Смена 2021
Digital Смена

Smena 2021
Digital Smena

Команда "ВЕМА"

VEMA Project

Создание игры с мини квестами, для решения которых задействуются новые интерактивные механики с отслеживанием жестов рук, мимики лица, произношением команд.

Спицын Павел, Осипов Михаил, Цыканова Ксения, Суханов Артём, Криштопа Денис

Канал связи: p@glav.tatar

Механики

1) Hand Tracking

В настоящее время всё больше исследований направлено на решение задач с применением компьютерного зрения и искусственного интеллекта. Наиболее частыми являются решения и подходы с использованием распознавания жестов на основании инфракрасных сенсоров или нейронных сетей.

Актуальность рассматриваемой тематики обусловлена возможностью применения предлагаемого подхода для управления работой объектов без тактильного контакта и голосовой идентификации команд, а также своей простотой с точки зрения конечного пользователя.

В настоящей работе проанализированы существующие способы распознавания жестов. Рассмотрены методы и подходы, а также их реализация, исследованы преимущества и недостатки рассмотренных методов. На их основе составлена таблица с тезисной информацией и предложена собственная архитектура сверточной нейронной сети для решения классификации жестов. Проведена оценка точности работы сети. На основе полученных данных проведен двухфакторный анализ зависимости сложности жеста, его дальности и точности полученного алгоритма.

По полученной зависимости построены графики изменения точности работы сверточной нейронной сети. Проанализирован характер изменения точности для различных факторов.

Ключевые слова: распознавание жестов, компьютерное зрение, сверточные нейронные сети, обучение, support vector machine, классификация, Keras, Tensorflow,

ВВЕДЕНИЕ

Распознавание жестов играет важную роль во взаимодействии человека с машиной из-за его естественного и дружественного семантического выражения. Для использования этой технологии машины должны быстро и точно их определять, чтобы пользователи чувствовали себя комфортно и были готовы взаимодействовать с ЭВМ. Распознавание жестов остается слож-

ной задачей из-за их разнообразия, сходства форм и сложности сценариев применения.

РАСПОЗНАВАНИЕ ЖЕСТОВ В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ СЕТИ ПЕРЕКАЛИБРОВКИ ФУНКЦИЙ С МНОГОМАСШТАБНОЙ ИНФОРМАЦИЕЙ

Использование сверточной нейронной сети для решения задачи распознавания и классификации жестов.

Существует две проблемы в процессе распознавания жеста, расположенного на большом расстоянии от камеры. Во-первых, жесты с различным соотношением размеров трудно идентифицировать; во-вторых, существует разная информация между признаками низкого уровня и признаками высокого уровня. Несмотря на то что низкоуровневые признаки с высоким разрешением содержат больше деталей и информацию о позиции в кадре, это не способствует выявлению нужных признаков разных размеров при обнаружении. И, напротив, высокоуровневые признаки больше подходят для классификации по категориям, но имеют более низкое восприятие деталей из-за их более низкого разрешения.

В некоторых сетях, таких как SegNet и Unet, точность и надежность сети улучшаются путем объединения нескольких наборов низкоуровневых признаков и остальных признаков, но эти методы неэффективны для практического использования.

В основе данной работы лежит объединение признаков в разных масштабах. Чтобы лучше извлечь контекстную информацию разных масштабов извлекается информация о характеристиках разных масштабов при помощи сверточного ядра с размером шага 2 и размером 3 × 3 и 5 × 5 соответственно.

Полная структура сверточной нейронной сети изображена на рис. 1.

Модуль FPA используется для объединения различных масштабов контекстной информации для объединения локальной и глобальной информации. Кроме того, применение ядра свертки эффективно уменьшает количество параметров в тренировочном процессе и повышает скорость

распознавания жестов. Для обучения сети было собрано 9289 изображений, содержащих 26 жестов. Затем был использован метод изменения насыщенности оттенка и вращения изображений, чтобы увеличить размер набора данных, тем самым увеличив устойчивость нейронной сети. По сравнению с популярными в настоящее время сверточными сетями метод может распознавать жесты в режиме реального времени, обеспечивая при этом высокую точность.

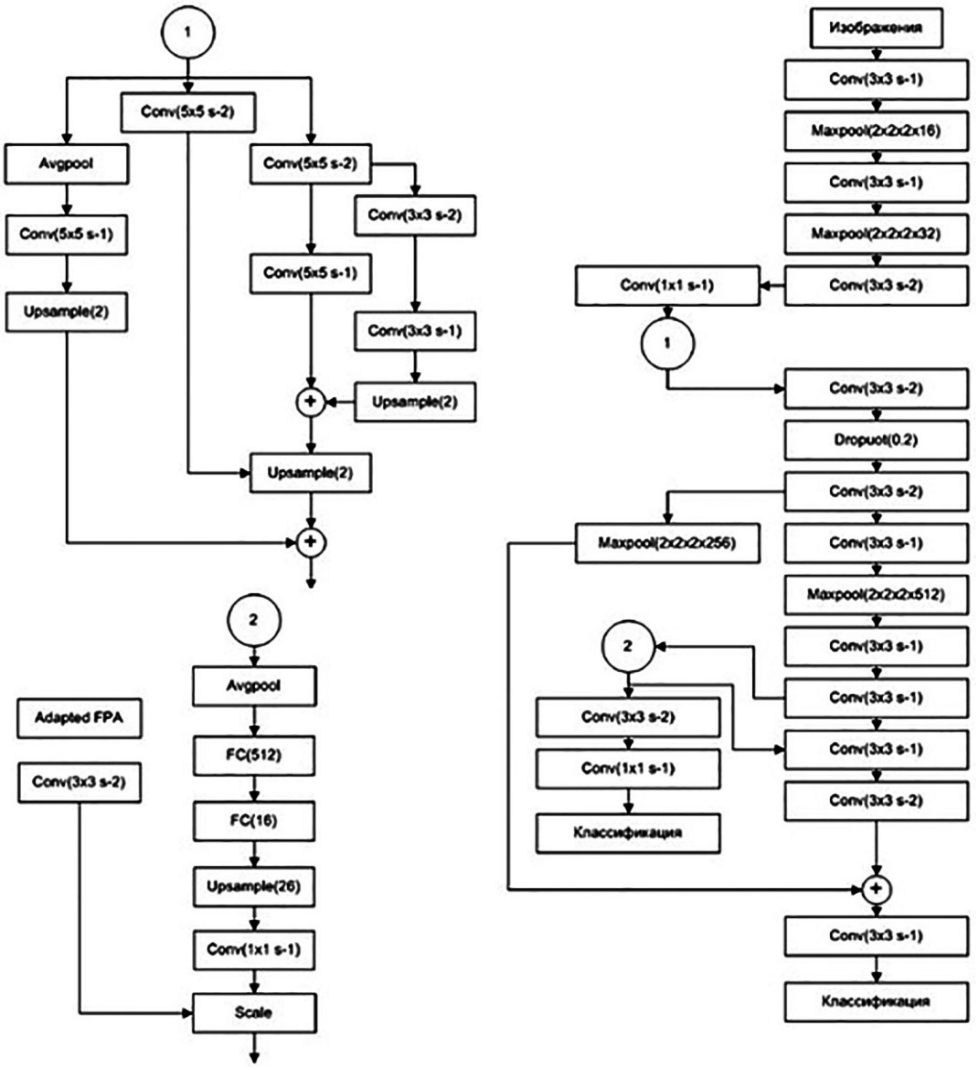


Рис. 1. Общая структура нейронной сети

Fig. 1. The gesture recognition network structure diagram

РАСПОЗНАВАНИЕ ЖЕСТОВ ПРИ ПОМОЩИ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ С НЕБОЛЬШОЙ АРХИТЕКТУРОЙ

Для увеличения быстродействия предложен алгоритм, в основе которого находится сверточная нейронная сеть с небольшой архитектурой.

