МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

“ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ”

Факультет компьютерных наук

Кафедра цифровых технологий

Изучение вопроса качества аппроксимации нейронных сетей в зависимости от вида архитектуры

Курсовая работа

02.03.01 Математика и компьютерные науки

Распределенные системы и искусственный интеллект

Зав. кафедрой \_\_\_\_\_\_\_\_\_\_\_ С.Д. Кургалин, д.ф.-м.н., профессор \_\_.\_\_.2024

Обучающийся\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, 3 курс, д/о

Руководитель \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Воронеж 2024

**СОДЕРЖАНИЕ**

**Введение…………………………………………………………………**

**Мотивация………………………………………………………………**

**Обзор литературы………………………………………………………**

**Эксперимент…………………………………………………………….**

**Заключение……………………………………………………………...**

**Мотивация**

В современном мире нейронные сети играют все более важную роль в различных сферах нашей жизни. Эти мощные инструменты машинного обучения применяются для решения широкого спектра задач, включая распознавание образов, обработку естественного языка и прогнозирование временных рядов.

Качество аппроксимации нейронных сетей, то есть их способность приближаться к неизвестным функциям, имеет решающее значение для их эффективности в этих задачах. Различные архитектуры нейронных сетей, такие как сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) и трансформаторы, обладают разной способностью аппроксимировать различные типы функций.

Понимание того, как выбор архитектуры влияет на качество аппроксимации, имеет решающее значение для разработки нейронных сетей, способных эффективно решать конкретные задачи.

В этой курсовой работе мы изучим вопрос качества аппроксимации нейронных сетей в зависимости от вида архитектуры. В частности, мы рассмотрим:

• Полносвязные нейронные сети (FNN): базовая архитектура, которая используется для решения различных задач машинного обучения.

• Сверточные нейронные сети (CNN): специализированная архитектура, предназначенная для обработки данных с пространственной структурой, таких как изображения.

Мы проанализируем влияние следующих гиперпараметров на качество аппроксимации:

• Количество скрытых слоев

• Количество сверток

• Функции активации

• Размер сверток

Мы проведем обширный обзор литературы по этой теме и разработаем экспериментальную методологию для сравнения качества аппроксимации различных архитектур нейронных сетей на задаче классификации изображений на датасете MNIST.

Наши результаты помогут выявить архитектуры нейронных сетей, которые обеспечивают наилучшее качество аппроксимации для различных задач, и предоставят ценную информацию для разработки и применения нейронных сетей в реальных приложениях.

**Список литературы**

1. Jozsef Suto. The effect of hyperparameter search on artificial neural network in human activity recognition. 2021
2. Clay Lafayette Mcleod. The Effect Of Hyperpar ect Of Hyperparameters In The Activ ameters In The Activation La ation Layers Of Deep ers Of Deep Neural Networks. University of Mississippi. 2016
3. Aleksandr Beknazaryan. Deep neural network approximation of analytic functions. University of Twente. 2021
4. Alexios Koutsoukas, Keith J. Monaghan, Xiaoli Li and Jun Huan. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data.

# Yongtao Lu. Influence of the parameters of the convolutional neural network model in predicting the effective compressive modulus of porous structure.

1. Р.А. Ешенко. Исследование влияния параметров нейронной сети на качество распознавания простых образов.
2. Daniel Jönsson. Visual Analysis of the Impact of Neural Network Hyper-Parameters
3. В. А. Лоренц. Эффект влияния количество нейронов сети на параметры ее обучаемости.

**Обзор литературы**

В ИНС функция активации определяет выходной сигнал нейрона. В простейших ситуациях такая функция может быть двоичной, то есть нейрон либо возбуждается и выдаёт ненулевой сигнал, либо нет. Такая функция называется пороговой. Для решения нетривиальных задач необходимо использовать нелинейные функции активации, чтобы уловить сложные зависимости в данных. Также функция должна быть дифференцируемой, так как алгоритм обратного распространения ошибки работает с производными.

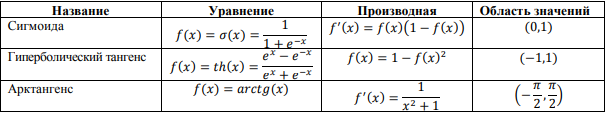
В работе [6] был проведен эксперимент, в котором рассматривались некоторые функции активации 

рисунок 1 - Функции активации

Видно, что производная сигмоиды и гиперболического тангенса выражаются через саму функцию, что позволяет уменьшить количество вычислений в методе обратного распространения ошибки.

Был проведен эксперимент обучения сети с разных функциями активации со скрытым слоем и без него.

