

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ**

**УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**Факультет Математики, Информационных и Авиационных технологий**

**Кафедра Информационных технологий**

**КУРСОВАЯ РАБОТА**

**по дисциплине «Web технологии»**

Разработка архитектуры системы для перевода с языка жестов в режиме реального времени

(название темы)

Прикладная информатика (Бакалавриат) - 09.03.03

(наименование и номер специальности(направления))

Проект выполнил студент \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

группа подпись, дата Ф.И.О.

Научный руководитель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

должность подпись, дата Ф.И.О.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

оценка

У Л Ь Я Н О В С К

2020 г.

Оглавление

[***Введение 3***](#_gjdgxs)

[**Актуальность исследования 3**](#_30j0zll)

[**Цели и задачи 3**](#_1fob9te)

[***1***](#_3znysh7) ***Предметная область исследований 4***

[**1.1**](#_2et92p0) **Основные понятия 4**

[***2***](#_1t3h5sf) ***Исследование данных 6***

[**2.1**](#_4d34og8) **Анализ данных 13**

[***3***](#_26in1rg) ***Подготовка данных 20***

[**3.1**](#_lnxbz9) **Заполнение пропусков 20**

[**3.2**](#_35nkun2) **Очистка данных 21**

[**3.3**](#_1ksv4uv) **Конструирование признаков (Feature Engineering) 21**

[**3.4**](#_44sinio) **Удаление лишних параметров 22**

[**3.5**](#_2jxsxqh) **Подготовка категориальных данных 22**

[***4***](#_4i7ojhp) ***Тренировка и тестирование классификатора 24***

[**4.1**](#_2xcytpi) **Создание модели 24**

[**4.2**](#_1ci93xb) **Программа 26**

[***Заключение 27***](#_3whwml4)

[***Список литературы 28***](#_2bn6wsx)

[***Приложение 29***](#_qsh70q)

[**Листинг программы 29**](#_3as4poj)

# Введение

Общение - это бесконечно важный навык, который позволяет людям обмениваться информацией. Мы общаемся с помощью речи, жестов, языка тела, письма или с помощью визуальных средств, причем речь является одним из наиболее часто используемых. Однако, к сожалению, не все люди способны эффективно коммуницировать между собой.

Люди, у которых имеются нарушения слуха и речи используют язык жестов. Язык жестов - это самостоятельный язык, состоящий из жестов, каждый из которых производится руками в сочетании с мимикой, формой или движением рта и губ, а также в сочетании с положением корпуса тела. Эти языки в основном используются в культуре глухих и слабослышащих с целью коммуникации.

В рамках данной работы я постараюсь создать систему, которая поможет преодолеть барьер между теми, у кого есть проблемы со слухом и речью, и теми, у кого таких проблем нет.

## Актуальность

Последние 10 лет бурного развития технологий машинного обучения, и компьютерного зрения в частности, вселяют оптимизм.

Буквально месяц назад (декабрь 2019 г.) компания Oculus выпустила систему воссоздания рук в виртуальной реальности. С помощью камеры и системы распознавания пользователь может взаимодействовать с объектами виртуальной реальности, просто двигая своими руками в пространстве. Данная технология уже используется для управление интерфейсом и в качестве основного способа взаимодействия в видеоиграх для виртуальной реальности.

Очевидно, что на данном этапе развития компьютерного зрения можно создать систему, которая будет осуществлять перевод жестов.

## Цели и задачи

**Цель:** Спроектировать архитектуру системы, которая будет осуществлять перевод с языка жестов.

**Задачи:**

1. Спроектировать архитектуру системы для перевода с языка жестов.
2. Создать демонстрационное веб-приложение.

# Предметная область

Перед созданием системы перевода с языка жестов стоит детально рассмотреть данный способ коммуникации. Обозначить сходства и различия с языками вербальными.

## Жестовый язык

В данной работе будут рассматриваться только сложные жестовые системы широкого применения. Такие системы, как вспомогательная жесты вербальных языков, жесты военных, жесты в различных видах спорта и т.п. имеют слишком узкую направленность, хотя, безусловно, элементы многих простых жестовых систем являются частью жестовых языков широкого применения.

К сложным жестовым системам широкого применения следует отнести жестовые системы общения некоторых племен североамериканских индейцев и австралийцев, членов ряда монашеских орденов (принимающих обет молчания) и, разумеется, дактильную и жестовую речь глухих.

Одним из главных неправильных представлений о жестовых языках является представление, что они каким-то образом зависят от словесных (звуковых и письменных) языков или произошли от них, что эти языки были придуманы слышащими, однако это не так.

