

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

**РЕАЛИЗАЦИЯ СИСТЕМЫ РАСПОЗНАВАНИЯ ЖЕСТОВ РУК С
ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студентки 4 курса 441 группы

направления 02.03.03 - Математическое обеспечение и администрирование
информационных систем

факультета компьютерных наук и информационных технологий

Куприяновой Валерии Владимировны

Научный руководитель

Кандидат ф.-м. наук, доцент

М.В. Огнева

Зав. кафедрой

Кандидат ф.-м. наук, доцент

М.В. Огнева

Саратов, 2023

ВВЕДЕНИЕ

Актуальность темы. Потеря слуха, даже частичная, создает барьер между человеком и обществом, затрудняет овладение знаниями и специальностью, ограничивает трудовую и общественную деятельность, задерживает развитие личности. Средства для коммуникации людей с помощью жестов мало развиты на сегодняшний день. Основная цель коммуникации заключается в достижении взаимопонимания и обеспечении взаимодействия людей.

Поэтому в настоящее время много исследований направлено на решение задач распознавания жестового языка с применением компьютерного зрения и искусственного интеллекта. Достаточно частыми являются решения и подходы с использованием нейронных сетей [1, 2].

Актуальность рассматриваемой темы заключается в том, что возможность управления работой объектов без голоса и тактильных команд полезна и достаточно востребована [2]. Такие системы распознавания способны улучшить качество жизни людей с различными проблемами со здоровьем – зрением, нервной системой, слухом и речью.

Цель бакалаврской работы – реализация системы распознавания языка жестов на изображении с помощью машинного обучения.

Задачи:

1. Исследовать существующие решения.
2. Рассмотреть инструменты машинного обучения.
3. Изучить основные понятия языка жестов.
4. Изучить подходы к реализации модели, распознающей язык жестов на изображении.
5. Подготовить и проанализировать данные для обучения.
6. Реализовать и обучить модели.

7. Исследовать параметры моделей.
8. Оценить результаты и сделать выводы.

Методологические основы реализации систем глубокого обучения представлены в работах Жерона О. [3], Граса Дж. [4], Бенджио И., Курвилль А., Гудфеллоу Я. [5], Рашки С. [6], Булыгин Д.А., Мамонова Т.Е. [7].

Практическая значимость бакалаврской работы. Была реализована сверточная нейронная сеть для распознавания жестов рук, подобраны оптимальные гиперпараметры для нее, собраны данные для проверки. Нейронная сеть достигла точности распознавания в 78.37% на собранных данных.

Структура и объем работы. Бакалаврская работа состоит из введения, 8 разделов, заключения, списка использованных источников и приложения. Общий объем работы – 69 страниц, из них 56 страниц – основное содержание, включая 35 рисунков, список использованных источников информации – 21 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Обзор готовых решений для задачи распознавания жестов рук» содержит описание существующих решений для задачи распознавания жестов рук.

Второй раздел «Глубокое обучение и нейронные сети» посвящен описанию архитектуры и свойств нейронной сети.

Подраздел 2.1 посвящен понятию искусственной нейронной сети.

Подраздел 2.2 содержит описание архитектуры нейронной сети. Здесь рассмотрены понятия «слои», «веса», «входные» и «выходные» нейроны.

Подраздел 2.3 посвящен математическому описанию искусственной нейронной сети. С математической точки зрения, искусственный нейрон — это взвешенный сумматор, определенный следующим образом:

$$y = f(\sum_{i=1}^n \omega_i x_i + \omega_0 x_0),$$

где ω_i — веса входных нейронов, x_i — значения сигналов входных нейронов, функция f — функция активации нейрона, устанавливаемая в зависимости от типа задачи и, как правило, нелинейна, и y — выходной сигнал (значение) нейрона.

Обучение нейросети заключается в нахождении весовых коэффициентов входных нейронов для каждого нейрона в зависимости от разницы между действительным и полученным ответами.

Третий раздел «Сверточная нейронная сеть» посвящен описанию архитектуры и свойств сверточной нейронной сети.

В подраздел 2.1 описывается архитектура сверточной нейронной сети и ее слои. Сверточная нейронная сеть состоит из трех основных видов слоев: сверточный слой, субдискретизирующий слой и выходной слой (чаще всего полносвязный). Сверточный и субдискретизирующий (пулинговые) слои считаются слоями двумерной и трехмерной размерности, а выходной слой, как правило, представляет собой вектор из пространства \mathbb{R} .