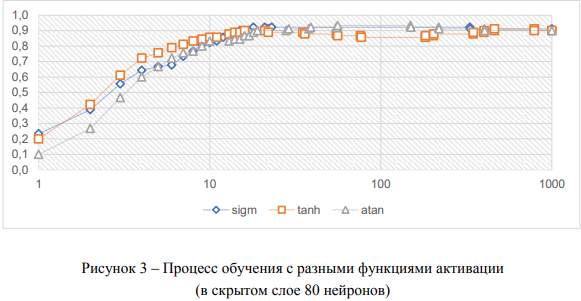


рисунок 2 - Процесс обучения с разными функциями активации

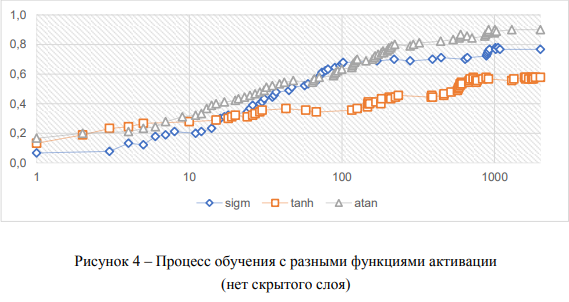


рисунок 3 - Процесс обучения с разными функциями активации

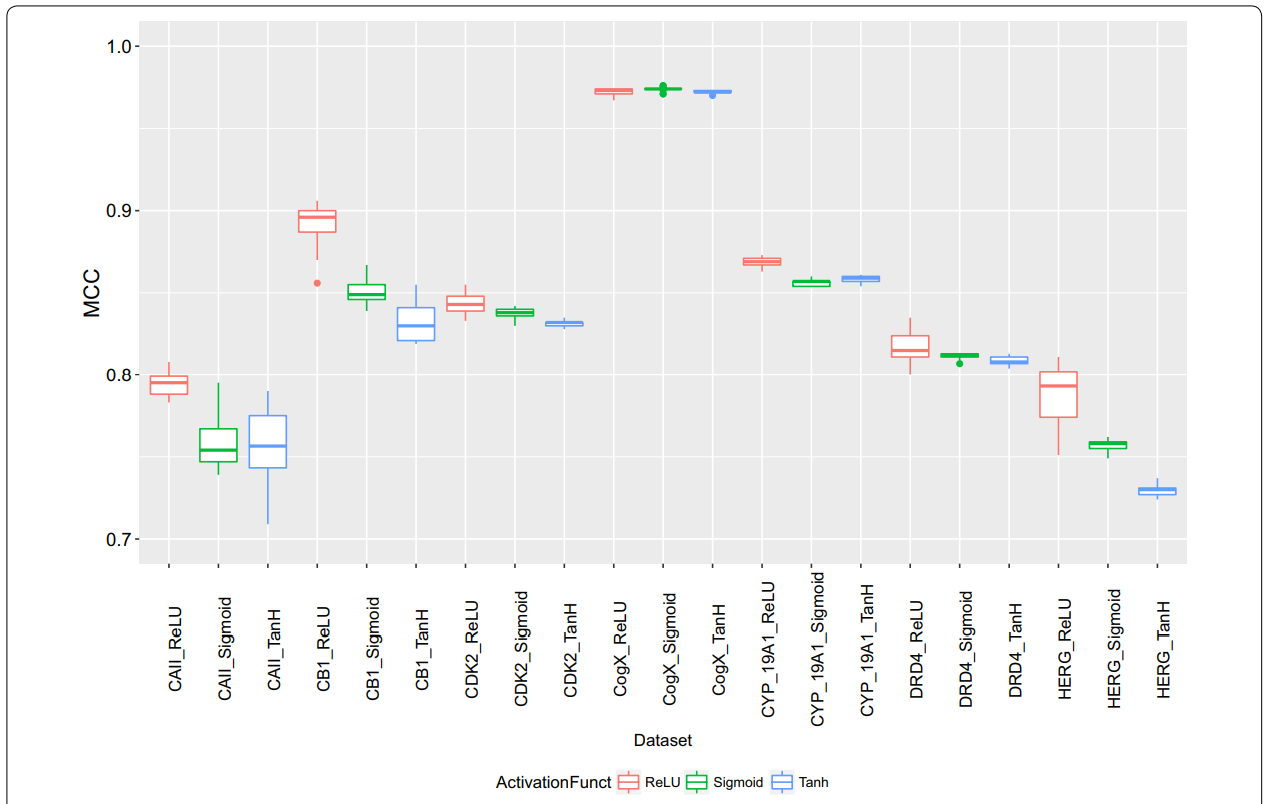
На рисунке 3 можно видеть, что ситуация иная. В ИНС без использования скрытого слоя процесс обучения при заданном шаге длится большее количество эпох . При этом разные функции активации выходят на разную эффективность. Из этого мы можем понять, что, если потребуется простая ИНС, не требующая ни одного скрытого слоя, эффективнее будет использовать функцию арктангенса.

В статье [4] был приведено сравнение производительности функций активации. Наборы данных о биоактивности, собранные из репозитория ChEMBL были использованы в этом исследовании. Использовались следующие функции активации ReLU, Sigmoid and Tanh.

