{ijk} x_{k,j-1} \right), \quad (1)$$

где  $w_{ijk}$  – коэффициенты линейной комбинации, а  $f_j(x)$  – функция активации, например, сигмоид:

$$\sigma(x) = \frac{1}{1+e^{-x}}. \quad (2)$$

В дальнейшем будем использовать матричное представление, в соответствии с которым выход  $j$ -го слоя будет равен

$$x_j = f_j(Wx_{j-1}). \quad (3)$$

В качестве функции активации предлагается выбирать любую нелинейную функцию, но на практике из-за некоторых функций активаций (например, сигмоид), может быстро затухать градиент в глубоких сетях. Борьба с этим можно, например, используя функции вроде ReLU или её модификаций. Для обучения сети, то есть для поиска параметров  $W_j$ , будем сравнивать отклонение значений выходного слоя  $x_n$  на некотором входе  $x_1$  от ожидаемых значений  $z$ . Для этого введём функцию потерь  $E(x_n, z)$  и минимизируем её по параметрам  $W_j$  для  $j \in [2, n]$ , а градиенты вычислим при помощи алгоритма обратного распространения, заключающийся в применении правила дифференцирования сложной функции.

Для поиска минимума обычно используется либо стохастический градиентный спуск, либо его модификации вроде RMSProp [3] или Adam [4], которые позволяют регулировать темп обучения (learning rate) по координатам.

У НС прямого распространения есть проблема — много параметров. Например, если взять нейросеть из 3 скрытых слоев, которой нужно обрабатывать картинки  $100 \times 100$  ps, это значит, что на входе будет 10 000 ps, и они заводятся на 3 слоя. В общем, если честно посчитать все параметры, у такой сети их будет порядка миллиона. Это на самом деле много. Чтобы обучить нейросеть с миллионом параметров, нужно очень много обучающих

примеров, которые не всегда есть. На самом деле сейчас примеры есть, а раньше их не было — поэтому, в частности, сети не могли обучать, как следует. Кроме того, сеть, у которой много параметров, имеет дополнительную склонность переобучаться. Она может заточиться на то, чего в реальности не существует: какой-то шум Data Set. Даже если, в конце концов, сеть запомнит примеры, но на тех, которых она не видела, потом не сможет нормально использоваться.

## 2.4 Сети с обратными связями

При обучении сети прямого распространения неявно предполагается, что элементы обучающей выборки являются взаимно независимыми. Следовательно, такая модель не подходит для работы с данными, у которых есть, например, временная зависимость. В связи с этим были предложены сети Элмана [5], которые впоследствии стали известны как рекуррентные нейронные сети (RNN).

Сети с обратными связями (англ. Recurrent neural network) — искусственные нейронные сети, в которых выход нейрона может вновь подаваться на его вход. В более общем случае это означает возможность распространения сигнала от выходов к входам.

В сетях прямого распространения выход сети определяется входным сигналом и весовыми коэффициентами при искусственных нейронах. В сетях с обратными связями выходы нейронов могут возвращаться на входы. Это означает, что выход какого-нибудь нейрона определяется не только его весами и входным сигналом, но еще и предыдущими выходами (так как они снова вернулись на входы).

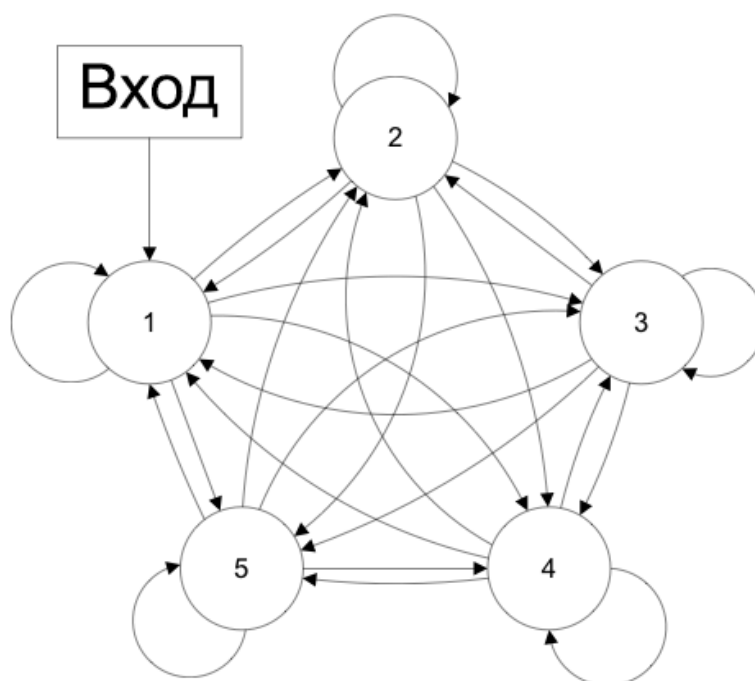


Рис. 4. Схема сети с обратными связями

Будем рассматривать сеть с одним скрытым слоем. Отличие такой сети от сети прямого распространения заключается в добавлении к входам скрытого слоя выходов из скрытого слоя в предыдущий момент времени. Обозначив за  $x_t$ ,  $h_t$  и  $y_t$  выходы входного, скрытого и выходного слоёв соответственно в момент времени  $t$ , определим рекуррентную нейронную сеть:

$$h_t = f_j(Ux_t + Wh_{t-1}) , \quad (4)$$

$$y_t = f_j(Vh_t) . \quad (5)$$

На практике при обучении стандартных рекуррентных нейронных сетей могут возникать проблемы, связанные с экспоненциальным затуханием или ростом градиента. Они были подробно изучены в работе «Модели и концепции революционной кибернетики» [6].

