# Слайд 1

Добрый день, уважаемая Государственная аттестационная комиссия! Вашему вниманию представляется выпускная квалификационная работа на тему "Система распознавания жестов на видеопоследовательности для реализации интерфейсов человеко-машинного взаимодействия». Работу выполнил студент группы ИУ1-81Б Юдаков Дмитрий.

# Слайд 2

В последнее время концепция взаимодействия между людьми и роботизированными машинами вызвала множество научных интересов. Разработка каналов связи и создание структурированных способов общения даст возможность сформировать человеко-машинную команду, компенсирующую недостатки каждого субъекта по отдельности, что позволит сохранить рабочие места, освободить людей от физически тяжелых задач и повысить общую эффективность. Одним из актуальных и развивающихся способов человеко-машинного взаимодействия является взаимодействие на основе жестов.

# Слайд 3

Общая модель человеко-машинного взаимодействия на основе распознавания жестов, исследованная на основе упрощённой модели обработки информации, содержит пять основных этапов – это сбор данных с датчиков, идентификация жеста, отслеживание жеста, классификация жестов и отображение жеста. Каждый этап может быть реализован с помощью нескольких технологий, имеющих свои недостатки и преимущества.

# Слайд 4

Целью работы является разработка и реализация системы детектирования кисти, идентификации и классификации жеста. В соответствии с поставленной целью необходимо решить следующие задачи: сделать обзор существующих подходов и средств для распознавания жестов, провести их сравнение, выделить особенности, **||** подробно описать методы детектирования кисти человека и методы последующей идентификации жеста, **||** разработать и реализовать систему распознавания жестов, **||** провести её анализ, выявить недостатки, описать пути их устранения.

# Слайд 5

Одним из самых активно развивающихся направлений в настоящее время является машинное обучение и искусственные нейронные сети. Нейросетям уже нашли множество применений в самых разных областях науки и техники, не исключение и использование нейросетей в распознавании жестов. Нейросети имеют множество архитектур, одними из которых являются свёрточная нейронная сеть и рекуррентная нейронная сеть с её улучшением -- LSTM, позволяющие эффективно работать с изображениями и с последовательностями соответственно.

# Слайд 6

Одним из подходов в задаче определения кисти на изображении и положения ключевых точек является подход MediaPipe Hands. Для его реализации используется нейросетевой Single Shot детектор, обладающий хорошим соотношением скорости и точности работы. Поскольку обнаружение всей кисти вместе с пальцами является достаточно сложной задачей из-за множества количества её конфигураций, то для определения начального положения рук была разработана модель SSD, оптимизированная для использования в реальном времени за счёт обучения обнаружения лишь ладоней. Так же этот подход позволяет определять 21 ключевую точку на кисти руки с помощью регрессии. С учётом всех оптимизаций точность обнаружения ладони составляет 95,7 процента.

# Слайд 7

Для разработки интерфейса человеко-машинного взаимодействия необходимо привести схему системы распознавания жестов. Итак, изображение, считанное одиночной камерой, поступает на вход детектора, определяющего существование кисти на изображении и её положение. Затем, на выделенном изображении кисти производится поиск ключевых точек с помощью нейросети регрессии. При удачном определении расположения всех ключевых точек, их массив сохраняется в буфер и одновременно с этим поступает на вход нейросети классификации статических жестов. После наполнения буфера массивы точек поступают на вход нейросети классификации динамических жестов. Фильтр, присутствующий на выходах нейросетей, пропускает лишь кодирующие жесты сигналы, отсекая «шум».

# Слайд 8

Для инициализации набора с координатами точек жестов было разработано GUI приложение, позволяющее эффективно собирать данные и в то же время производить мониторинг для выравнивания их объема по классам. Обозначив три класса статических жестов с одним, двумя и трёмя пальцами и три класса динамических жестов, был собран набор данных, состоящий из 3000 записей по статическим и 500 записей по динамическим жестам соответственно.

# Слайд 9

Подготовка данных для обучения модели заключается в преобразовании координат к нормализованному виду. Поскольку изначальные координаты точек абсолютны, то их необходимо отвязать от расположения на изображении, придав им относительный характер. Для этого все точки кисти вычитаются из точки 0, находящейся на запястье, и нормируются.

После этого набор массивов координат подаётся на вход многослойного перцептрона со структурой, изображённой на слайде. Все скрытые слои, кроме последнего имеют функции активации ReLU, последний -- Softmax.