Основой сверточной нейронной сети являются слои свертки. Каждыйслой включает в себя фильтры для каждого канала. Они обрабатывают предыдущий слой по частям (путем суммирования матричных фрагментов). Все веса ядра свертки заранее неизвестны и изменяются в процессе обучения в зависимости от входных данных. В конце слоя свертки всегда стоит функ- ция активации.

Для передачи информации с одного слоя на другой необходима функция активации. Она преобразует информацию (численные значения) со всех нейронов предыдущего слоя в определенное значение для нейрона текущего слоя. Выход зависит от функции активации и может быть как действительным, так и целым. Значение выхода – это показатель того, насколько активировался нейрон текущего слоя.

Так как каждому фильтру свертки соответствует одна карта признаков, то это позволяет нейронной сети научиться выделять признаки независимо от их расположения во входном изображении.

Пуллинг можно сформулировать так: если на предыдущей операции свертки были обнаружены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уменьшается в размерности, т. е. уплотняется до менее подробного. К тому же фильтрация уже ненужных деталей уменьшает переобучение. Слой пуллинга, как правило, вставляется после слоя свертки перед слоем следующей свертки.

За счет слоя пуллинга сеть становится наиболее устойчивой к изменениям входного изображения, например, к его сдвигам. Также уменьшается размерность последующих слоев.

Полносвязный слой (многослойный перцептрон) – скрытый слой, соединенный со всеми нейронами предыдущего слоя. Последним слоем многослойного перцептрона является один или несколько нейронов, количество которых равно количеству классов. Проще говоря, на вход всей сверточной нейронной сети подается изображение, а на выходе сеть выдает класс, к которому это изображение относится.

Сверточные нейронные сети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям. Общая топология изображена на рис. 2.

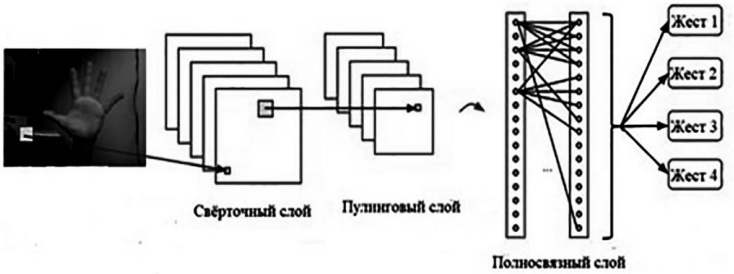


Рис. 2. Общая структура нейронной сети

Fig. 2. The gesture recognition network architecture

В результате была спроектирована сверточная нейронная сеть с параметрами, указанными в табл. 1.

Таблица 1

Table 1

Результаты проектирования нейронной сети

Results of the neural network design

Параметр		Характеристика
Количество сверточных слоев и слоев пуллинга		3
Размер фильтров	Первый слой	5×5
	Второй слой	3×3
Количество фильтров для каждого слоя		8, 12, 16
Вероятность Drop-out слоя		50 %
Количество нейронов в полносвязном слое		128

Данная сверточная нейронная сеть обучена на выборке из 12 000 изображений, соответствующих двадцати шести классам. Количество эпох обучения равно трем. При большем количестве эпох сеть начинает «выучивать» данные с изображений и становится не способна работать с окружением, отличным от того, которое присутствует в обучающей выборке, т. е. происходит переобучение.

Общая точность обучения на тестовой выборке, полученная при помощи функций библиотеки для машинного обучения Keras, примерно равна 95 %. Результаты успешных определений жестов приведены на рис. 3.

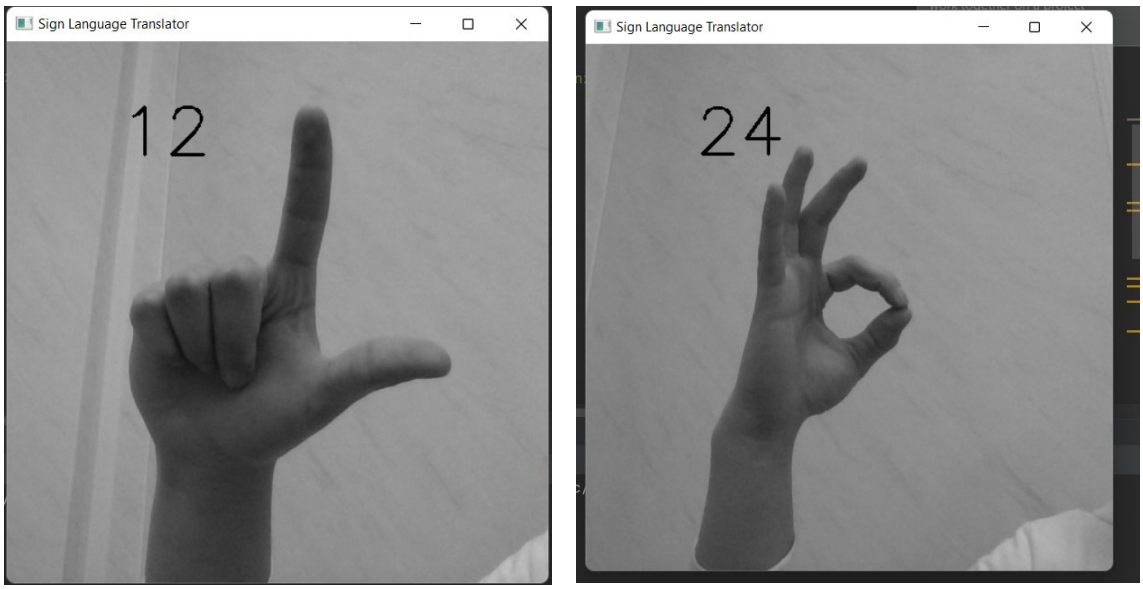


Рис. 3. Успешное обнаружение жестов нейронной сетью

Fig. 3. Successful gesture detection by neural network

ЗАКЛЮЧЕНИЕ

Современные методы решения распознавания жестов имеют ряд недостатков, таких как низкая скорость распознавания, низкая скорость и низкая производительность при распознавании нескольких целей или целей на большом расстоянии в сложных условиях. Ввиду вышеупомянутых проблем был предложен подход распознавания жестов сверточной нейронной сетью с меньшей архитектурой.

Предложенная сверточная нейронная сеть имеет наименьшее время обучения, скорость работы и требует меньших затрат вычислительных мощностей.

В результате была разработана нейронная сеть, состоящая из трех сверточных слоев, распознающая двадцать шесть жестов или же их отсутствие.