Однослойные скрытые нейронные сети с различными количество нейронов {5, 10, 50, 100, 200, 500, 700, 1000, 1500, 2000, 2500, 3000, 3500}, learning rate = 0.1.

Производительность измерялась с помощью MCC

Результат эксперимента представлен на рисунке \*



Видно, что ReLU показывает лучшие результаты производительности, это следует из того что ReLU имеет очень простой вид.

В работе [8] рассматривалась зависимость качества нейронной сети от количества нейронов.

Решалась задача предсказания функции sin(x).

В ходе проведенного компьютерного эксперимента процесс обучения нейронных сетей запускался при разных значениях числа нейронов с прочими одинаковыми параметрами. Полученные результаты представляли в нормированном виде. Затем по сгенерированным данным были построены зависимости средней ошибки обучения, средней ошибки прогноза и среднего темпа обучения нейронной сети от числа нейронов, представленные на рис. 1.

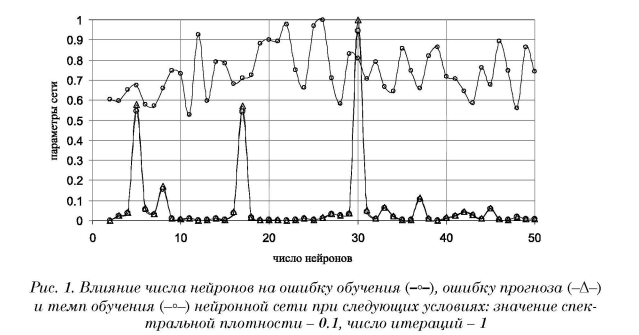
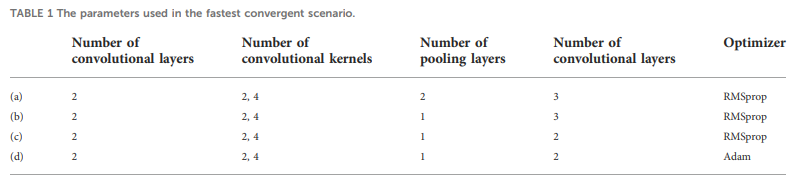
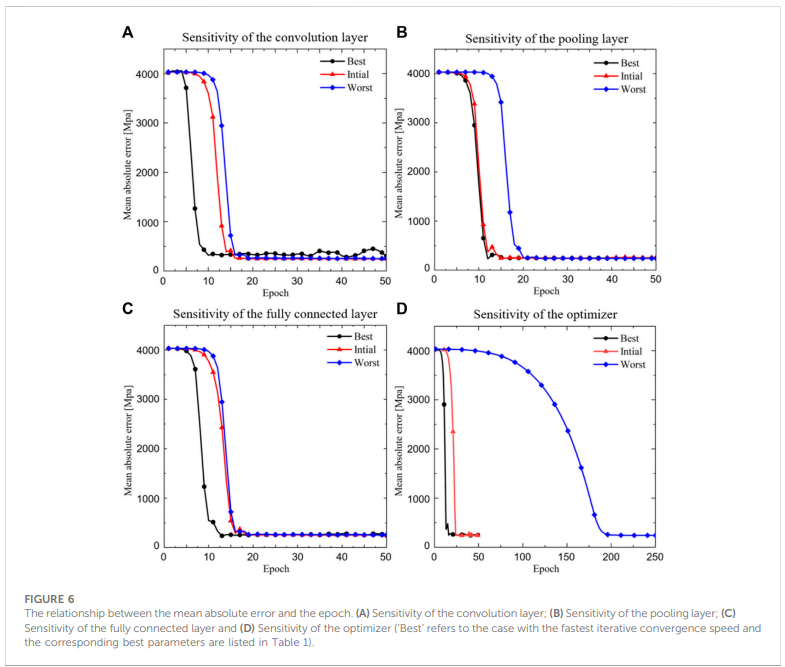


рисунок 5 - Влияния числа нейронов на ошибку обучения, ошибку прогноза и темп обучения нейронной сети при следующих условиях

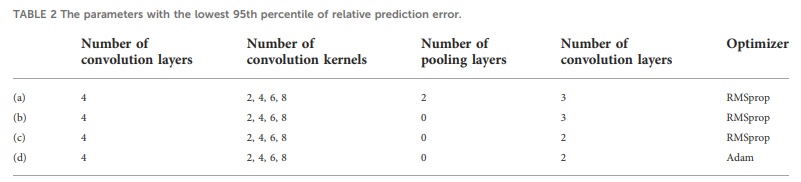
Кривая среднего темпа обучения, представленная на рис. 5, отражает отсутствие направленного влияния числа нейронов на темп обучения нейронной сети, так как средний темп обучения с ростом числа нейронов сети не демонстрирует явного тренда.

