

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**ОПРЕДЕЛЕНИЕ ЛИЦА В ВИДЕО ПОТОКЕ И УПРАВЛЕНИЕ
ПРИЛОЖЕНИЯМИ С ПОМОЩЬЮ МИМИКИ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 411 группы

направления 02.03.02 — Фундаментальная информатика и информационные
технологии

факультета КНиИТ

Минаева Ивана Вячеславовича

Научный руководитель

д. ф.-м. н., профессор

В. А. Романов

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2021

ВВЕДЕНИЕ

Идея создания искусственного интеллекта начала покорять людей еще с древних времен, так например в Греческих мифах прослеживается идея создания робота-андроида, который мог выполнять звучащие современно задачи. К сожалению, задача создания и обучения искусственного интеллекта оказалась непосильной для наших предков и серьезно начала развиваться с момента технического прорыва XX века. История по развитию технологий машинного обучения начинается, примерно, с 1950-х годов, когда были разработаны первые программы для игры в шашки. Прошло достаточно много времени, но основной принцип не изменился. Однако, в связи с ростом вычислительных мощностей современных компьютеров и ростом потребностей человека в развитии технологического прогресса, усложнились и алгоритмы построения нейронных сетей и алгоритмы их обучения. Современные нейронные сети в состоянии обрабатывать огромный поток информации, автоматизировать сложные для человека действия классифицировать объекты и т.д.

Одним из ответвлений нейронных сетей является компьютерное зрение. Работа в этом направлении позволяет решать множество задач, таких как: Помощь в управлении компьютером людям с ограниченными возможностями; Помощь государствам в обеспечении безопасности с помощью мониторинга улиц и распознавания лиц; Обеспечение личной безопасности, с помощью уменьшения физического контакта с аппаратными средствами управления компьютером и другие.

2020 и 2021 года, из-за пандемии вируса, дали большой толчок развитию нейронных сетей и их обучения. Люди ищут все больше путей для минимизации физического контакта с общедоступными поверхностями, такими как, например, терминалы, банкоматы и другие. На помощь в таких ситуациях может прийти управление компьютерами с помощью мимики лица и жестов.

Попытки управления интерфейсом компьютера предпринимались множество раз, однако в большинстве случаев они упирались в различные ограничения, такие как, например, необходимость иметь специальные устройства для фиксации направления взгляда. Но современные алгоритмы машинного обучения позволяют решить данную проблему с помощью простой web-камеры, имеющейся в доме практически у каждого.

Целью данной работы является изучение нейронных сетей и алгоритмов

машинного обучения, а в частности — создание web-приложения, которое может распознавать лицо в видео потоке, что позволит управлять страницей браузера с помощью мимики лица и без помощи рук, а также интеграция данного web-приложения в другие приложения с целью демонстрации возможностей его применения в различных сферах деятельности человека. Актуальность данного приложения подтверждается в первую очередь его доступностью и простотой использования. От пользователя не требуется никаких дополнительных настроек и использования сторонних приложений и устройств, кроме web-камеры.

В данной работе по указанной теме представлены 3 главы:

- Глава 1 — глава с теоретической частью работы. В данной главе рассказаны основные принципы работы нейронных сетей, история нейронных сетей, способы машинного обучения, применение нейронных сетей в распознавании лиц.
- Глава 2 — 1-ая практическая глава работы. В данной главе приводится последовательное описание создания приложения для распознавания мимики лица и управления страницей браузера без помощи рук.
- Глава 3 — 2-ая практическая глава работы. В данной главе рассказывается о внедрении созданного приложения в другие приложения для управления ими.

1 Основные теоретические выкладки

1.1 Типы задач машинного обучения

Все задачи, решаемые с помощью машинного обучения относятся к одному из следующих типов задач [1]:

1. Задача регрессии — прогноз на основе выборки объектов с различными признаками.
2. Задача кластеризации — разделение заданных объектов по группам по некоторым признакам.
3. Задача классификации — получение категориального ответа на основе набора признаков.
4. Задача уменьшения размерности — уменьшение количества признаков для удобства их последующей визуализации.
5. Задача выявления аномалий — отделение аномалий от стандартных случаев.
6. Задача распознавания объектов на изображении и их классификация.