Жестовые языки почти полностью независимы от словесных и они продолжают развиваться: появляются новые жесты, отмирают старые — и чаще всего это мало связано с развитием словесных языков. Количество жестовых языков в стране не связано с количеством в ней словесных языков. Даже в одной стране, где присутствуют несколько словесных языков, может быть единственный общий жестовый язык, и в некоторых странах даже с одним словесным языком могут сосуществовать несколько жестовых.

### Дактильная речь

Дактильная речь — это особая кинетическая система. Движения рук (жесты) в данном случае обозначают не сами явления окружающего мира, а буквы алфавитов национальных языков. Но ведь буквы сами служат для называния этих явлений, так как они являются элементами слов (единиц письменной речи). Следовательно, по структуре дактильная речь — это речь при помощи жестовых обозначений букв, иначе говоря, обозначения обозначений. Используя набор дактильных знаков, говорящий следует грамматике словесного языка (русского, английского и т. д.). Таким образом, дактильная речь — своеобразная кинетическая форма словесной речи, вербальной коммуникации.

Многие глухие используют преимущественно именно дактильную речь ввиду сложности освоения языка жестовой речи.

Дактильная речь используется в жестовых языках в основном для произнесения имен собственных, географических названий, а также специфичных терминов, взятых из словесных языков



### Жестовая речь глухих

Система жестового общения глухих имеет сложную структуру, включает две разновидности жестовой речи: русскую и калькирующую. Русская жестовая речь (РЖР) — это общение при помощи средств русского жестового языка — самобытной лингвистической системы, обладающей своеобразной лексикой, грамматикой и т.д. Калькирующая жестовая речь (КЖР) — калькирует лингвистическую структуру словесного языка. Калькирующая жестовая речь — вторичная знаковая система, которая усваивается на базе и в процессе изучения глухим ребенком словесной речи. Жесты здесь являются эквивалентами слов, а порядок их следования такой же, как в обычном предложении. [1]

# Компьютерное зрение

Под компьютерным зрением подразумевается технология создания машин, которые могут производить обнаружение и классификацию объектов. Данные технологии призваны автоматизировать процессы, которые человек выполняет с помощью своей зрительной систем.

На данном этапе развития компьютерного зрения машины уже превосходят человека в скорости и точности при решении определенных задач: классификация животных, обнаружение признаков заболеваний на медицинских снимках, анализ снимков космоса. [2]

Наиболее популярной на данной момент технологией для решения задач компьютерного зрения являются сверточные нейронные сети.

Рассмотрим сверточные нейронные сети и способы их применение.,

## Сверточные нейронные сети

Сверточные нейронные сети (Convolutional neural networks - CNN) - это подвид Нейронных сетей с прямой связью (Feedforward neural network - FFNN) - сетей без циклов или петель. Сверточные нейронные сети позволяют снизить количество параметров, без значительной потери качества модели. Сверточные нейронные сети применяются при обработке изображений и текста. [3]

Данная архитектура нейронных сетей была предложена Яном Лекуном в 1988 г.

Архитектура CNN аналогична структуре связности нейронов в человеческом мозге и была вдохновлена организацией зрительной коры головного мозга. Отдельные нейроны реагируют на стимулы только в ограниченной области зрительного поля, которые называются рецептивными. Набор таких полей перекрывается, чтобы охватить всю визуальную область.

Алгоритм получает на вход изображение, присваивает важность (обучаемые веса - weights и константный сдвиг - bias) различным объектам на изображении. Предварительная обработка, требуемая в CNN, значительно ниже по сравнению с другими алгоритмами классификации. В то время как в примитивных методах фильтры создаются вручную, при достаточном обучении CNN могут настраивать эти фильтры самостоятельно. [4]

### Архитектура сверточных нейронных сетей

В сверточной нейронной сети выходы промежуточных слоев образуют матрицу (изображение) или набор матриц (несколько слоев изображения). Так, например, на вход сверточной нейронной сети можно подавать три слоя изображения (R-, G-, B-каналы изображения). Основными видами слоев в сверточной нейронной сети являются сверточные слои (англ. convolutional layer), пулинговые слои (англ. pooling layer) и полносвязные слои (англ. fully-connected layer).