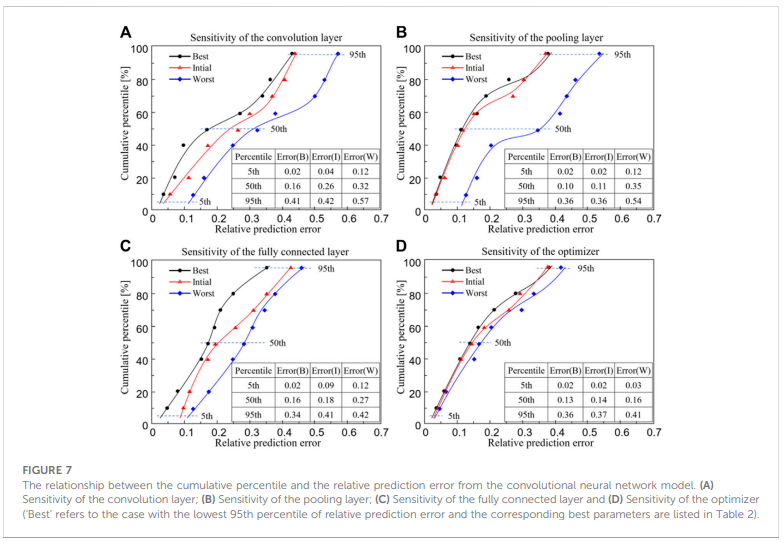
В статье [5] были проведены рассматривался вопрос влияние параметров на скорость сходимости и предсказательную модель CNN.





На рисунке представлена связь между средней абсолютной ошибкой и эпохой. (А) Чувствительность слоя свертки. (В) чувствительность объединяющего слоя. (С) чувствительность полносвязного слоя и (D) чувствительность оптимизатора





На рисунке представлена Связь между совокупным процентилем и относительной ошибкой прогнозирования модели сверточной нейронной сети. (А) Чувствительность слоя свертки; (B) чувствительность объединяющего слоя; (C) чувствительность полносвязного слоя и (D)

чувствительность оптимизатора

**Эксперимент**

В этом разделе рассматриваются различные архитектуры полносвязных и сверточных нейронных сетей (НС) для задачи классификации изображений. В качестве обучающих данных использовался широко распространенный набор данных MNIST, состоящий из черно-белых изображений рукописных цифр. Весь код НС был реализован на Python с использованием популярной библиотеки глубокого обучения PyTorch.

**Полносвязные нейронные сети**

Для задачи классификации изображений была разработана простая полносвязная нейронная сеть, содержащим 1024 нейрона. Сеть была обучена с использованием оптимизатора стохастического градиентного спуска (SGD) с начальным коэффициентом обучения 0,01 и моментумом 0,9.

Была рассмотрена ошибка (loss) и доля правильных ответов (accuracy) на валидации и на тесте для сетей с разным количеством слоев, но с зафиксированным количеством нейронов.

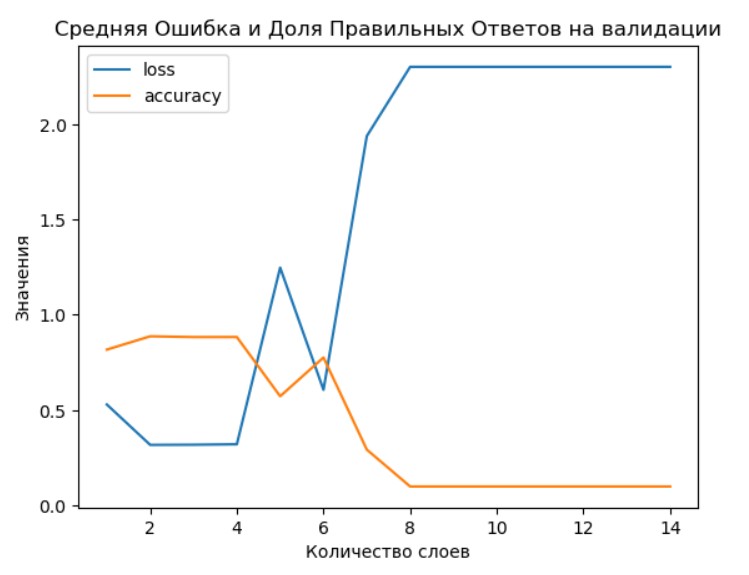
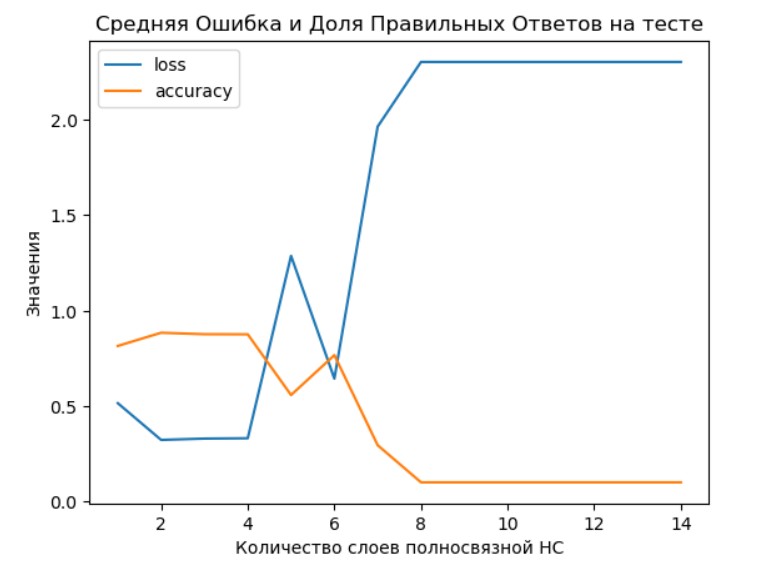


