МРНТИ 28.23.15
УДК 004.93, 004.89

https://doi.org/10.51889/2959-5894.2023.81.1.018

Д.Ж. Сатыбалдина[1], К.А. Калымова[2], Д.М. Сыдыков[1],*

[1]Евразийский национальный университет им. Л.Н. Гумилёва, г.Астана, Казахстан
[2]Казахский национальный женский педагогический университет, г.Алматы, Казахстан
[*]e-mail: gulzia_kalymova@mail.ru

ПРИМЕНЕНИЕ ТРАНСФЕРА ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

Аннотация

В последние годы машинное обучение применяется в различных областях и приложениях, и его использование продолжает увеличиваться. Сложные задачи требуют большого количества данных и времени для обучения, и данные нужно пометить для обучения с учителем. Трансферное обучение позволяет сократить объем исходных данных и время обучения или повысить точность решения интеллектуальной задачи. В работе представлены результаты экспериментальных исследований по распознаванию статических жестов рук на основе предлагаемой нами модели глубокой нейронной сети, с традиционным полным обучением на всех параметрах, и сверточной нейронной сети архитектуры VGG-16, предобученной с использованием концепции трансфера обучения. Программная реализация системы распознавания жестов выполнена с использованием Python-библиотек обработки изображений, полученных с глубинного сенсора захвата изображений. Производительность предлагаемой в работе модели глубокой нейронной сети сравнивается с моделью трансферного обучения для модифицированной архитектуры VGG-16.

Ключевые слова: глубокое обучение, сверточная нейронная сеть, трансфер обучения.

Аңдатпа

Д.Ж. Сатыбалдина[1], К.А. Калымова[2], Д.М. Сыдыков[1],
[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан
[2] Қазақ ұлттық қыздар педагогикалық университеті, Алматы қ., Казахстан

БЕЙНЕЛЕРДІ КЛАССИФИКАЦИЯЛАУ ҮШІН НЕЙРОНДЫҚ ЖЕЛІЛЕРГЕ ТРАНСФЕРТТІК ОҚЫТУДЫ ҚОЛДАНУ

Соңғы жылдары машиналық оқыту әртүрлі салаларда және қолданбаларда қолданылады және оны пайдалану артып келеді. Күрделі тапсырмаларды шешу үшін көп деректер мен уақыт қажет, ал деректер бақыланатын оқыту үшін белгіленуі керек. Трансферттік оқыту бастапқы деректер көлемін және оқу уақытын қысқартуға немесе интеллектуалды мәселені шешудің дәлдігін арттыруға мүмкіндік береді. Бұл мақалада біз ұсынған терең нейрондық желі үлгісіне негізделген статикалық қол қимылдарын тану бойынша эксперименталды зерттеулердің нәтижелері келтірілінген, барлық параметрлер бойынша дәстүрлі толық оқыту және трансферттік оқыту тұжырымдаманы пайдалана отырып алдын ала дайындалған VGG-16 архитектурасының конволюциялық нейрондық желісі алынып отыр. Қимылдарды тану жүйесін бағдарламалық қамтамасыз етулуі Python кескінді өңдеу кітапханалары арқылы терең суретке түсіру сенсорынан алынып жасалыңды. Ұсынылған терең нейрондық желі моделінің өнімділігі модификацияланған VGG-16 архитектурасы үшін трансферттік оқыту моделімен салыстырылады.

Түйін сөздер: терең оқыту, конволюциялық нейрондық желі, трансферттік оқыту.