2) Speech recognition

Речевое общение является естественным и удобным для человека. Задача распознавания речи состоит в том, чтобы убрать посредника в общении человека и компьютера. Управление машиной голосом в реальном времени, а также ввод информации посредством человеческой речи намного упростит жизнь современного человека. Научить машину понимать без посредника тот язык, на котором говорят между собой люди – задачи распознавания речи.

ВВЕДЕНИЕ

Системы распознавания голоса – это вычислительные системы, которые могут определять речь говорящего из общего потока. Эта технология связана с технологией распознавания речи, которая преобразует произнесенные слова в цифровые текстовые сигналы, путем проведения процесса распознавания речи машинами. Обе эти технологии используются параллельно: с одной стороны для идентификации голоса конкретного пользователя с другой стороны для идентификация голосовых команд посредством распознавания речи. Распознавание голоса используется в биометрических целях безопасности, чтобы определить голос конкретного человека. Эта технология стала очень популярной в мобильном банкинге, который требует идентификации подлинности пользователей, а также для других голосовых команд, чтобы помочь им совершать сделки.

КЛАССИФИКАЦИЯ СИСТЕМ РАСПОЗНОВАНИЯ РЕЧИ

Системы распознавания речи классифицируются:

- по размеру словаря (ограниченный набор слов, словарь большого размера);
- по зависимости от диктора (дикторозависимые и дикторонезависимые системы);
- по типу речи (слитная или раздельная речь);
- по назначению (системы диктовки, командные системы);
- по используемому алгоритму (нейронные сети, скрытые Марковские модели, динамическое программирование);
- по типу структурной единицы (фразы, слова, фонемы, дифоны, аллофоны);
- по принципу выделения структурных единиц (распознавание по шаблону, выделение лексических элементов).

Для систем автоматического распознавания речи, помехозащищённость обеспечивается, прежде всего, использованием двух механизмов:

- Использование нескольких, параллельно работающих, способов выделения одних и тех же элементов речевого сигнала на базе анализа акустического сигнала;
- Параллельное независимое использование сегментного (фонемного) и целостного восприятия слов в потоке речи.

АРХИТЕКТУРА СИСТЕМ РАСПОЗНОВАНИЯ РЕЧИ

Одна из архитектур систем автоматической обработки речи, основанной

на статистических данных, может быть следующей.

Модуль очистки шума и отделение полезного сигнала.

Акустическая модель — позволяет оценить распознавание речевого сегмента с точки зрения схожести на звуковом уровне. Для каждого звука изначально строится сложная статистическая модель, которая описывает произнесение этого звука в речи.

Языковая модель — позволяют определить наиболее вероятные последовательности слов. Сложность построения языковой модели во многом зависит от конкретного языка. Так, для английского языка, достаточно использовать статистические модели (так называемые N-граммы). Для высокофлективных языков (языков, в которых существует много форм одного и того же слова), к которым относится и русский, языковые модели, построенные только с использованием статистики, уже не дают такого эффекта — слишком много нужно данных, чтобы достоверно оценить статистические связи между словами. Поэтому применяют гибридные языковые модели, использующие правила русского языка, информацию о части речи и форме слова и классическую статистическую модель.

Декодер — программный компонент системы распознавания, который совмещает данные, получаемые в ходе распознавания от акустических и языковых моделей, и на основании их объединения, определяет наиболее вероятную последовательность слов, которая и является конечным результатом распознавания слитной речи.

ЭТАПЫ РАСПОЗНОВАНИЯ РЕЧИ

Обработка речи начинается с оценки качества речевого сигнала. На этом этапе определяется уровень помех и искажений.

Результат оценки поступает в модуль акустической адаптации, который управляет модулем расчета параметров речи, необходимых для распознавания.

В сигнале выделяются участки, содержащие речь, и происходит оценка параметров речи. Происходит выделение фонетических и просодических вероятностных характеристик для синтаксического, семантического и прагматического анализа. (Оценка информации о части речи, форме слова и статистические связи между словами.)

Далее параметры речи поступают в основной блок-системы распознавания — декодер. Это компонент, который сопоставляет входной речевой поток с информацией, хранящейся в акустических и языковых моделях, и определяет наиболее вероятную последовательность слов, которая и является конечным результатом распознавания.

Первый компонент распознавания речи – это речь. Речь должна быть преобразована из физического звука в электрический сигнал с помощью микрофона, а затем в цифровые данные с помощью аналого-цифрового преобразователя. После оцифровки можно использовать несколько моделей для преобразования звука в текст.

Большинство современных систем распознавания речи основаны на так называемой скрытой марковской модели (HMM). Этот подход работает на предположении, что речевой сигнал при просмотре в достаточно коротком временном масштабе (скажем, десять миллисекунд) может быть разумно аппроксимирован как стационарный процесс, то есть процесс, статистические свойства которого не меняются с течением времени.

В типичном HMM речевой сигнал делится на 10-миллисекундные фрагменты. Спектр мощности каждого фрагмента, который, по сути, представляет собой график зависимости мощности сигнала от частоты, отображается на вектор действительных чисел, известный как кепстральные

коэффициенты. Размерность этого вектора обычно мала - иногда всего 10, хотя более точные системы могут иметь размерность 32 или более. Конечным результатом HMM является последовательность этих векторов.

Чтобы преобразовать речь в текст, группы векторов сопоставляются с одной или несколькими фонемами - фундаментальной единицей речи. Этот расчет требует обучения, поскольку звук фонемы варьируется от говорящего к говорящему и даже меняется от одного высказывания к другому одним и тем же говорящим. Затем применяется специальный алгоритм для определения наиболее вероятного слова (или слов), которые образуют данную последовательность фонем.

Можно представить, что весь этот процесс может быть дорогостоящим в вычислительном отношении. Во многих современных системах распознавания речи нейронные сети используются для упрощения речевого сигнала с использованием методов преобразования признаков и уменьшения размерности перед распознаванием HMM. Детекторы голосовой активности (VAD) также используются для уменьшения звукового сигнала до тех частей, которые могут содержать речь. Это предотвращает трату времени распознавателем на анализ ненужных частей сигнала.

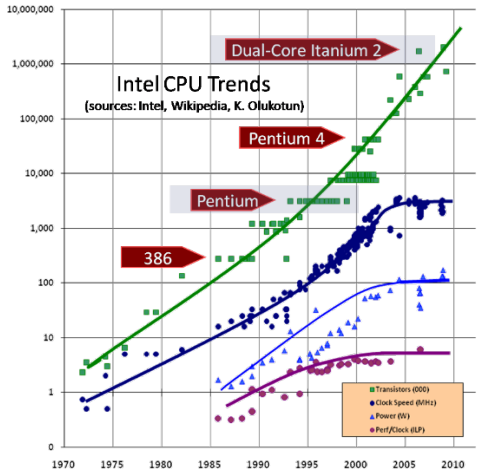
ПРОБЛЕМЫ СИСТЕМ РАСПОЗНОВАНИЯ РЕЧИ

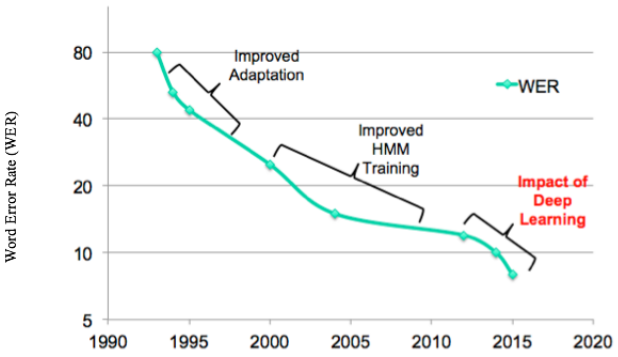
Положение дел в развитии технологии сегодня выражается целью: от распознавания речи к пониманию. Для этой цели выбран и ключевой показатель – процент ошибок в распознавании. Стоит сказать, что такой показатель применяется и в распознавании речи одного человека другим. Мы пропускаем часть слов, принимая во внимания другие факторы, например, контекст. Это позволяет нам понимать речь даже без понимания значений отдельных слов. Для человека показатель ошибки распознавания равен 5,1%.