рисунок 1 - Зависимость средней ошибки и доли правильных ответов на валидации и тесте от количества слоев полносвязной нейронной сети

Из рис. 1 видно, что лучшее качество показывают сети с 2, 3, 4 слоями. Также видно, что при увеличении глубины нейронная сеть не обучается.

**Функции активации**

На рис. 2 рассмотрены средняя ошибка и доля правильных ответов на валидационной выборке для различных функций потерь (ReLU, Sigmoid, Tanh). Из рис. 2 видно, что тангенциальная гиперболическая функция (tanh) последовательно показывает наилучшую точность на всех глубинах сети.

С другой стороны, функция ReLU демонстрирует наименьшую ошибку на всех глубинах сети. Это связано с тем, что ReLU не страдает от проблемы исчезающего градиента, которая может возникать в других функциях активации.

Сигмоидальная функция активации имеет наихудшие показатели как по точности, так и по ошибке, особенно в более глубоких сетях. Это связано с тем, что сигмоидальная функция страдает от проблем с насыщением, что затрудняет обучение глубоких сетей.

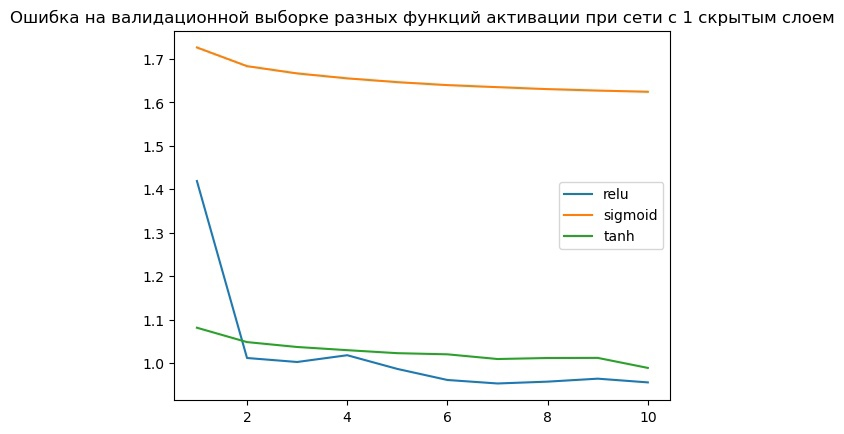
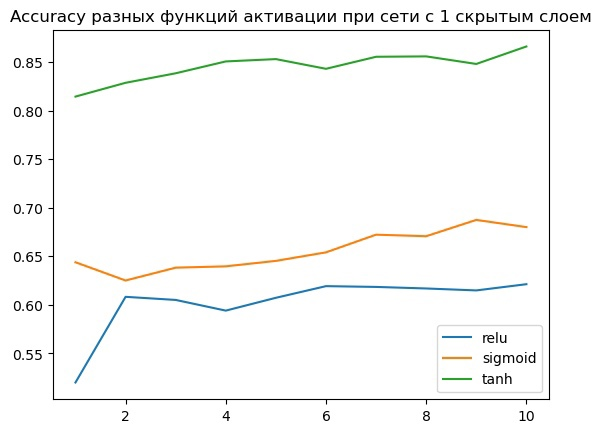


рисунок 2.1 - Ошибка и доля правильных ответов (accuracy) на валидации разных функций активации в нейронной сети с одним скрытым слоем