# Слайд 10

В качестве инструментов для разработки моделей нейросетей и их обучения были выбраны язык программирования Python версии 3.9, обладающий большим объёмом функций машинного обучения, и библиотека TensorFlow второй версии, позволяющая быстрее и качественнее других производить обучение и его мониторинг.

Процесс обучения нейронной сети и основные результаты представлены на слайде, за 97 эпох были получены значения точности и потерь 0,97 и 0,14 соответственно.

# Слайд 11

Поскольку для реального взаимодействия человека и машины трёх типов жестов может быть недостаточно, то было решено расширить модель бОльшим количеством классов. Для расширения модели была взят набор данных HANDS, содержащий почти 16 тыс. изображений с 13 жестами, выполненными левой и правой рукой, и нейронная сеть с улучшенной архитектурой, содержащей бОльшее количество скрытых слоёв и нейронов в них.

# Слайд 12

Процесс обучения расширенной модели представлен на слайде, за 78 эпох точность и потери на тестовой выборке составили 0,99 и 0,03 соответственно. На основе показателей метрик были выделены две пары жестов, определяемых ключевыми точками одинаково, но на деле являющихся разными. Это говорит о том, что в работе модели жесты в каждой из пар не могут использоваться вместе.

# Слайд 13

Для подготовки к обучению модели распознавания динамических жестов каждую запись требуется привести к одному общему размеру, поскольку динамический жест является последовательностью координат различной длины. В данном случае каждая запись дополняется нулями справа до максимального размера среди всех записей.

# Слайд 14

В архитектуре нейросети распознавания динамических жестов используется слой LSTM, поскольку динамический жест является последовательностью. После окончания процесса обучения, представленного на слайде, за 100 эпох общая точность и потери на тестовой выборке составили 0,99 и 0,02 соответственно.

# Слайд 15

В качестве практического применения интерфейса и тестирования системы распознавания жестов была разработана система управления курсором мыши. В этой системе курсор мыши привязан к расположению точки на указательном пальце, а жесты 2 и 3 являются левой и правой кнопкой мыши соответственно. В результате тестирования были выявлены ряд недостатков и предложены пути их решения.

# Слайд 16

В результате работы была разработана система детектирования и идентификации кисти и классификации динамических и статических жестов с использованием искусственных нейронных сетей. Был сделан обзор существующих подходов распознавания жестов, проведено их сравнение, выделены особенности. Подробно описаны методы детектирования кисти человека и методы последующей идентификации жеста. Разработана и реализована система распознавания жестов. Проведён анализ разработанной системы, выявлены недостатки и описаны пути их устранения.

# Слайд 17

Доклад окончен, спасибо за внимание!

ВОПРОСЫ

**Система** – комбинация взаимодействующих элементов, организованных для достижения одной или нескольких поставленных целей.

**Интерфейс** – граница между двумя функциональными объектами, требования к которой определяются стандартом; совокупность средств, методов и правил взаимодействия (управления, контроля и т. д.) между элементами системы.

**Исключение** или **дропаут** — метод [регуляризации](https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%B3%D1%83%D0%BB%D1%8F%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_(%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0)) [искусственных нейронных сетей](https://ru.wikipedia.org/wiki/%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C), предназначен для уменьшения переобучения сети за счет предотвращения сложных коадаптаций отдельных нейронов на тренировочных данных во время обучения. Характеризует исключение определённого процента случайных нейронов (находящихся как в скрытых, так и видимых слоях) на разных итерациях (эпохах) во время обучения нейронной сети.

**ReLU** – линейна в первом квадранте, во втором.

**Softmax** – обобщение логистической функции для многомерного случая

**Выборки**: Обучающая – для настройки и оптимизации параметров алгоритма, Тестовая (контрольная) -- для оценки качества модели, Проверочная (валидационная) – для выбора наилучшей модели машинного обучения.

**Пояснения для этапов модели процесса:** Это ***сбор данных с датчиков***, когда данные жеста фиксируются датчиками, ***идентификация жеста***, когда в каждом кадре жест локализуется, ***отслеживание жеста***, когда локализованный жест отслеживается во время движения, ***классификация жестов*,** когдаотслеживаемый жест классифицируется в соответствии с заранее определёнными типами и ***отображение жеста***, когда результат распознавания жестов преобразуется в управляющие команды.

**Точность** системы в пределах класса – это доля объектов, действительно принадлежащих данному классу относительно всех объектов, которые система отнесла к этому классу. **Полнота** системы – это доля найденных классификатором объектов, принадлежащих классу относительно всех объектов этого класса в тестовой выборке.