Другими сложностями в обучении системы распознавания речи пониманию языка будут эмоции, неожиданная смена темы разговора, использование сленга и индивидуальные особенности говорящего: темп речи, тембр, произношение звуков.

РАЗВИТИЕ СИСТЕМ РАСПОЗНОВАНИЯ РЕЧИ

Идея распознавания речи выглядела многообещающе во все времена. Но уже на этапе распознавания чисел и самых простых слов исследователи столкнулись с проблемой. Суть распознавания сводилась к построению акустической модели, когда речь представлялась как статистическая модель, которая сравнивалась с готовыми шаблонами. Если модель соответствовала шаблону, то система принимала решение о том, что команда или число распознано. Рост словарей, которые могла распознать система, требовал увеличения мощностей вычислительных систем.





ЗАКЛЮЧЕНИЕ

На данный момент системы распознавания речи развиваются большими темпами, изобретаются и совершенствуются. В данных система до сих пор присутствуют некие проблемы.

Эта технология занимает достойное в IT индустрии. Сегодня есть множество общедоступных проектов для использования данных систем, среди них google speech API, Yandex.SpeechKit.

3) Emotion detection recognition (EDR)

Эмоциональные явления – давний предмет философских изысканий и не иссякающая тема психологических исследований. Значительная часть работ касается проблемы распознавания эмоциональных состояний, которая на настоящий момент выступает полем для дискуссий в обсуждении как фундаментальных вопросов психологии эмоций, так и методических аспектов организации эмпирических исследований. Отдельные вопросы, связанные с распознаванием эмоций, получают разработку в других областях психологического знания, в частности, в социальной, дифференциальной психологии, психологии развития. Ключевой вопрос настоящего исследования заключается в том, каким образом люди распознают эмоциональные состояния других людей и какие характеристики познавательной сферы определяют содержание и эффективность этих оценок.

В современной психологии широко исследуются проблемы распознавания эмоций в различных областях деятельности человека. В процессе общения и взаимодействия людей друг с другом очень значимо то, как мы воспринимаем эмоциональное состояние партнера. Информация о закономерностях распознавания эмоций взрослыми на лицах взрослых людей широко представлена в современной психологической науке.

Способность к распознаванию эмоций своих или другого лица относится к сфере «эмоционального интеллекта» и означает, что человек может установить факт наличия эмоционального переживания у себя или другого человека и найти ему словесное определение. Эта способность является важной составляющей невербальной (несловесной) коммуникации, которая осуществляется в процессе речевого общения параллельно с вербальной и составляет независимый информационный канал в системе общения.

Невербальная коммуникация - важнейшее, наряду со звуковой речью, и вместе с тем недостаточно изученное средство общения и взаимопонимания людей. Звуковая речь несет слушателю, независимо от семантики слова, т.е. как бы «между слов», невербально, весьма значительную и важнейшую для слушателя информацию о говорящем, его отношение к собеседнику, к предмету разговора, к самому себе и т.п. Невербальная информация может как значительно усилить семантическое значение слова, так и ослабить вплоть до полного его отрицания субъектом восприятия. Ввиду значительной непроизвольности и подсознательности восприятия невербальной информации слушатель ориентируется не только по вербальному, но и невербальному смыслу сообщения.

Ключевые слова: распознавание эмоций, классификация, машинное обучение, глубокое обучение, сверхточные нейронные сети, регуляризация.

ВВЕДЕНИЕ

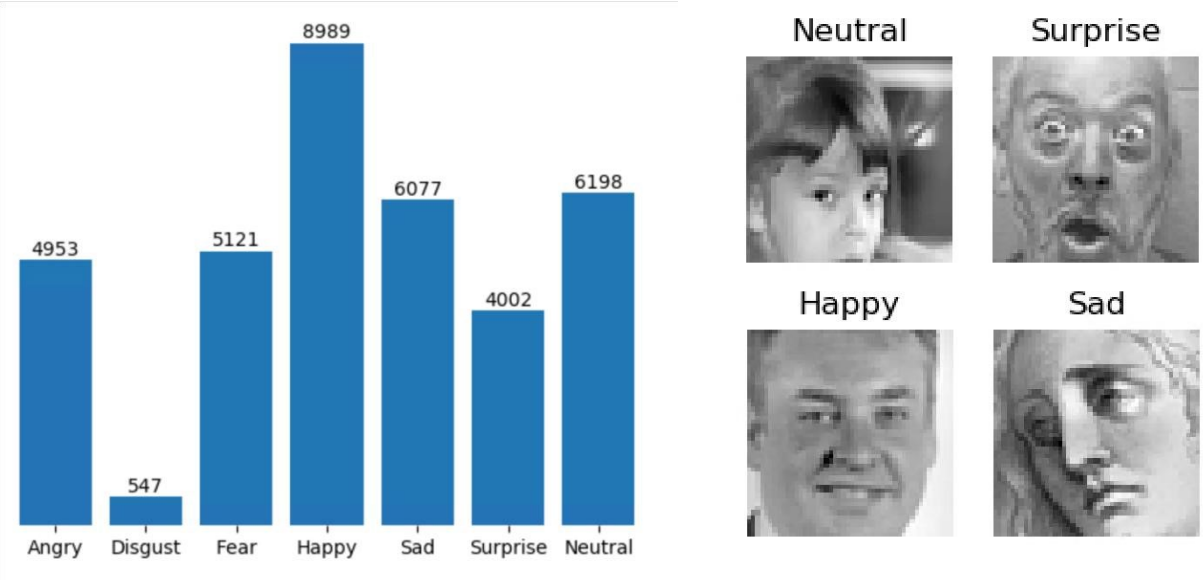
Изучением эмоций и их проявления ученые занимаются достаточно давно. Ведь эмоции являются неизбежной частью любой межличностной коммуникации, выражают отношение человека к окружающему миру, сложившейся вокруг него ситуации, к самому себе. Вместе с тем, в последнее время потребность в выявлении человеческих эмоций еще более возросла. В первую очередь, это связано с расширением сферы применения задачи распознавания эмоций. В настоящее время это и мониторинг состояния водителя за рулем, и системы видео аналитики «умного города», и маркетинговые исследования, и системы безопасности.

Эмоции могут быть выражены разными способами: мимикой, голосом, поведением, реакциями систем организма. Наибольший интерес из них представляет распознавание эмоций человека по выражению его лица. Эта задача является достаточно популярной в настоящее время по ряду причин: такие изображения несложно получить, они содержат много полезной информации для распознавания эмоций, собрать большой набор данных в виде изображений лиц достаточно легко (по сравнению с другим материалом для распознавания: речью или образцами почерка).