В подразделе 3.2 описывается математическая модель сверточной нейронной сети. Здесь даны математические описания сверточного, субдискретизирующего и выходного слоев.

Схема сверточного слоя изображена на рисунке 1.

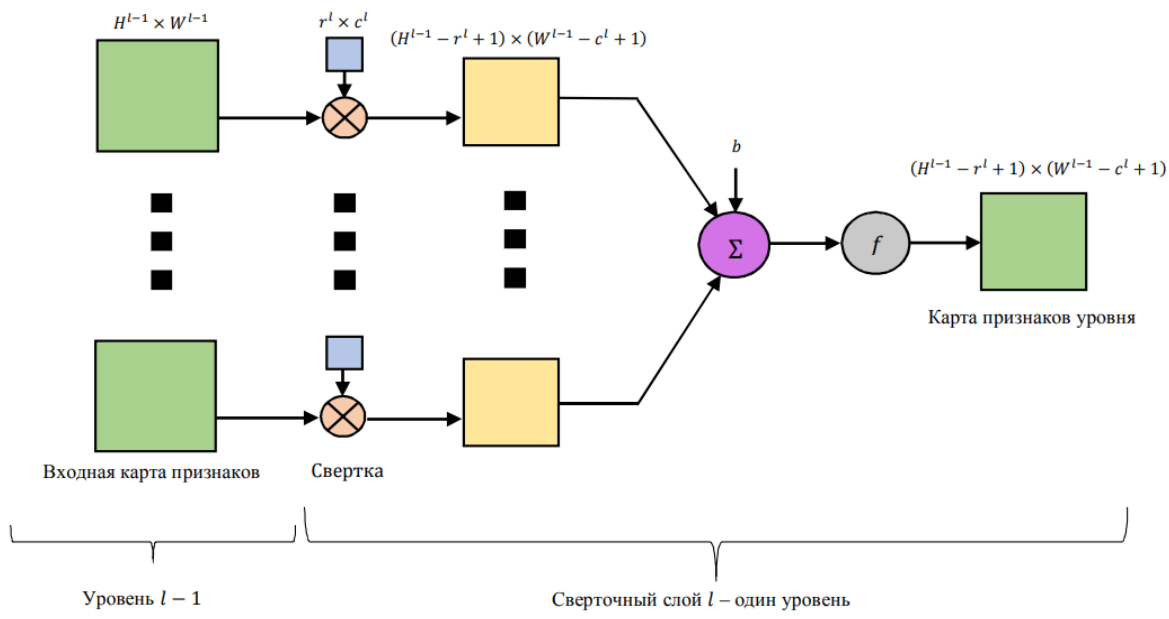


Рисунок 1 – Схема сверточного слоя

На рисунке 2 изображена схема субдискретизирующего слоя.

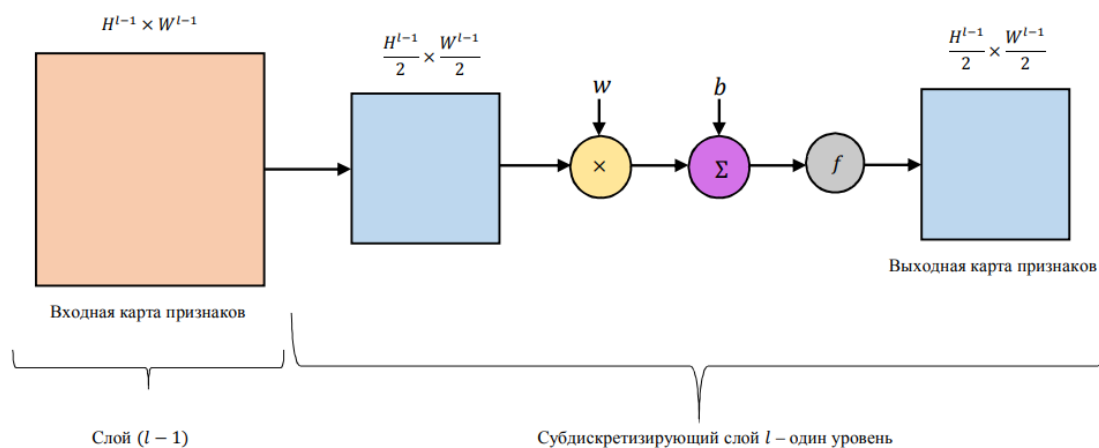


Рисунок 2 – Схема субдискретизирующего слоя

Четвертый раздел «Обучение нейронной сети» посвящен описанию алгоритма обучения нейронной сети, который заключается в минимизации функции ошибки.