1.2 Биологические нейронные сети

За основу искусственных нейронных сетей была принята идея биологических нейросетей человеческого мозга. Человеческий мозг состоит из огромного количества нейронов. Каждый нейрон имеет один длинный отросток — аксон и множество коротких отростков — дендритов, который связываются с аксонами других нейронов, тем самым образуя сложные связи — синапсы. В среднем, человеческий мозг содержит порядка ста миллиардов нейронов, каждый из которых обладает примерно 7000 связей.

Изучение строения и возможностей биологических нейронных сетей послужило сильным толчком для создания искусственных нейросетей. Так основные принципы строения и передачи сигналов были переняты с работы человеческого мозга.

1.3 Устройство искусственной нейронной сети

Компьютерные нейронные сети представляют собой математическую модели реальных нейросетей живых существ. Для того чтобы перевести такую модель в программу, в языке программирования обязательно должны быть математические модули.

Нейронная сеть — это последовательность, связанных между собой слоев нейронов, получающих и передающих информацию. По отдельности нейроны не играют важной роли, они важны, когда из них выстраивается цепь.

К нейрону поступают сигналы, каждый из которых имеет свой вес. Сигнал умножается на свой вес и результаты складываются, получается число, которое передается активационной функции и уже она решает, передавать ли сигнал дальше.

1.4 Многослойные нейронные сети

Так как нейроны поодиночке не способствуют решению сложных задач, на практике применяются многослойные нейронные сети. Нейроны таких сетей связаны своими входами с нейронами предыдущего слоя, а выходами — с нейронами следующего слоя. Активационными функциями для таких нейронов служат линейные, сигмоидные и пороговые функции. Обучаемыми параметрами нейронной сети являются веса признаков нейронов, а статическими гиперпараметрами являются количество слоев, количество нейронов в слоях, используемые функции активации и т.д.

1.5 Глубокие нейронные сети

В последнее время стала популярна теория глубоких нейронных сетей. Идея создания глубоких нейронных сетей заключается в рассмотрении нейросети как граф вычислений с наложением слоев нейронов друг на друга и дифференцировании композиции функций.

Основным отличием глубоких нейронных сетей является то, что они обучаются выделяя основные признаки самостоятельно, перебирая их и представляя их с различным приоритетом для более точного решения задачи. Такой подход, в котором нейронная сеть сама выбирает основные признаки, является гораздо более ресурсозатратным.

1.6 Обучение нейронных сетей

Сами по себе нейроны с ненастроенными весами не приносят никакой пользы в решении конкретных задач, поэтому для того, чтобы нейросеть могла прогнозировать ответ на заданный вопрос, ее необходимо для начала обучить. Машинное обучение может проходить с учителем или без него.

В реальной жизни эти два подхода используются для разных задач, так, например, нейросеть хорошо обученная с учителем может распознавать образы, сортировать данные по заданным признакам, играть в игры. Но цели поставленные перед такими нейросетями ограничены самими людьми. Обучение без учителя — это парадигма обучения, в которой испытываемая нейросеть "награждается" за наивное изучение зависимостей между данными без каких-то конкретных целей, что помогает открывать что-то новое.

Самыми известными и популярными являются алгоритмы:

1. Дерево принятия решений.
2. Наивная байесовская классификация.
3. Метод наименьших квадратов.
4. Логистическая регрессия.
5. Метод опорных векторов.
6. Метод ансамблей.
7. Алгоритмы кластеризации.
8. Анализ главных компонент.
9. Анализ независимых компонент.

1.7 Оптимизация машинного обучения

Одним из важнейших аспектов машинного обучения является его вероятностная интерпретация. Машинное обучение — это наука о том, как на основе входных данных делать выводы и предсказания. При этом сделать безошибочные выводы невозможно, так как процессы порождения данных слишком сложны даже в простейших задачах. Поэтому задача машинного обучения — свести к минимуму шанс ошибки. Если рассматривать желаемые результаты как функции, то задача машинного обучения — максимизировать функцию правдоподобия, которая показывает, насколько полученное решение соответствует поставленной задаче. Функция оценки максимального правдоподобия для параметра θ строится по выборке элементов $X = (x_1, \dots, x_n)$:

$$\Lambda_{\theta}(X) = \prod_{i=1}^n f_{\theta}(x_i) \rightarrow \max$$

После того как мы получили реальную выборку элементов и знаем функцию максимального правдоподобия, мы сможем высчитать функцию ошибки как разницу между максимальным правдоподобием и реальным результатом.