НАБОР ДАННЫХ

В качестве набора данных для обучения глубоких сетей был выбран Facial Expression Recognition 2013 (FER2013), который был представлен на конференции International Conference on Machine Learning 2013. Этот набор данных содержит 35 887 изображений с разрешением 48×48пикселей, большинство из которых сделаны в произвольных условиях. База данных была создана с использованием инструментов поиска изображений Google.

Каждое изображение классифицировано одним из семи видов эмоций: удивление (surprise), страх (fear), счастье(happy), гнев (angry), отвращение (disgust), грусть (sad) и нейтральное состояние или спокойствие (neutral). FER имеет большое число вариаций в изображениях, включая частичное закрытие лица (в основном, с помощью руки), низко контрастные изображения и лица в очках. Распределение данных по разным классам эмоций и примеры изображений лиц с указанием классов, к которым они отнесены, представлены на рис. 1.



а) б)

Рис. 1. Набор данных FER 2013

а – диаграмма распределения данных по различным классам эмоций б – примеры изображений лиц с указанием классов

В работе весь набор данных FER разделен на три части: обучающий набор, валидационный набор и тестовый набор. Первые два участвуют при обучении сети: обучающий набор используется для оптимизации весов модели, а валидационный набор предоставляет метрики после каждой эпохи обучения, которые помогают оценить качество обучения модели. Тестовый набор необходим для сравнения точности распознавания среди разных моделей.

АРХИТЕКТУРА И ОСОБЕННОСТИ НЕЙРОННЫХ СЕТЕЙ

Для распознавания эмоций в работе используется архитектура сверточной нейронной сети (Convolutionalneuralnetwork, CNN). Схематично CNN представляет собой последовательность слоев. Каждый слой преобразует один активационный объем в другой с помощью дифференцируемой функции. Для организации сверточной нейронной сети применяется 3 основных слоя: свертка (convolution), пулинг (иначе слой подвыборки или субдискретизации, англ. pooling) и полносвязный (fullyconnected, FC) слой. Слои свертки и пулинга используются для извлечения карты признаков из исходного изображения, а полносвязные слои используются для конечной классификации изображения по извлеченным признакам.

Размер входного слоя сети равен 48×48×1, в соответствии с размером изображений из набора данных. Выходной слой сети – это вектор из 7 элементов, соответствующих вероятностям принадлежности входного изображения к каждому из классов. В результате входное изображение относится к классу, имеющему максимальное значение вероятности.

В процессе исследования были построены две модели CNN. Первая модель содержит 2 слоя свертки, 2 слоя подвыборки и 4 полносвязных слоя. Подробная иллюстрация первой модели: размерности слоев, их параметры и используемые функции активации представлены на рис. 2.

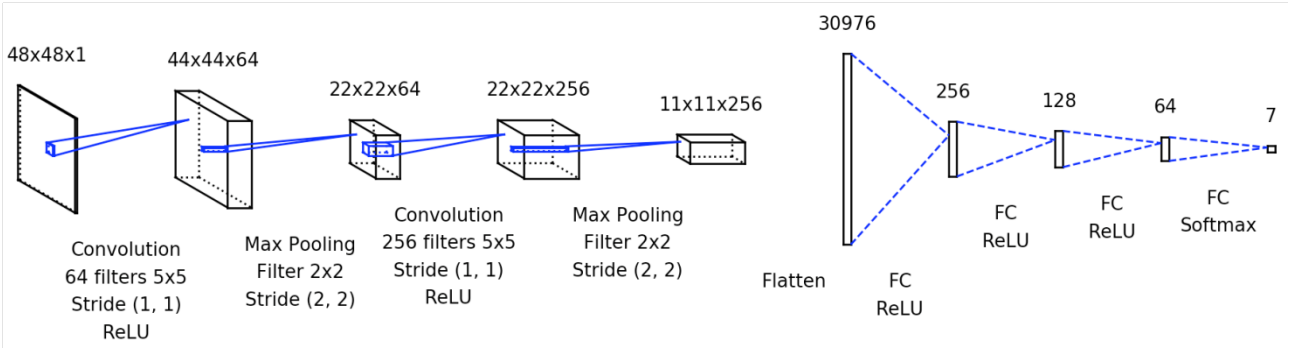


Рис. 2. Первая модель сверточной нейронной сети

Вторая модель является модернизацией первой модели и содержит 8 слоев свертки, 4 слоя подвыборки и 4 полносвязных слоя, а также механизм регуляризации. Иллюстрация второй модели представлена на рис. 3.

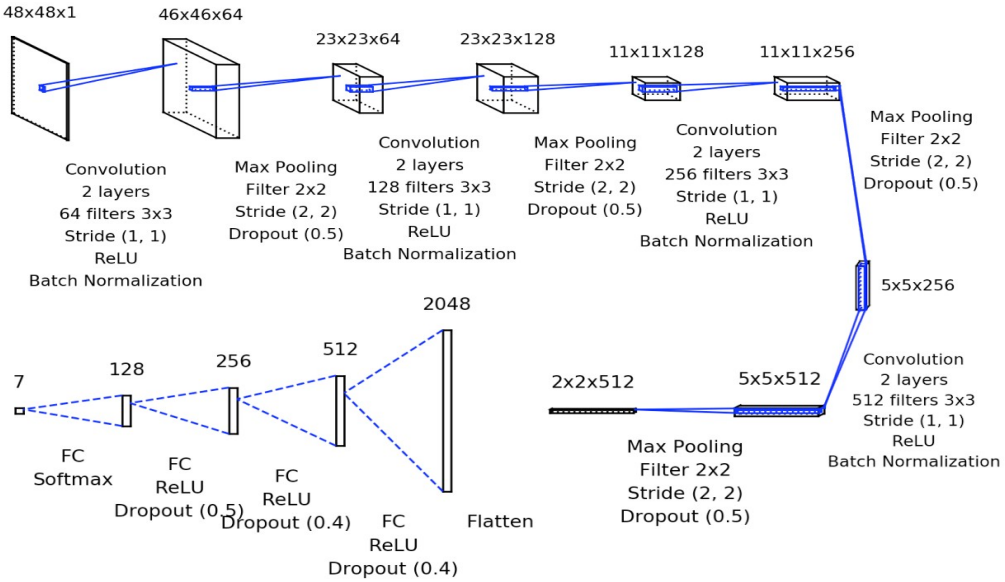


Рис. 3. Вторая модель сверточной нейронной сети

По сравнению с первой моделью, в ней большее количество сверточных слоев, которые имеют меньший размер матрицы свертки, что позволяет извлечь более детальную карту признаков. Механизм регуляризации позволяет избежать ситуации, называемой переобучением (overfitting). Характерным признаком переобучения является высокая точность распознавания на обучающей выборке и относительно низкая точность распознавания на тестовой выборке. Такая ситуация может возникнуть, если данные имеют много признаков, но при этом сам набор данных содержит мало примеров, либо в том случае, когда модель является слишком сложной для данных. Во второй модели сети для предотвращения ситуации переобучения используются такие механизмы регуляризации, как Batch Normalization и Dropout. Рассмотрим идеи, лежащие в основе этих механизмов.