Abstract

APPLYING OF TRANSFER LEARNING TO NEURAL NETWORKS FOR IMAGE CLASSIFICATION

Satybaldina D.ZH. [1], Kalymova K.A. [2], Sydykov D.M. [1]
[1] L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[2]Kazakh National Women's Teacher Training University, Almaty, Kazakhstan

In recent years, machine learning has been applied in various fields and applications, and its use continues to increase. Complex tasks require a lot of data and time to train, and the data needs to be labeled for supervised training. Transfer learning allows reducing the amount of raw data and learning time, and improving the accuracy of solving an intellectual problem. This paper presents the results of experimental studies on the static hand gesture recognition based on our proposed deep neural network model, with traditional full learning on all parameters, and a convolutional neural network of the VGG-16 architecture, pre-trained using the concept of transfer learning. Software implementation of the gesture

recognition system was made using Python image processing libraries obtained from the image capture sensor. The performance of the proposed deep neural network model is compared with the model for the modified VGG-16 architecture and the transfer learning.

Keywords: deep learning, convolutional neural network, transfer learning.

1 Введение

Создание и внедрение эффективных и точных систем распознавания жестов рук способствует развитию инновационных технологий человеко-машинного взаимодействия (a human-machine interaction, HMI) [1]. Системы HMI на основе жестов рук применяются в компьютерных играх [2] и приложениях с виртуальной реальностью [3], при управлении интеллектуальными устройствами «умного» дома [4], во взаимодействиях человека и робота (a human–robot interaction, HRI) [5] или человека с беспилотным летательным аппаратом. Для этих целей распознавание жестов рук представляет собой процесс отслеживания человеческих жестов, идентификации и преобразования их в семантически значимые команды для управления устройствами. Технологии отслеживания жестов для этих задач используют устройства захвата изображения, методы компьютерного зрения и алгоритмы машинного обучения, в том числе алгоритмы глубокого обучения для многослойных нейронных сетей.

Ранее нами в работах [6-9] были предложены подходы по проектированию и программной реализации системы распознавания статических и динамических жестов рук. Получены результаты экспериментальных исследований по распознаванию жестов рук на основе цифровой обработки видео потока в режиме реального времени, предобработки кадров, выделения признаков, идентифицирующих классы жестов, и классификации посредством модели сверточной нейронной сети. В качестве устройства захвата жестов использовались веб-камера Logitech HD Pro Webcam C920 и камера глубины Intel RealSense D435. Экспериментальные результаты показали, что точность распознавания жестов зависит как от условий демонстрации поз рук (освещенность, расстояние до камеры), так и типа используемого изображения (RGB или RGB-D, от веб-камеры и камеры глубинного зрения, соответственно). Средняя точность классификатора при обучении на 2000 изображениях, полученных из видео потока от камеры глубины при нормальном освещении и среднем расстоянии до камеры, на уровне 97,35 % и 91,31%, для статических и динамических жестов рук, соответственно. Точность распознавания изображений, полученных от RGB-камеры, ниже: 91,4 % и 84,8% при отслеживании статических и динамических жестов рук. На этапе тестирования максимально полученная точность распознавания статических жестов при увеличении расстояния до нескольких метром от сенсора захвата изображения в условиях плохой освещенности составляет около 78%. Таким образом, производительность реализованной системы распознавания жестов рук остается недостаточно высокой для практического применения в разных условиях.

Обучение глубоких нейронных сетей представляет собой процесс настройки весовых коэффициентов множества нейронов скрытых слоев с использованием алгоритма оптимизации, занимает много времени и требует наличия больших баз входных данных [10]. В работе [10] сформулировано следующее эвристическое правило: «алгоритм глубокого обучения с учителем достигает приемлемого качества при наличии примерно 5000 помеченных примеров на категорию и оказывается сопоставим или даже превосходит человека, если обучается на датасете, содержащем не менее 10 миллионов помеченных примеров».

Актуальным направлением исследований является поиск путей повышения эффективности модели глубокого обучения при работе с наборами данных меньшего размера.

В нескольких исследованиях было предложено концепция трансфера обучения для преодоления уменьшения времени на обучение и размера используемых обучающих выборок [11-12]. Методы трансферного обучения успешно применяются во многих реальных приложениях, включая определение спелости фруктов, тонкую настройки больших предварительно обученных моделей для классификации текстов, определения на медицинских изображениях участков кожи с злокачественными новообразованиями [11] или проявлений рака молочной железы [12].