Из-за первой проблемы такие сети не способны обучиться долгосрочным зависимостям. Для её решения предлагается использовать LSTM сети, которые описаны ниже.

Для решения второй проблемы предлагалось нормировать градиент, если его норма превосходит некоторую заранее заданную величину.

uRNN — ещё один подход, избавляющий от указанных проблем, идея которого заключается в том, что при любых параметрах матрица перехода будет унитарной. uRNN работает почти так же хорошо, как LSTM, но содержит меньше параметров [7].

## 2.5 Классификация сетей

На рис. 5 приведена классификация рассмотренных в данной работе методов реализации НС.

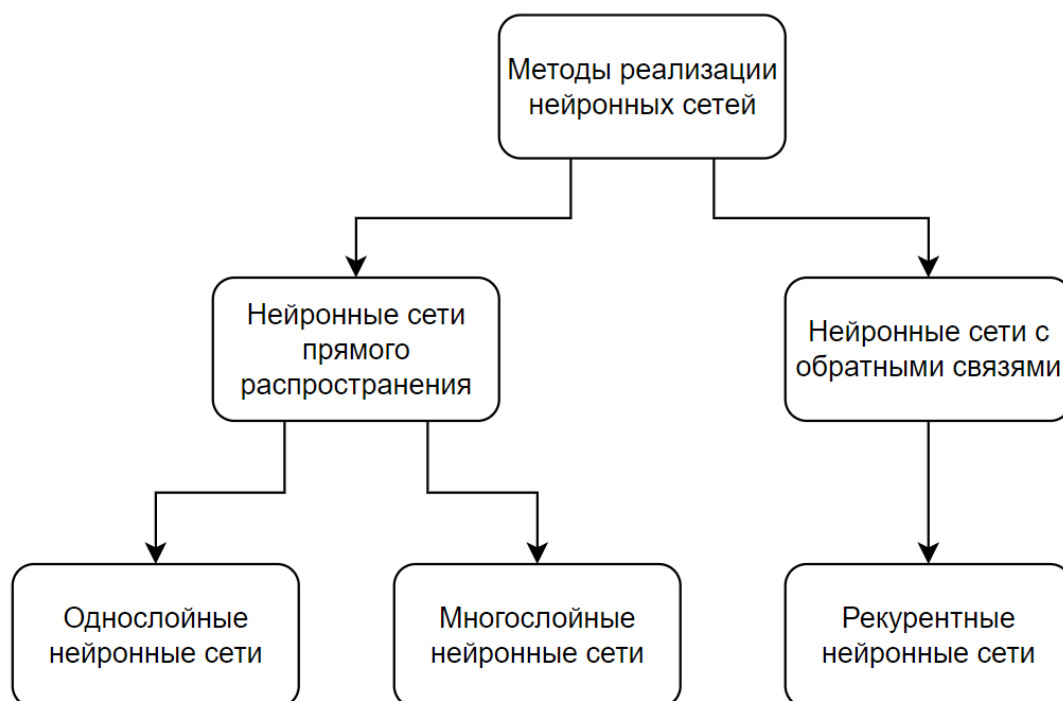


Рис. 5. Классификация методов реализации НС.

## 2.6 Оценка методов

На основании описанных выше видов нейронных сетей были выбраны такие критерии как сложность реализации, влияющий на скорость разработки сети, количество данных распределённых на нейронах и зависимость элементов выборки данных.

Результаты приведены в таблице 1.

Таблица 1. Сравнение методов реализации нейросетей

Название	Сложность реализации	Количество данных на нейронах	Зависимость элементов выборки данных
Однослойная	Простая	Низкое	Нет
Многослойная	Средняя	Высокое	Нет
Рекуррентная	Высокая	Высокое	Да

Однослойная сеть обладает излишней простотой, из-за чего ограничена в возможностях, а это значит, что она будет работать менее продуктивно на сложных объектах, то есть анализировать изображение куда дольше.

Рекуррентная сеть не подойдёт для данной задачи, так как элементы выборки должны быть независимыми. Пиксели объекта же могут быть зависимыми(иметь схожий цвет).

Поэтому наиболее подходящей выбрана многослойная нейронная сеть.



## **2.6 Выводы**

В данной части были рассмотрены виды нейронных сетей, был выбран метод, который лучше всего решает задачу определения статического объекта на изображении из рассмотренных.

## **ЗАКЛЮЧЕНИЕ**

В ходе данной работы были проанализированы некоторые методы реализации нейронных сетей; для анализа статических объектов на изображениях было предложено использовать многослойная нейронная сеть.

Задачи, решенные для достижения поставленной цели:

- 1) изучены существующие виды реализации нейронных сетей,
- 2) выбрана реализация, предположительно наиболее эффективно решающая задачу.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Классификация нейронных сетей / Круглов В.В., Борисов В.В. – Искусственные нейронные сети. Теория и практика, 2015, 35-40 с.
- [2] Нейронная сеть прямого распространения / Бишоп С.М. – Распознавание образов. Машинное обучение, 2006, № 1 – 25-40 с.
- [3] Функции и методы нейронных сетей / Тиелеман Т., Хинтон Д. – Нейронные сети. 2012. № 1 – 102-105 с.
- [4] Модификация Adam / Кингма Д., Ба Д. – Методы оптимизации, 2014, №1 – 110-120 с.
- [5] Нахождение структуры во времени / Элман Д.Л. – Когнитивная наука, 1990, № 14, часть 2 – 179-211 с.
- [6] Модели и концепции революционной кибернетики / Редько В.Г – Эволюция, нейронные сети, интеллект, 2019, №1 – 224 с.
- [7] Унитарная эволюция рекуррентных нейронных сетей / Арьевский М., Шах А., Бенджио Ю. – На Международной конференции по машинному обучению, 2016 – 1120-1128 с.