#### Сверточный слой

Сверточный слой нейронной сети представляет из себя применение операции свертки к выходам с предыдущего слоя, где веса ядра свертки являются обучаемыми параметрами. Еще один обучаемый вес используется в качестве константного сдвига - bias

#### Пулинговый слой

Пулинговый слой призван снижать размерность изображения. Исходное изображение делится на блоки размером w×h и для каждого блока вычисляется некоторая функция. Чаще всего используется функция максимума (англ. max pooling) или (взвешенного) среднего (англ. weighted average pooling). Обучаемых параметров у этого слоя нет. Основные цели пулингового слоя:

* уменьшение изображения, чтобы последующие свертки оперировали над большей областью исходного изображения;
* увеличение инвариантности выхода сети по отношению к малому переносу входа;
* ускорение вычислений. [5]

#### Полносвязный слой

Цель полносвязного слоя состоит в том, чтобы взять результаты процесса свертки/пулинга и использовать их для классификации изображения в метку (в простом примере классификации).

Выходные данные свертки / пулинга трансформируются в вектор значений, каждый из которых представляет вероятность того, что определенный объект принадлежит метке. Например, если изображение представляет собой кошку, то признаки, представляющие такие вещи, как усы или мех, должны иметь высокую вероятность для метки “кошка”. [6]

Обычно, этот слой идет последним в сверточных нейронных сетях.

### Определение позы

Для данного направления в компьютерном зрении еще нет устоявшегося перевода с английского Pose estimation. Суть состоит в том, чтобы из двухмерного изображения получить трехмерный объект. Например по фотографии человека получить его трехмерный скелет.



#### 

#### 

После получения скелета с ключевыми точками можно классифицировать не сам жест на изображении, а жест, как расстояние между ключевыми точками. Это значительно улучшит точность определения жеста.

В классификации важна общая поза, выражение лица, движения рук.

### Сегментация изображения

Под сегментацией понимается поиск нужного объекта на изображении. Для решения подобных задач применяются CNN, но, в отличие от классификации, на выходе у сегментирующей нейросети мы получим: координаты центра объекта, длину и ширину.

В машинном обучении это называется Bounding Box - рамка, с помощью которой можно будет обрезать изображение.

# Перевод с языка жестов и предсказание следующего жеста

Для работы с последовательностями используют рекуррентные нейронные сети. Это поможет предугадывать следующий жест, что, в свою очередь, приведет к более точной классификации жеста.

Более того, рекуррентные нейронные сети используются для перевода с одного языка на другой.

Рассмотрим рекуррентные нейронные сети и способы их применение.

## Рекуррентные нейронные сети

Рекуррентные нейронные сети — сети с циклами, которые хорошо подходят для обработки последовательностей. Обучение RNN аналогично обучению обычной нейронной сети. Мы также используем алгоритм обратного распространения ошибки (англ. backpropagation) но с небольшим изменением.

Поскольку одни и те же параметры используются на всех временных этапах в сети, градиент на каждом выходе зависит не только от расчетов текущего шага, но и от предыдущих временных шагов. Например, чтобы вычислить градиент для четвертого элемента последовательности, нам нужно было бы «распространить ошибку» на 3 шага и суммировать градиенты. Этот алгоритм называется «алгоритмом обратного распространения ошибки сквозь время» (англ. Backpropagation Through Time, BPTT). [7]

Рекуррентные нейронные сети основаны на работе Дэвида Румельхарта 1986г.

## Предсказание следующего жеста.

Предсказание следующего жеста относится к задачам области обработки естественных языков (англ. Natural Language Processing, NLP). Языки жестов являются естественными языками, наравне с русским, английским и т.д.

Подобная технология нашла широкое применение в наших повседневных задачах. Например набор текста. Нейронная сеть подсказывает следующее слово, которое вы бы, с наибольшей вероятностью, набрали.

Для решения подобных задач применяется модифицированная версия рекуррентных сетей, которая называется Долгая краткосрочная память (англ. Long short-term memory; LSTM).