Пятый раздел «Способы настройки гиперпараметров» содержит описание способов настройки гиперпараметров нейронной сети. Здесь в подразделе 5.1. приведено описание и сравнение следующих оптимизаторов градиентного спуска:

1. Метод стохастического градиентного спуска.

2. Модификация Momentum.
3. Метод Нестерова (NAG).
4. Модификация Avagrad.
5. RMSProp.
6. Модификация Adadelta.
7. Модификация Adam.

В подразделе 5.2. приведено описание и сравнение следующих функций активации:

1. Sigmoid
2. ReLU
3. Tanh
4. Elu
5. Softmax

Шестой раздел «Особенности языка жестов» посвящен описанию особенностей языка жестов и дактильного алфавита. Непосредственно в работе был использован американский жестовый алфавит (ASL).

Седьмой раздел «Библиотеки и инструменты» посвящен описанию библиотек для реализации систем машинного обучения.

Восьмой раздел «Реализация моделей для распознавания жестов» посвящен описанию процесса реализации нейронной сети, распознающей жесты рук, подбору оптимальных гиперпараметров для нее и проверке на новых данных.

Подразделе 8.1 содержит описание набора данных для обучения (Sign Language MNIST).

Подраздел 8.2 содержит анализ и описание предподготовки данных. Здесь визуализировано разбиение набора на классы, матрица корреляций, также визуализированы сами данные. Результат визуализации данных показан на рисунке 3.

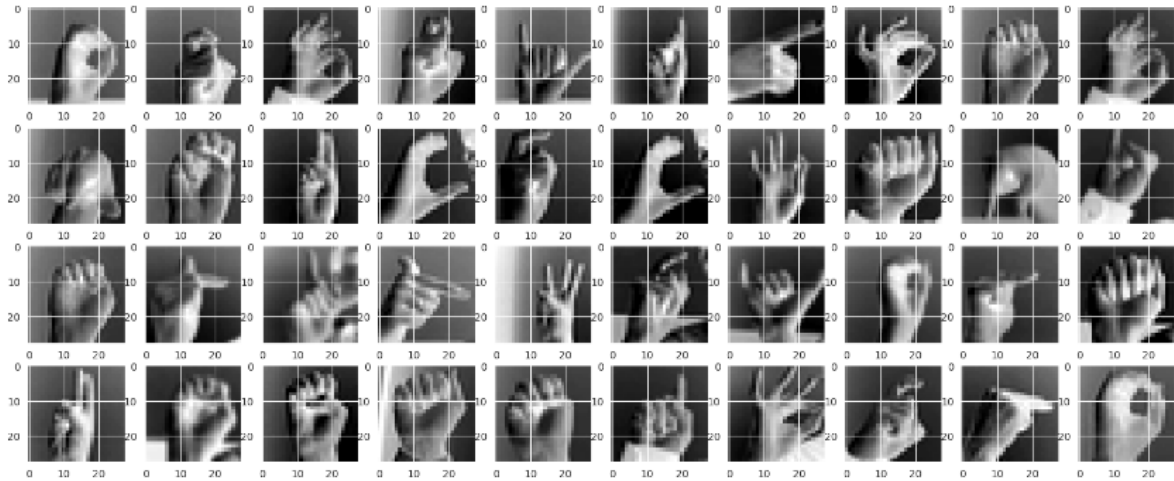


Рисунок 3 – Визуализация подготовленных данных

Анализ данных показал хорошее разбиение, отсутствие мультиколлинеарности и пропусков.

Подраздел 8.3 содержит описания процесса исследования гиперпараметров нейронной сети. В работе были исследованы следующие параметры: архитектура нейронной сети, оптимизаторы и функции активации.

Архитектуры всех сетей, участвующих в исследовании перечислены в таблице 1.