Задача машинного обучения заключается в минимизации функции ошибки или другими словами аппроксимации функции, полученной по реальной выборке к функции максимального правдоподобия.

1.8 Распознавание объектов нейронной сетью

Теория распознавания образов — раздел кибернетики, развивающий методы классификации объектов по конечному числу признаков.

Нейронные сети обращаются с изображениями как с наборами пикселей и главной сложностью для распознавания объектов является задача распознавание цельных объектов в наборе пикселей. Для обработки изображения используются разные алгоритмы, но большинство из них нацелены на нахождение особых точек — точек, которые отделяют объекты друг от друга и являются уникальными и последующего выделения предполагаемых объектов. Для нахождения особых точек используются детекторы, а для того чтобы однозначно описать особые точки используются дескрипторы.

Одними из популярных и эффективных алгоритмов для нахождения особых точек являются:

- Детектор углов Харриса.
- Алгоритм FAST

Однако детектирование особых точек является инструментом для нахождения объектов на изображении. Одной из главных структур для детекции объектов на изображении стала структура R-CNN, разработанная командой из UC Berkley для применения сверточных нейронных сетей.

Алгоритм R-CNN:

1. Выделение регионов-кандидатов с помощью алгоритмов выделения особых точек.
2. Преобразование региона в размер, принимаемый сверточной нейронной сетью CaffeNet.
3. Получение при помощи сверточной нейронной сети 4096-размерного вектора признаков.
4. Проведение N бинарных классификаций каждого вектора признаков при помощи N линейных SVM.
5. Линейная регрессия параметров рамки региона для более точного охвата объекта.

На данный момент существует множество улучшений алгоритма R-CNN,

так как этот алгоритм является одним из первых и имеет слабую производительность и избыточные вычисления.

1.9 Распознавание лиц

Ученые решают задачи по определению лиц уже больше 40 лет. В них входят:

- Поиск и лиц в видеопотоке или на фотографиях и их дальнейшее распознавание.
- Масштабируемость данных для идентификации личности.
- Корректная работа программ для распознавания лиц при изменении особенностей лица.
- Работа в реальном времени.
- Распознавание эмоций

Для начала системе распознавания лиц необходимо определить лицо на видео или фотографии и выделить данную область. Программное средство может выполнять эту задачу с помощью разных алгоритмов. Например, определение пропорций и цвета кожи, выделение контуров на изображении и сопоставление с контурами лиц, выделение изображения с помощью нейросетей [2].

Когда программа определяет лицо в видеопотоке, она находит контрольные точки — опорные элементы, с помощью которых в дальнейшем и будет отслеживаться мимика лица.

Одним из наиболее эффективных методов для распознавания лиц является метод Виолы-Джонса, который может использоваться в режиме реального времени. Метод использует признаки Хаара, представляющие собой набор черно-белых масок.

1.10 Возможности по внедрению распознавания лиц

Отраслей, в которые можно ввести технологии по распознаванию лиц, действительно много. Такие технологии могут быть полезны не только правоохранительным органам для отслеживания преступников, как считают многие. Например, с помощью технологий по распознаванию лиц, ритейлеры могут анализировать аудиторию магазинов: пол, возраст, количество повторных визитов. Разработчики смартфонов уже давно вводят такие технологии в свои

устройства, например, для проведения бесконтактной оплаты с помощью телефона, разблокировки телефона и многих других функций.

2 Основные выкладки практической части

2.1 Выбор средств для реализации задачи

В ходе анализа было принято решение использовать библиотеку `clmtrackr`, написанную на языке JavaScript [3]. Данная библиотека предоставляет доступ к обученной нейросети, которая, используя web-камеру компьютера определяет лицо человека в видео потоке или на статичной фотографии и выстраивает так называемые контрольные точки, привязанные к изображению лица. Нейросеть, предоставляемая библиотекой обучена на датасете из 3500 фотографий людей разного возраста, пола, национальности. Хотя библиотека и использует данные с web-камеры компьютера, с ее помощью мы можем получить относительно хорошую точность в определении лица на видео.

Помимо основного функционала по отслеживанию лица, разработчики данной библиотеки также предоставляют специализированную утилиту для переобучения нейросети, что может помочь в улучшении точности ее работы. Также библиотека предоставляет дополнительные возможности, такие как: Наложение маски на конечное изображение поверх распознанного лица; Замена лица на конечном изображении; Распознавание эмоций; Деформация лица на конечном изображении. Но эти дополнительные функции в работе применены не были, так как их функционал вытекает из основной функции - распознавания лица и просто демонстрирует возможности библиотеки.