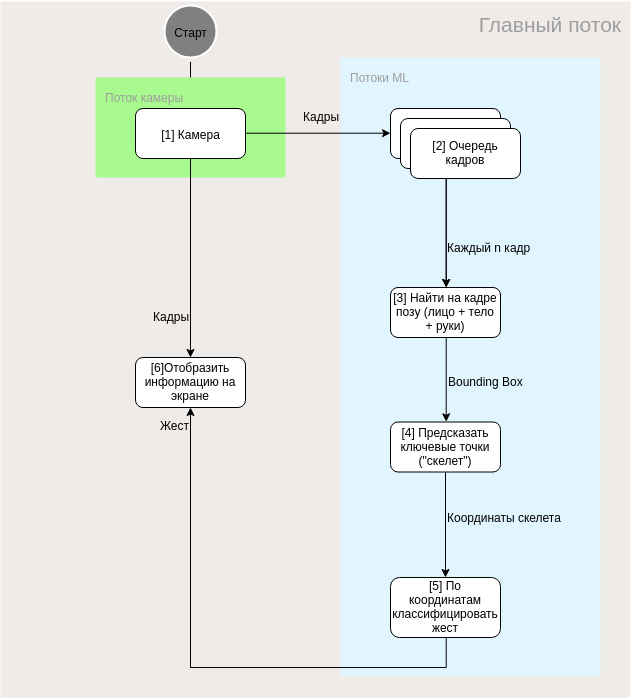
## Перевод с языка жестов

Машинный перевод (МТ) - это область компьютерной лингвистики, которая ориентирована на перевод текста с одного языка на другой. Задачи машинного перевода решаются с помощью алгоритма под названием Нейронный машинный перевод (англ. Neural Machine Translation, NMT).

В частности, именно этот алгоритм используется в Google Translate. NMT представляет собой алгоритм глубокого обучения, в котором множество наборов переведенных предложений используются для обучения модели, способной переводить между любыми двумя языками. Существует несколько вариантов NMT, которые в настоящее время исследуются и внедряются в промышленность. Одной из более старых и устоявшихся версий NMT является архитектура Кодер-Декодер. Эта архитектура состоит из двух рекуррентных нейронных сетей (RNN), используемых совместно в тандеме для создания модели - переводчика.[8]

# Архитектура приложения

Данная архитектура вдохновлена работой Виктора Мёнье. [9]



Рассмотрим каждый элемент поподробнее:

1. Кадр захватывается с камеры и помещается во входную очередь.
2. Рабочий(свободный поток) берет каждый n кадр из очереди и помещает кадр в память устройства.
3. Нейросеть находит на кадре человека и обрезает изображение (сегментация).
4. Нейросеть определяет позу и создает трехмерный объект - ключевые точки (скелет).
5. По взаимному расположению ключевых точек нейросеть предсказывает жест.
6. Изображение с камеры выводится на экран. По сигналу ML потока, главный поток получает перевод жеста и выводит информацию на экран.

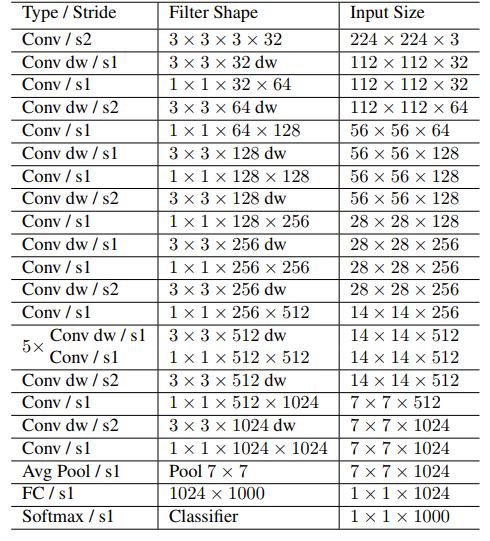
# Демо-вариант приложения

Данный вариант призван продемонстрировать работу алгоритмов сегментации изображений и их классификацию. Модель, которая была использована, была натренирована для классификации алфавита американского языка жестов.

Демо основано на работе Виктора Дибиа. [10] Он предлагает библиотеку для прототипирования и создания веб-приложений отслеживания положения кисти в режиме реального времени.

## Модель для сегментации

Модель для данного приложения выбрана как компромисс между точностью и скоростью. Обе эти характеристики, безусловно, важны. Базовой архитектурой стала “MobileNetv2”. Данная архитектура предназначена для работы на современных телефонах. Основные преимущества - это небольшой вес, высокая скорость работы и не слишком большая потеря точности, по сравнению с архитектурами, которые работают на удаленных серверах. [11]



### Набор данных для тренировки

Набор данных Ego Hands содержит 48 видеороликов Google Glass от первого лица. На видео люди производят различные взаимодействия. Основная цель Ego Hands состоит в том, чтобы предоставить качественные данные для приложений, в которых требуется распознать кисти рук.