Обычно для обучения нейронной сети выполняется некоторая предварительная обработка входных данных. Например, набор данных FER нормализуется таким образом, чтобы его данные напоминали нормальное распределение – имели нулевое математическое ожидание и единичную дисперсию. Такая обработка происходит для предотвращения раннего насыщения нелинейных функций активации слоев и обеспечения того, чтобы все входные данные находились в одном диапазоне значений. Но проблема возникает в промежуточных слоях, поскольку распределение значений, которое может иметь активационная функция, постоянно меняется в процессе обучения. Это замедляет процесс обучения, потому что каждый слой должен учиться приспосабливаться к новому распределению на каждом этапе обучения. Эта проблема известна как внутренний ковариантный сдвиг.

Суть метода Batch Normalization заключается в нормализации входных значений внутренних слоев нейронной сети и, таким образом, предотвращении возникновения внутреннего ковариантного сдвига. В процессе обучения, механизм Batch Normalization выполняет следующие действия.

Вычисляется математическое ожиданиеμBи дисперсия$\sigma B2$ входных значений слоя (1):

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i,$$
$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B). \quad (1)$$

Входные значения слоя нормализуются с помощью ранее рассчитанных статистических значений (2):

$$\overline{x_i} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2}}. \quad (2)$$

Нормализованные значения масштабируются и сдвигаются для того, чтобы избежать изменения представления данных в слое (3):

$$yi = \gamma \overline{xi} + \beta. \quad (3)$$

При этом параметры масштабирования γ и сдвига β настраиваются во время обучения совместно с другими параметрами сети.

Основная идея механизма Dropout состоит в том, чтобы случайно отбрасывать отдельные нейроны в слоях (вместе с их связями) из нейронной сети во время обучения. Так как отброшенные нейроны перестают вносить свой вклад в процесс обучения сети, то это становится равносильно обучению новой нейронной сети. Это предотвращает слишком большую адаптацию нейронов друг к другу. Каждый слой, использующий Dropout, имеет параметр, определяющий вероятность исключения нейрона из сети.

РЕЗУЛЬТАТЫ РАБОТЫ

После обучения первая сеть продемонстрировала точность распознавания эмоций 52 % на тестовом наборе данных. При этом на обучающем наборе данных точность распознавания составила 98 %. График изменения точности в процессе обучения модели представлен на рис. 4.

Матрица ошибок, построенная на тестовом наборе данных, представлена на рис. 5. В матрице ошибок строки и столбцы обозначены одним из семи классов эмоций. На пересечении указано количество вариантов, отнесенных к классу эмоций, обозначающему столбец, но реально принадлежащих к классу эмоций, обозначающему текущую строку. По матрице видно, что наименьшей ошибке подвержено распознавание эмоции «счастье» (23 % ошибок), наибольшей – распознавание эмоций «страх» и «гнев» (65 % ошибок).

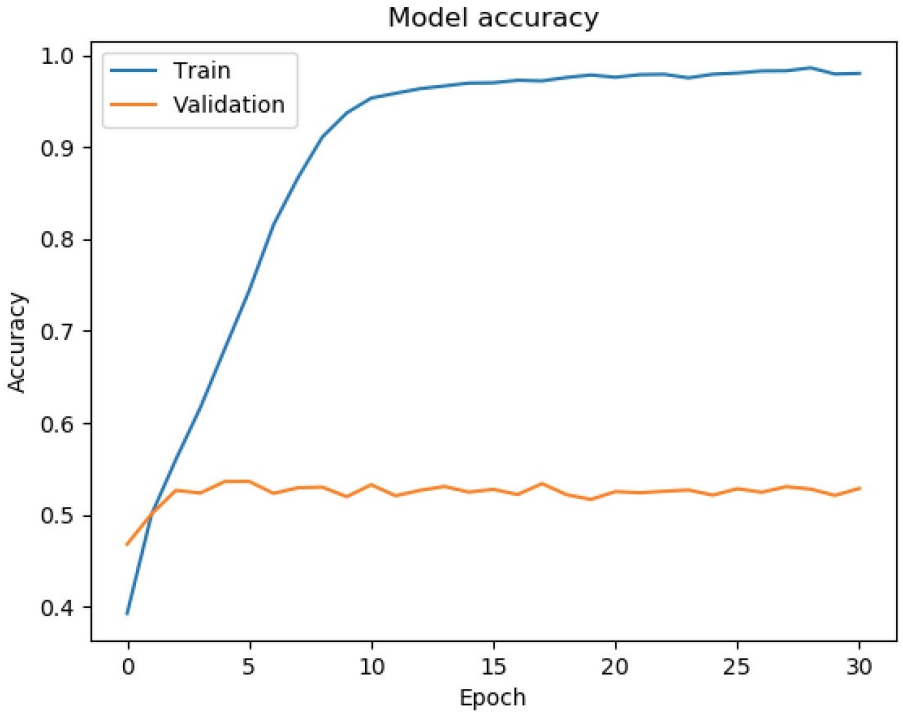


Рис. 4. График изменения точности распознавания модели в зависимости от эпохи обучения

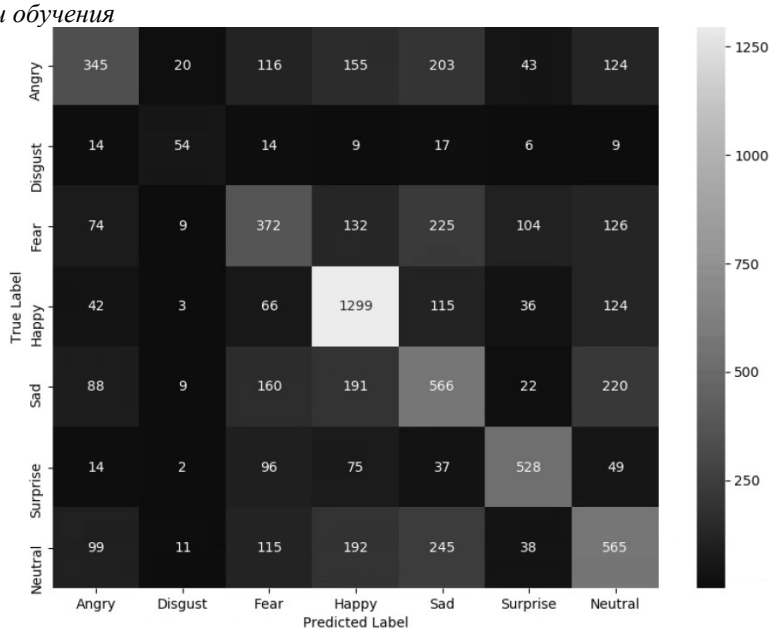


Рис. 5. Матрица ошибок первой модели

Для второй сверточной сети использовались тот же самый набор данных, что и для первой сети. В результате обучения сеть показывает точность распознавания 92 % на обучающем наборе данных, но при этом на валидационном наборе данных точность достигает 64 % (рис. 6). Корреляция между правильными и ошибочными распознаваниями представлена матрицей ошибок на рис. 7.

Высокая точность распознавания на обучающей выборке и относительно низкая точность распознавания на тестовом наборе являются признаком переобучения сети. Как было указано ранее, решением данной проблемы является механизм регуляризации, который добавлен во вторую сверточную сеть.