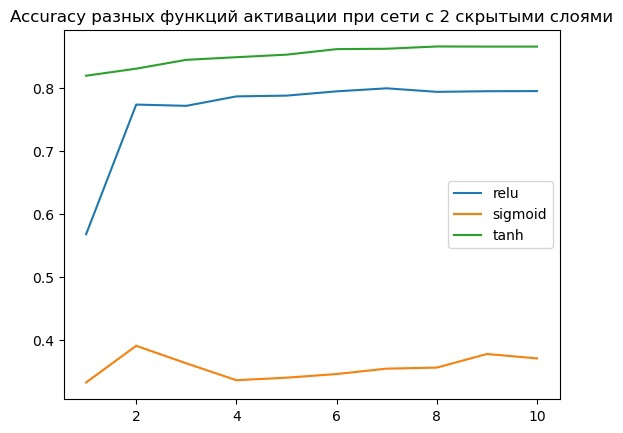
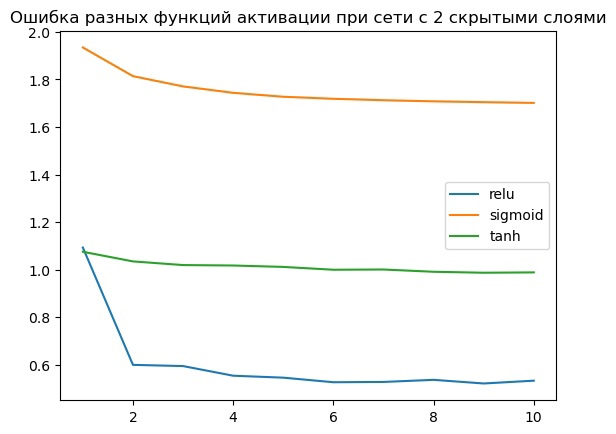


рисунок 2.2 - Ошибка и доля правильных ответов (accuracy) на валидации разных функций активации в нейронной сети с двумя скрытыми слоями

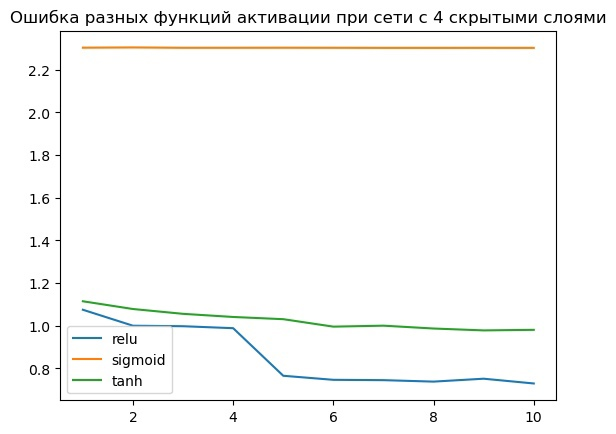
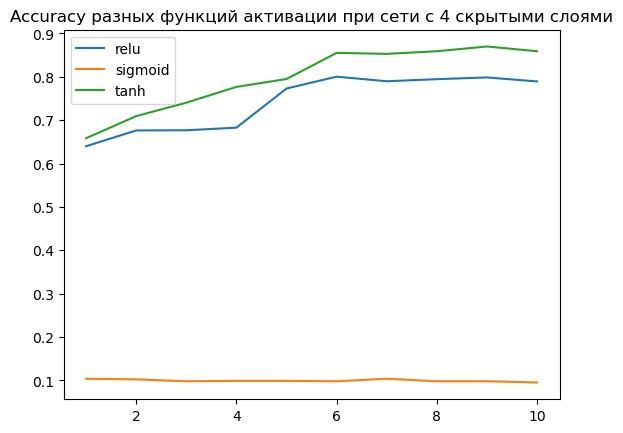


рисунок 2.3 - Ошибка и доля правильных ответов (accuracy) на валидации разных функций активации в нейронной сети с четырьмя скрытыми слоями

**Сверточные нейронные сети**

В этом разделе были исследованы различные архитектуры СНС для задачи классификации изображений. СНС используют сверточные слои для извлечения пространственных признаков из данных изображений, что делает их особенно подходящими для этой задачи

В сверточных нейронных сетях был использован оптимизатор SGD с learning rate = 0.01 и momentum = 0.9.