Таблица 1 – Архитектуры нейронных сетей. Здесь СС – слой свертки, СД – слой дискретизации, ПС – полносвязный слой, Ф – слой уменьшения размерности (flatten), Р – слой регуляризации, Н – слой нормализации, В – выходной слой.

Архитектура\№ слоя		1	2	3	4	5	6	7	8	9
Тип слоя	Сеть с одним сверточным слоем	СС	Н	СД	Ф	В				
	Сеть с двумя сверточными слоями	СС	Н	СД	СС	Р	Н	СД	Ф	В
	Многослойный персептрон	Ф	ПС	ПС	В					
	Сеть с двумя сверточными	СС	СД	СС	Р	СД	Ф	В		

	слоями без нормализации									
--	----------------------------	--	--	--	--	--	--	--	--	--

В результате исследования сеть с двумя сверточными слоями и слоем нормализации показала наивысшую точность на 8 эпохе обучения.

Затем было исследовано влияние оптимизатора на качество распознавания. В исследовании участвовали следующие оптимизаторы:

1. Adam
2. Adadelta
3. SGD
4. Adagrad
5. RMSprop

Исходя из полученных результатов исследования, был сделан вывод, что оптимизатор Adam справляется лучше всех в решении данной задачи.

Затем было исследовано влияние функции активации на качество распознавания. В исследовании участвовали следующие функции активации для сверточных слоев в сверточной нейросети и полносвязного слоя в многослойном перцептроне:

1. ReLU
2. Sigmoid
3. Tanh
4. SeLu
5. Elu

Наилучшие результаты показали следующие сочетания: функция активации ReLu и оптимизатор Adam или оптимизатор RMSprop.

Подраздел 8.4 содержит описание модели с лучшими гиперпараметрами и подведенные итоги. В результате проверки моделей на тестовой выборке набора данных MNIST максимальной точности в 98.63% и потери 0.0394 на тестовой выборке после рекордных 8 эпох обучения из 10 достигла сверточная нейронная сеть с двумя слоями свертки, слоем нормализации, оптимизатором

градиентного спуска Adam и функцией ReLu в качестве функции активации на сверточных слоях.

Подраздел 8.5 посвящен оценке модели на собственных собранных данных. В процессе работы были собраны собственные данные для тестирования нейронной сети: сделаны фото жестов рук разных людей, приведены к нужному формату и обработаны до формата входных данных нейронной сети с помощью библиотеки Pillow. Пример сделанных изображений представлен на рисунке 4.

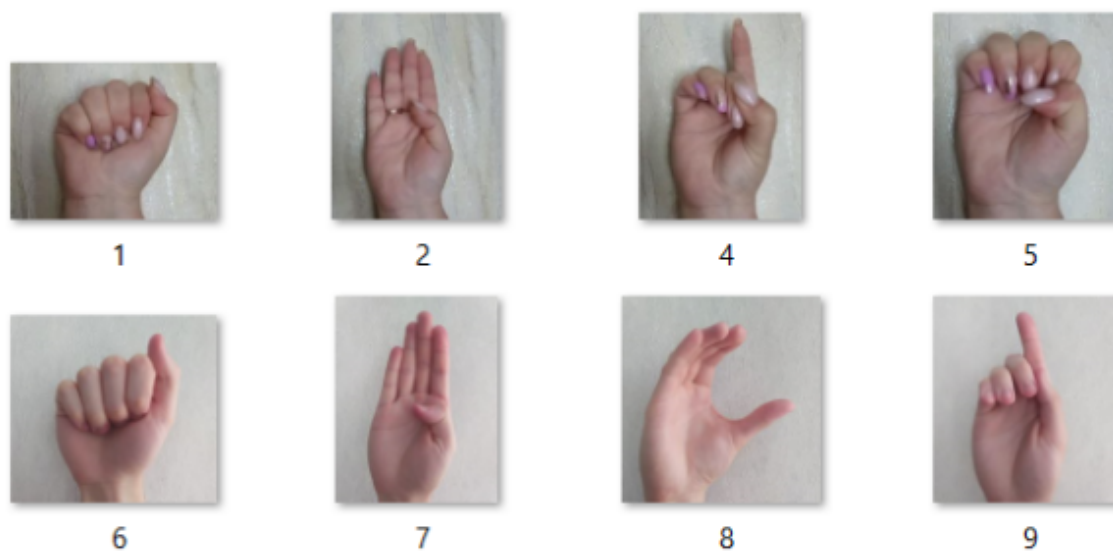


Рисунок 4 – Фрагмент изготовленных данных

Общее количество объектов в наборе – 148.