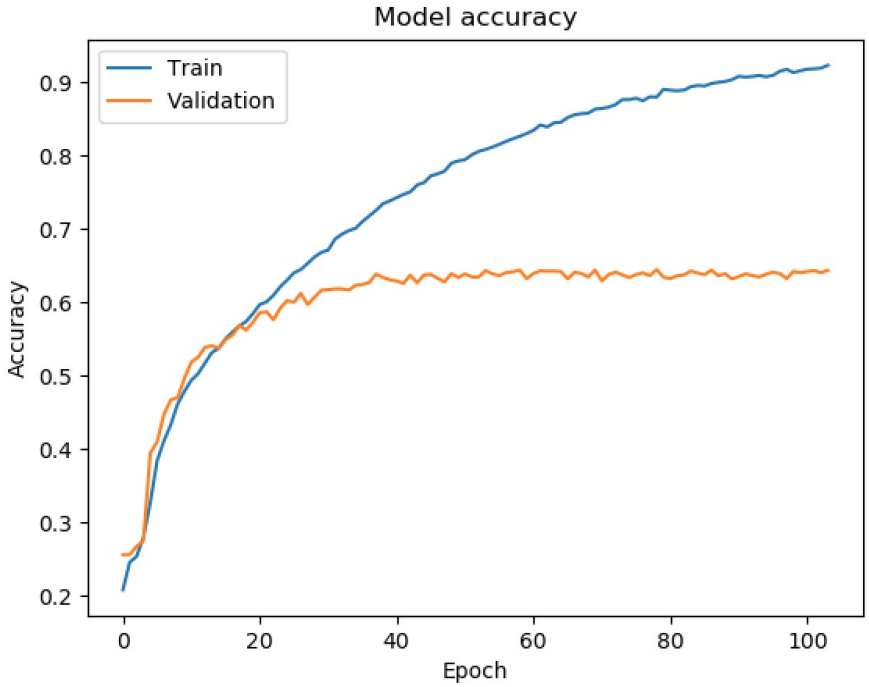


Рис. 6. График зависимости точности распознавания модели от номера эпохи обучения

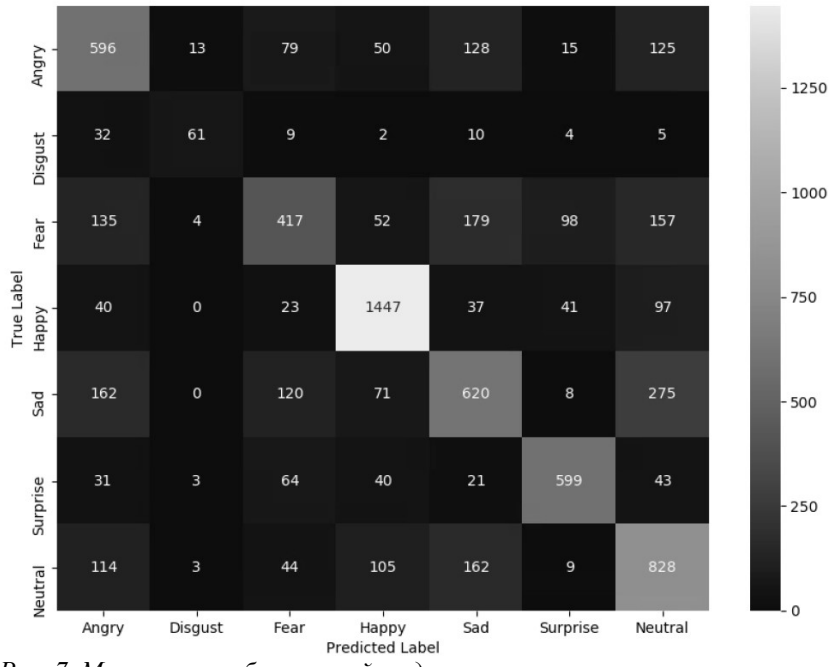


Рис. 7. Матрица ошибок второй модели

Анализ результатов позволяет сделать вывод, что регуляризация совместно с добавлением новых сверточных слоев в модели улучшила точность распознавания на 12 %.

РЕЗУЛЬТАТЫ РАБОТЫ

Для тестирования обученной модели на произвольных данных было разработано вспомогательное приложение, которое позволяет классифицировать эмоции на заданном изображении или видео. В качестве источника данных может выступать как заранее записанное видео, так и видео, поступающее с камеры в реальном времени.

Для декодирования и покадровой обработки видео используется библиотека OpenCV. Поиск лиц на отдельном кадре осуществляется методом Виолы-Джонса. Данный метод демонстрирует высокую точность поиска лица на изображении вместе с быстрой скоростью работы. Также существуют альтернативные методы поиска лица, основанные на сверточных нейронных сетях, но они требуют большего количества ресурсов для обработки изображения, вследствие чего метод Виолы-Джонса является более приемлемым вариантом для классификации эмоций в реальном времени с высокой частотой кадров.

После выполнения поиска лиц по методу Виолы-Джонса все найденные лица на кадре подвергаются ряду преобразований для улучшения точности дальнейшей классификации.

Выравнивание положения лица по вертикали и горизонтали.

Гамма-коррекция.

Объединение нескольких цветовых каналов в один для получения изображения в градациях серого.

Изменение размера изображения до 48×48 пикселей.

Далее преобразованный набор лиц передается на вход классификатору – обученной модели. После завершения классификации каждому изображению лица будет присвоен соответствующий класс эмоции. Завершающим этапом обработки кадра является визуализация полученных классов – каждое найденное лицо на кадре обозначается цветной рамкой и маркируется названием присвоенной ему эмоции.

Схематично процесс обработки кадра изображен на рис. 8.