Особенности Ego Hands:

* Высокое качество, пиксельные сегментации рук.
* Возможность семантически различать руки наблюдателя и чужие руки, а также левую и правую руку.
* Множество вариаций различных действий руками. [12]

## Модель для классификации

Модель основана на работе Ранжита Джейн. [13]

Архитектура модели:

model = Sequential()

model.add(Conv2D(64, kernel\_size=(3,3), activation = 'relu', input\_shape=(28, 28 ,1) ))

model.add(MaxPooling2D(pool\_size = (2, 2)))

model.add(Conv2D(64, kernel\_size = (3, 3), activation = 'relu'))

model.add(MaxPooling2D(pool\_size = (2, 2)))

model.add(Conv2D(64, kernel\_size = (3, 3), activation = 'relu'))

model.add(MaxPooling2D(pool\_size = (2, 2)))

model.add(Flatten())

model.add(Dense(128, activation = 'relu'))

model.add(Dropout(0.20))

model.add(Dense(num\_classes, activation = 'softmax')

### Набор данных для тренировки

Для тренировки использовался набор Sign Language MNIST. Он содержит 24 класса - жесты американской азбуки. Без букв “J” и “Z” (требуют движение). [14]

## Пайплайн

1. Камера передает изображение на экран
2. Изображение отражается по вертикали и выводится на экран
3. Нейросеть определяет границы кисти (центр ладони, размер вертикали и горизонтали).
4. По полученным данным изображение обрезается и сохраняется на устройстве.
5. Изображение классифицирует сверточная нейронная сеть.
6. Перевод жеста выводится на экран.

## Библиотеки и языки

* Языки: HTML, JS, CSS
* Библиотеки: Tensorflow

# Заключение

В ходе работы было выполнено:

* Спроектирована архитектура системы для перевода с языка жестов.
* Создано демонстрационное веб приложение.

Дальнейшие пути развития:

* Создать модель, которая будет опираться на ключевые точки, а не кадры.
* Создать модель, которая на вход принимает последовательность кадров (видео), чтобы классифицировать жесты, требующие движения.
* Использовать рекуррентные нейросети, для улучшение точности предсказания жеста.
* Создать систему, которая сможет переводить не отдельные жесты, а формировать из жестов фразы.
* Перенести приложение на мобильные устройства для удобства потенциальных пользователей.

# 

# Список литературы

1. Жестовая речь. Дактилология: Учеб. для студ. высш. учеб. заведений /

Зайцева Г.Л. Гуманит. изд. центр ВЛАДОС; 2000 г. 192 c.

1. Kevin Hartnett - Computers and Humans ‘See’ Differently. Does It Matter? www.quantamagazine.org/computers-and-humans-see-differently-does-it-matter-20190917/ дата обращения – (09.01.2020)
2. The Hundred-page Machine Learning Book / Andriy Burkov; 2019 160 c.
3. A Comprehensive Guide to Convolutional

Neural Networks — the ELI5 way - Sumit Saha www.towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 дата обращения – (02.01.2020)

1. Сверточные нейронные сети - Университет ИТМО www.neerc.ifmo.ru/wiki/index.php?title=%D0%A1%D0%B2%D0%B5%D1%80%D1%82%D0%BE%D1%87%D0%BD%D1%8B%D0%B5\_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5\_%D1%81%D0%B5%D1%82%D0%B8 дата обращения – (05.01.2020)
2. Fully Connected Layers in Convolutional Neural Networks www.missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/ дата обращения – (05.01.2020)
3. Рекуррентные нейронные сети - Университет ИТМО www.neerc.ifmo.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%BA%D1%83%D1%80%D1%80%D0%B5%D0%BD%D1%82%D0%BD%D1%8B%D0%B5\_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5\_%D1%81%D0%B5%D1%82%D0%B8 дата обращения – (05.01.2020)
4. Neural Machine Translation - Quinn Lanners www.towardsdatascience.com/neural-machine-translation-15ecf6b0b дата обращения – (14.01.2020)
5. HandPose - Victor Meunier www.github.com/MrEliptik/HandPose дата обращения – (14.01.2020)
6. How to Build a Real-time Hand-Detector using Neural Networks (SSD) on Tensorflow - Victor Dibia www.medium.com/@victor.dibia/how-to-build-a-real-time-hand-detector-using-neural-networks-ssd-on-tensorflow-d6bac0e4b2ce дата обращения – (14.01.2020)
7. MobileNetV2: Inverted Residuals and Linear Bottlenecks - Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen www.arxiv.org/pdf/1801.04381.pdf дата обращения – (14.01.2020)
8. Ego Hands Dataset www.vision.soic.indiana.edu/projects/egohands/ дата обращения – (14.01.2020)
9. Deep learning using sign language - Ranjeet Jain www.kaggle.com/ranjeetjain3/deep-learning-using-sign-langugage дата обращения – (14.01.2020)
10. Sign Language MNIST www.kaggle.com/datamunge/sign-language-mnist/kernels дата обращения – (14.01.2020)
11. Handtrack.js - Victor Dibia www.github.com/victordibia/handtrack.js дата обращения – (14.01.2020)

# Приложение

## Листинг программы