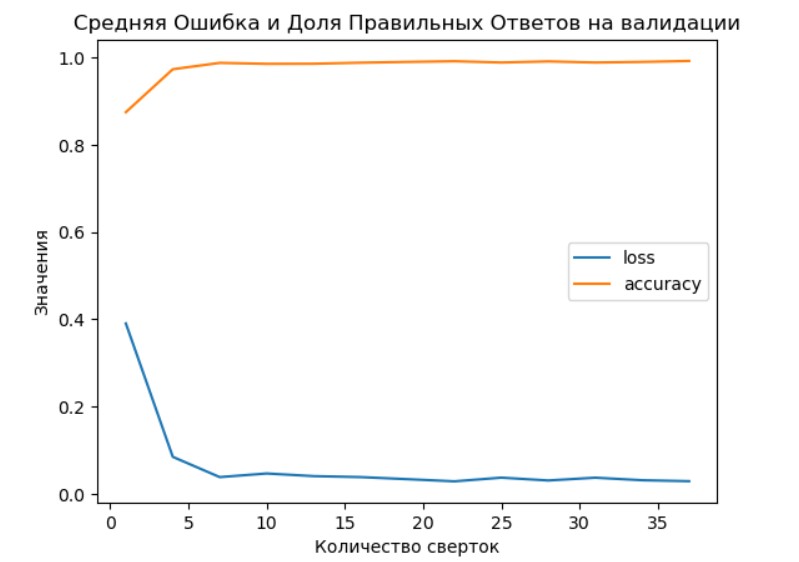
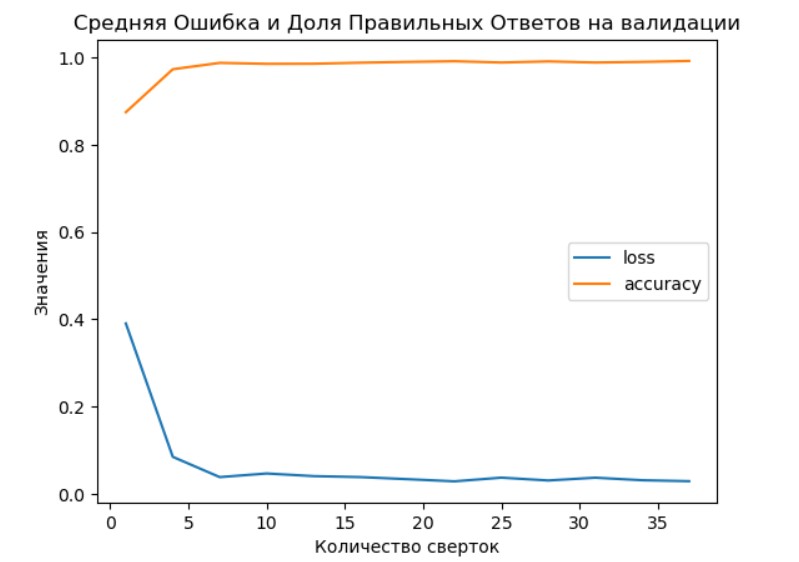


рисунок 3 - Ошибка и доля правильных ответов (accuracy) на валидации в зависимости от количества сверток

Для начала было рассмотрено качество нейронной сети в

зависимости от количества свёрток в слоях. Из рис. 3 видно что

увеличение количества сверток особо не влияет на качество сети. Оптимальным количеством оказалось от 7 сверток на слой. Большее количество сверток находит больше паттернов в изображении, следовательно, позволяет сети улавливать более широкий спектр признаков. Но увеличение количества сверток на слой также увеличивает вычислительные затраты модели.

В следующих экспериментах фиксируем количество весовых коэффициентов. Рассмотри качество в зависимости от размера сверток, а также от количества слоев

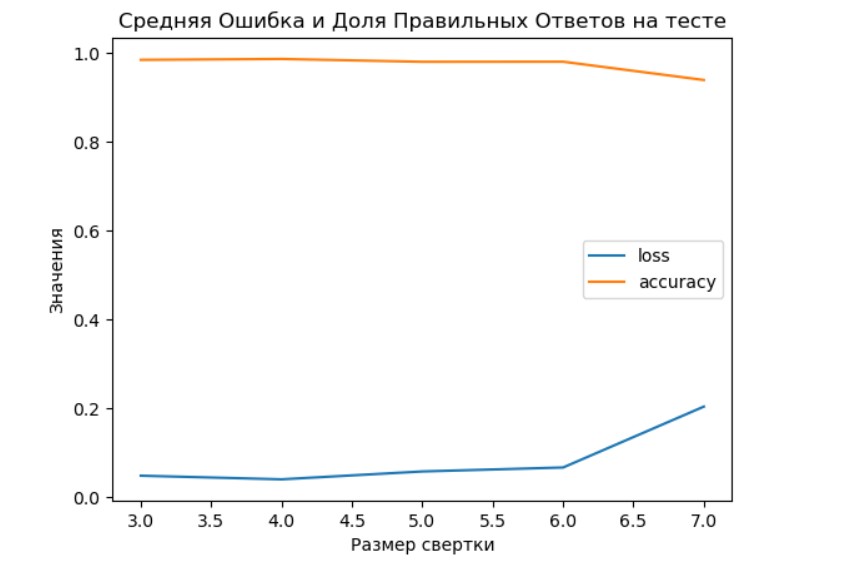
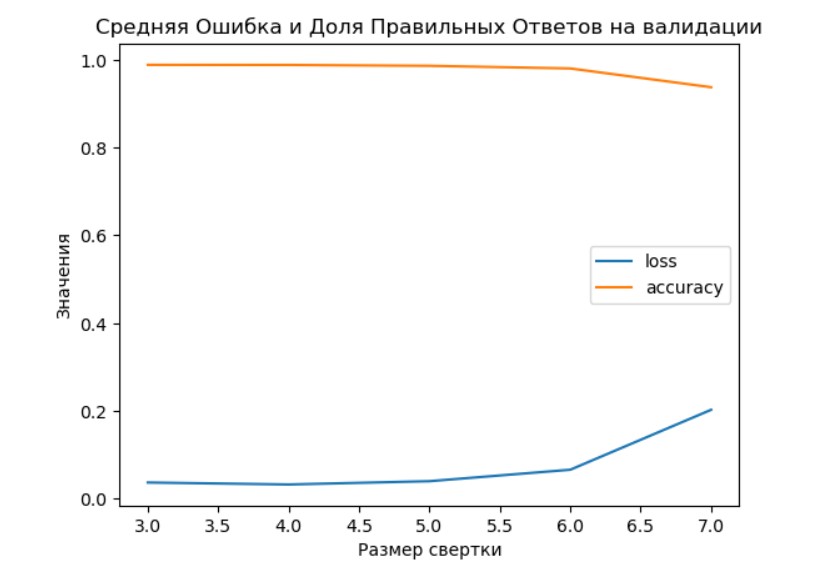