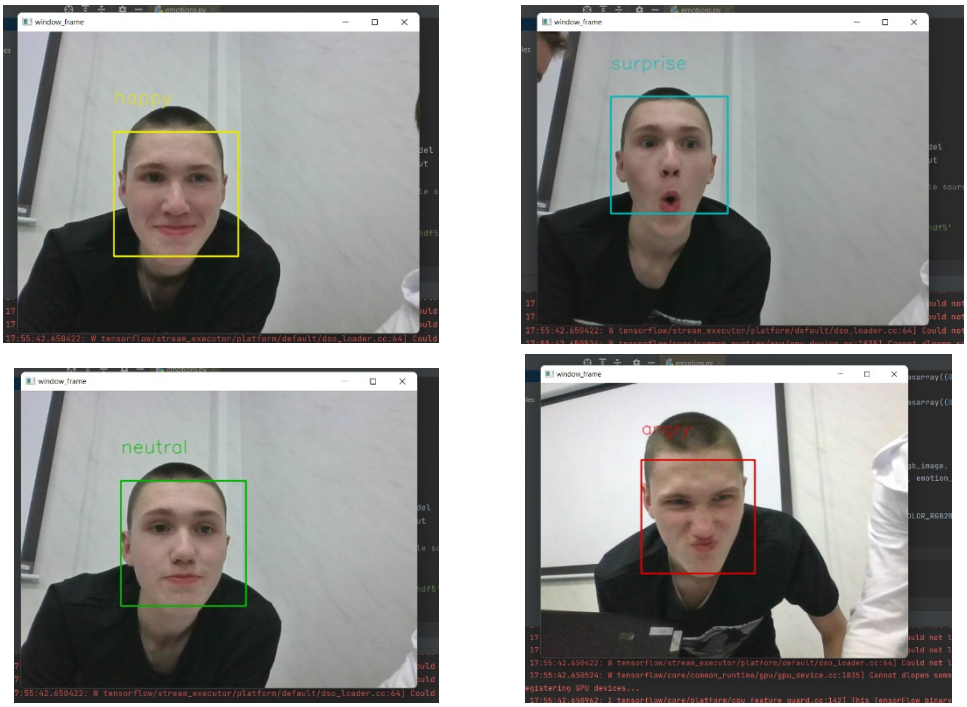


Рис. 8. Процесс обработки кадра

ЗАКЛЮЧЕНИЕ

В результате проведенного проектирования сетей и их последующего обучения наилучшая полученная точность классификации эмоций по изображению лица составила 64 %. При этом построенная матрица ошибок демонстрирует, что полученная точность классификации в первую очередь обусловлена неравномерным распределением данных по классам в исходном наборе данных. Так, количество изображений, отнесенных к классу «отвращение», в 16 раз меньше, чем количество изображений, отнесенных к классу «счастье».

Тестирование модели на произвольных данных, не относящихся к набору данных FER, позволило качественно оценить точность распознавания эмоций. Было выявлено, что из-за низкого разрешения входного изображения модели возникает погрешность в распознавании.

СПИСОК ЛИТЕРАТУРЫ

1. Takahashi T., Kishino F. Hand gesture coding based on experiments using a hand gesture interface device // ACM SIGCHI Bulletin. – 1991. – Vol. 23 (2). – P. 67–74.
2. Lee C., Xu Y. Online, interactive learning of gestures for human/robot inter-faces // Proceedings of the IEEE International Conference on Robotics and Automation. – 2002. – Vol. 4. –
P. 2982–2987.
3. Smedt Q. De, Wannous H., Vandeborre J.-P. Heterogeneous hand gesture recognition using
3D dynamic skeletal data // Computer Vision and Image Understanding. – 2019. – Vol. 181. –
P. 60–72.
4. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot
interaction in real time / J. Dong, Z. Xia, W. Yan, Q. Zhao // Journal of Visual Communication and
Image Representation. – 2019. – Vol. 63. – P. 102583.
5. Real-time gesture recognition based on feature recalibration network with multi-scale information / Z. Cao, X. Xu, B. Hu, M. Zhou, Q. Li // Neurocomputing. – 2019. – Vol. 347. – P. 119–130.
6. Голиков И. Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество. – URL: https://habr.com/ru/post/348000
7. Функции активации в нейронных сетях. – URL: http://www.aiportal.ru/articles/neuralnetworks/activation-function.html
8. Библиотека Keras. Слой пуллинга. – URL: https://keras.io/layers/pooling/
9. Библиотека Keras. Полносвязный слой. – URL: https://keras.io/layers/core
10. Библиотека Keras. Model class API. – URL: https://keras.io/models/model/
11. Основы планирования эксперимента: методическое пособие / сост. К.М. Хамханов. –
Улан-Удэ, 2001. – URL: http://window.edu.ru/resource/438/18438/files/Mtdukm8.pdf
12. Проверка адекватности регрессионной модели. – URL: https://helpstat.ru/proverkaadekvatnosti-regressionnoj-modeli/
13. Системы распознавания речи – URL: https://ru.wikipedia.org/wiki/Распознавание_речи#Классификация _систем_распознавания_речи
14.Распознавание голоса. О технологии и ее значении для маркетологов – URL: https ://www.comagic.ru/blog/posts/may/raspoznavanie_golosa_o_tekhnologii_i_ee_znachenii_dlya_marketologov/
15. Herb Sutter. The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software
16. И. А. Шалимов, М. А. Бессонов. Анализ состояния и перспектив развития технологий определения языка аудиосообщения
17. Как устроена технология распознавания речи Yandex SpeechKit от Яндекса | Хабрахабр – URL: https://habr .com/ru/company/yandex/blog/198556/
17. Davies , K.H., Biddulph, R. and Balashek, S. (1952) Automatic Speech Recognition of Spoken Digits, J. Acoust. Soc. Am. 24 (6) pp. 637—642
18. Pedro Domingos, A Few Useful Things to Know about Machine Learning – URL: http://homes.cs.washington. edu/~pedrod/papers/cacm12.pdf
19. Pedro Domingos, “The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World”, 2015 (на русском: “Верховный алгоритм. Как машинное обучение изменит наш мир”)
20. Ian Goodfellow, Yoshua Bengio, Aaron Courville, “Deep Learning”, MIT
Press, 2016, http://www.deeplearningbook.org/
21. “Evaluating Machine Learning Models. A Beginner's Guide to Key Concepts and Pitfalls”, 2015 – URL: http://www.oreilly.com/data/free/evaluating-machine-learning-models.csp
22. http://eclass.cc/courselists/4_machine_learning
23. http://eclass.cc/courselists/117_deep_learning
24.http://eclass.cc/courselists/42_data_science

25. Gaind, B. Emotion Detection and Analysis on Social Media / B. Gaind, V. Syal, S. Padgalwar // Global Journal of Engineering Science and Researches (ICRTCET-18). 2019. – P. 78-89.
26. Facial Expression Recognition Challenge// Deeplearning URL: http://deeplearning.net/icml2013workshop-competition/challenges/ (датаобращения: 22.12.2019).
27.Schmidhuber, J. Deep Learning in Neural Networks: An Overview / J. Schmidhuber // Neural Networks. – 2015. – №61. – P. 85-117.
28.Salman, S. Overfitting Mechanism and Avoidance in Deep Neural Networks [Электронный ресурс] / S. Salman, X. Liu // arXiv.org. 2019. URL: https://arxiv.org/abs/1901.06566 (дата обращения: 17.01.2020).
29.Ioffe, S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [Электронный ресурс] / S. Ioffe, C. Szegedy // arXiv.org. 2015. URL: https://arxiv.org/abs/1502.03167 (дата обращения: 11.01.2020).
30.Srivastava, N. Dropout: A Simple Way to Prevent Neural Networks from Overfitting / N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov // Journal of Machine Learning Research. – 2014. – №15. – P. 1929-1958.
31.Culjak, I. A brief introduction to OpenCV / I. Culjak, D. Abram, T. Pribanic, H. Dzapo, M. Cifrek // 2012 Proceedings of the 35th International Convention MIPRO, Opatija. 2012. – P. 1725-1730.
32.Viola, P. Rapid object detection using a boosted cascade of simple features / P. Viola, M. Jones // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. – 2001. – T. 1.
33.Murillo, P.C.U. Comparison between CNN and Haar classifiers for surgical instrumentation classification / P.C.U. Murillo, R.J. Moreno, J.O.P. Arenas // Contemporary Engineering Sciences. – 2017. – T. 10. – № 28. – P. 1351-1363.
34.Anila, S. Preprocessing Technique for Face Recognition Applications under Varying Illumination Conditions / S. Anila, N. Devarajan // Global Journal of Computer Science and Technology Graphics & Vision. – 2012. – T. 12. – № 11.