2.2 Создание приложения для управления страницей браузера без помощи рук

В ходе работы были созданы несколько HTML-страниц, которыми будем управлять пользователь без помощи рук. В HTML-файле подключается JS скрипт с библиотекой `clmtrackr`, наше будущее приложение будет использовать данную библиотеку [4]. Для того чтобы использовать изображение пользователя с камеры компьютера, необходимо запросить доступ у пользователя. После того как получен доступ, в правом верхнем углу страницы выводится изображение с видеокamеры и отрисованная маска зеленого цвета, которая показывает как лицо «видит» нейросеть библиотеки `clmtrackr`. Данная библиотека предоставляет разработчику массив контрольных точек, привязанных к определенным частям лица, с использованием этих контрольных точек и будет осуществлено управление страницей.

В качестве основных управляющих функций реализуем такие функции как:

1. Управление курсором — так как курсор на странице нельзя перемещать программно, создадим объект — имитацию курсора. Его перемещение привязано к перемещений контрольной точки, соответствующей кончику носа.
2. Скроллинг страницы — осуществляется наклоном головы. Наклон вправо — скроллинг вниз, наклон влево — скроллинг вверх.
3. Нажатие на кнопку, на которую указывает курсор — при наведении курсора на кнопку, вызов события кнопки осуществляется приподнятием бровей.

2.3 Интеграция управления без помощи рук в другие приложения

В качестве примера приложений для интеграции, мной были выбраны 2 игры, так как игровая форма проста в понимании и работу приложения можно проверить наглядно.

2.3.1 Интеграция в 2-D игру

Первой игрой для интеграции управления без рук была выбрана 2-D игра — «Flappy Bird». Исходники данной игры были взяты мной из интернета и переработаны под необходимые задачи. Мной было принято решение немного отойти от начальной концепции нажатия на клавишу для перемещения птицы по оси X, хотя реализация этого метода также возможна. Вместо этого, птица будет повторять движения головы игрока вверх или вниз. Суть данной игры останется той же — игрок должен проводить птицу между преград, набирая очки, но делая это не с помощью нажатия на клавишу, а с помощью движений головы.

2.3.2 Интеграция в 3-D игру

Второй игрой стала 3-D игра, в которой игроку необходимо перемещать персонажа в трехмерном пространстве и проходить уровни, собирая золотые монеты. С помощью мыши игрок направляет своего персонажа в пространстве, а с помощью клавиатуры перемещает его.

Так как вопрос перемещения персонажа в пространстве оказался более простым, то в первую очередь я начал заниматься решением этого вопроса.

После анализа кода игры было выявлено, что нажатие на клавиши клавиатуры изменяет скорость героя в разных направлениях, поэтому мне было необходимо придумать другой вариант изменения параметров скорости героя. За основу для перемещения было взято движения лица игрока — приподнятие бровей. Когда игрок приподнимает брови — персонаж начинает двигаться вперед, а когда опускает брови — персонаж двигается назад.

Для поворота героя в пространстве использовалась компьютерная мышь, поэтому необходимо было придумать замену, которая могла бы также вращать героя в пространстве. За основу для этого действия было взято движение игрока — поворот головы. Таким образом игрок поворачивает голову в одну из сторон и персонаж в игре начинает поворачиваться в ту же сторону.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены основы работы нейронных сетей, их обучения, оптимизации, а также возможности их применения для решения практических задач. Также в практической части работы разработано web-приложение для управления страницами браузера с помощью мимики и движений лица. А также были показаны примеры внедрения данного приложения в другие приложения.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Галушкин, А.* Нейронные сети: основы теории / А. Галушкин. — Москва: Горячая линия - Телеком, 2012. — С. 412.
- 2 *Брилюк, Д.* Распознавание человека по изображению лица нейросетевыми методами / Д. Брилюк. — Москва: Препринт / Ин-т техн. кибернетики НАН Беларуси, 2002. — С. 54.
- 3 *Mayers, M.* A Smarter Way to Learn JavaScript: The new approach that uses technology to cut your effort in half / M. Mayers. — California: CreateSpace Independent Publishing Platform, 2014.
- 4 *Atencio, L.* Functional Programming in JavaScript / L. Atencio. — Shelter Island: Manning, 2016.