На данных, после их предобработки, были протестированы все реализованные нейросети, архитектура с полученными точностями распознавания которых изображены на таблице 8.

Таблица 8 – Метрики качества для каждой архитектуры нейронной сети.

Архитектура	Accuracy, %	Precision, %	Recall, %	F1-score, %
Сеть с одним сверточным слоем	71,62	81,60	71,62	72,52

Сеть с двумя сверточными слоями с нормализацией	78,37	86,60	78,37	79,58
Многослойный перцептрон	48,64	61,97	48,64	50,27
Сеть с двумя сверточными слоями без нормализации	70,27	81,45	70,27	72,39

По результатам тестирования на новых данных наивысшая точность была получена на сверточной нейронной сети с двумя слоями свертки – 78,37%. Таким образом на сырых данных, которые представляют собой фото жестов рук разных людей, точность распознавания ощутимо ниже, но остается достойной для многоклассовой классификации.

Также в данном подразделе были описаны выделенные группы жестов с худшим и лучшим качествами распознавания.

ЗАКЛЮЧЕНИЕ

В ходе проделанной работы были рассмотрены инструменты глубокого машинного обучения для реализации распознавательных систем и готовые решения. Была выполнена цель работы, которая заключалась в реализации искусственной нейронной сети, распознающей жесты рук. Также были исследованы ее гиперпараметры и найдены оптимальные. Помимо этого для тестирования был собран собственный набор проверочных данных. Для решения задачи была использована библиотека Keras и другие инструменты, а результатом работы стала нейронная сеть, которая достигла точности распознавания жестов рук в 98.63% на подготовленной тестовой выборке набора MNIST и 78.37% на новых данных.

В конечном итоге, разработки и исследования в данном направлении могут быть использованы в более масштабных системах, которыми смогут пользоваться люди с ограниченными возможностями. Реализованная нейронная сеть имеет большой потенциал для дальнейшего улучшения и оптимизации, в

том числе и расширения набора жестов рук для еще более точного распознавания.

Основные источники информации:

1. Бычков А. Г., Киселева Т. В., Маслова Е. В. Использование сверточных нейросетей для классификации изображений / Вестник СГИУ. 2023. №1 (43).
2. Хнюнин М. В., Гриф М. Г. Система распознавания жестов в реальном времени с использованием Mediapipe / материалы XXIII Международной научно-практической конференции им. Э.К. Алгазинова, Воронежский государственный университет. 2023 -54-57с.
3. Жерон О., Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow — М.: ООО "Диалектика" 2020. -1040с.
4. Грас Дж. Наука о данных с нуля 2-е изд.: ВHV, 2022. -336 с.
5. Бенджио И., Курвилль А., Гудфеллоу Я. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК. Пресс, 2018.
6. Рашка С., Мирджалили В. Python и машинное обучение. Машинное и глубокое обучение с использованием Python, scikit-learn : Диалектика 2020. -848 с.
7. Булыгин Д.А., Мамонова Т.Е. Распознавание жестов рук в режиме реального времени / Научный вестник НГТУ Science Bulletin of the NSTU том 78, 2020. № 1
8. De Smedt Q., Wannous H., Vandeborre J-P. Heterogeneous hand gesture recognition using 3D dynamic skeletal data. / Computer Vision and Image Understanding. 2019. №181. - Яз. англ.
9. Dong J., Xia Z., Yan W., Zhao Q. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time Journal of Visual Communication and Image Representation, 2020. № 2. - Яз. англ.
10. Раднаев Ч. Б. Оптимизация гиперпараметров сверточной нейронной сети / Молодежь и современные информационные технологии : сборник трудов

XV Международной научно-практической конференции студентов, аспирантов и молодых учёных, 2018. -81-82с.

11. Зайцева Г. Л. Жестовая речь. Дактилология: Учебник для студ. высш. учеб. заведений. — М.: Гуманит. изд. центр ВЛАДОС, 2004. — 192 с.