рисунок 4 - Ошибка и доля правильных ответов (accuracy) на валидации в зависимости от размера сверток

Из рисунка 4 следует, что оптимальный размер ядра свертки 3x3 или 4x4. Это связано с тем, что более крупные размеры свертки могут улавливать слишком много контекста из данных изображения, что может привести к переобучению и снижению обобщающей способности модели.

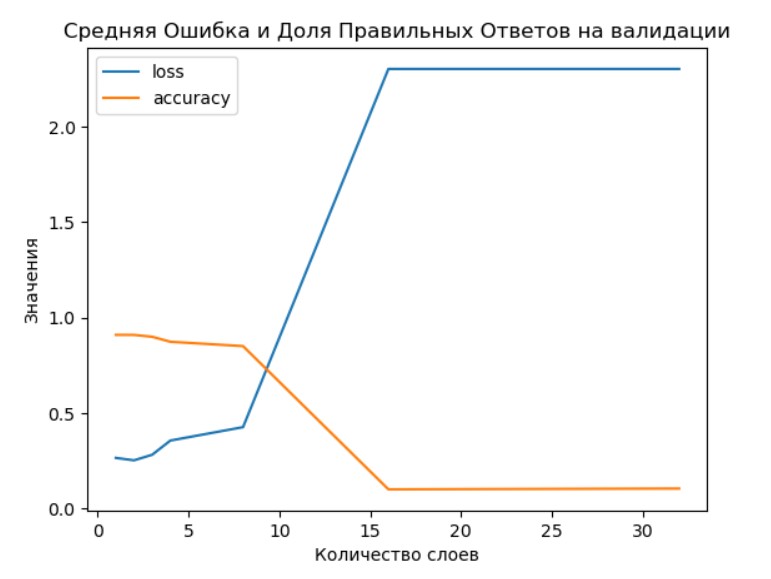
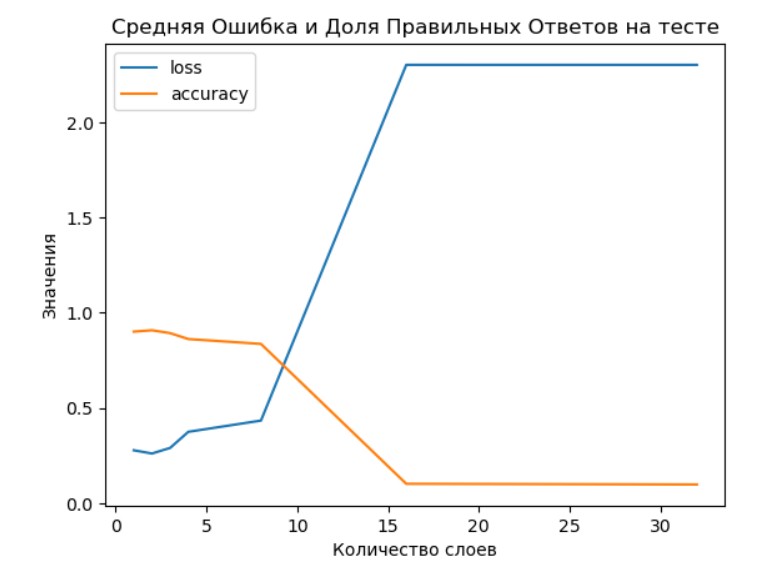


рисунок 5 - Ошибка и доля правильных ответов (accuracy) на валидации в зависимости от количества слоев

Из рис. 5 видно, что оптимальное количество слоев от 2 до 3. Качество после 2, 3 слоя уменьшается. Это может быть связано с переобучением, так как исходные данные обладают довольно просто структурой, также о плохом качестве может судить затухающий градиент

**Заключение**

В ходе курсовой работы был изучен вопрос о качестве аппроксимации нейронных сетей от вида архитектуры. Изучение проводилось на основе анализа научных статей и собственного эксперимента.

В результате проведенного исследования были рассмотрены следующие гипперпараметры:

Количество нейронов и слоев в полносвязных сетях, а также количество, размер сверток, количество сверточных слоев в сверточных сетях

Проведенный эксперимент подтвердил влияние гипперпараметров на качество нейронной сети. Было показано, что при правильном подборе гипперпараметров можно добиться существенного повышения точности сети.

Данное исследование может быть полезно для разработки более эффективных архитектур нейронных сетей.