В связи с этим, целью настоящей работы является развитие подходов для реализации систем распознавания жестов, позволяющих сократить время и объем данных на обучение нейросетевого классификатора, повысить точность детектирования жестов на основе комбинации современных сенсоров глубинного зрения и модели трансфера обучения сверточной нейронной сети. Основное внимание в этом исследовании уделяется влиянию методов сокращения переобучения на трансферное обучение с использованием сверточных нейронных сетей с архитектурой VGG16 с заменой и

переобучением полносвязных слоев, тонкой настройкой нейронов скрытых слоев для классификации статических жестов. Были использованы две модели сетей с различным числом обучаемых параметров с одним набором входных данных для обучения. Это исследование показывает, что производительность системы распознавания жестов рук с использованием трансферного обучения лучше, чем машинное обучение с традиционным извлечением признаков, и достигается при использовании простого метода оптимизации.

Остальная часть этой статьи структурирована следующим образом. В разделе 1.1 представлены краткие сведения по принципам работы алгоритмов машинного обучения, моделей глубокого обучения и трансферного обучения глубоких нейронных сетей. Методология исследования, подготовка датасетов и описание программной системы распознавания жестов рук приведены в разделе 2. В разделе 3 и 4 соответственно представлены результаты экспериментов и обсуждение производительности обученных сверточных нейронных сетей при варьировании расстояния до сенсора захвата изображений. Заключение и будущие исследования представлены в разделе 5.

1.1 Концепция трансфера обучения

Машинное обучение (a machine learning) представляет собой способ решения сложно формализуемых интеллектуальных задач на основе поиска сложных закономерностей и паттернов во входных данных и ответных реакций на входные сигналы [13]. Система машинного обучения не программируется в явном виде, а обучается. В классическом программировании используется парадигма символического искусственного интеллекта, когда пользователи вводят данные и правила (программа), данные обрабатываются в соответствии с правилами, получая ответы как результат решения интеллектуальной задачи (см. рисунок 1*a*). При машинном обучении пользователи вводят данные, а также ответы, ожидаемые от этих данных, и получают на выходе правила, которые можно применить к новым данным для получения новых ответов (см. рисунок 1*b*).

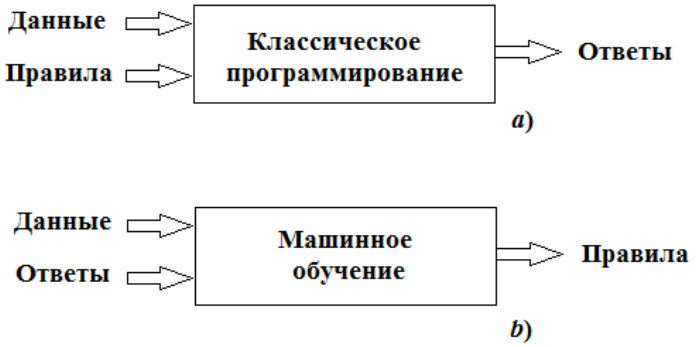

Рисунок 1. Парадигмы программирования интеллектуальных систем [13]: a) классическое программирование; b) машинное обучение.

Машинное обучение показывает хорошую применимость в задачах, связанных с многомерными данными, такими как классификация, регрессия и кластеризация. Изучая предыдущие вычисления и извлекая закономерности из массивных баз данных, алгоритмы машинного обучения помогают получить надежные и воспроизводимые решения. По этой причине успешно применяются во многих областях, таких как распознавание речи и изображений или обработка естественного языка, обнаружение аномалий в сетевом трафике и оценка кредитоспособности.

Для реализации алгоритма машинного обучения требуются следующие компоненты модели [10]:

– машинные представления входных данных (например, для задачи классификации изображений необходимы файлы изображений с заданным разрешением);

– примеры ожидаемых результатов (в задаче распознавания изображений ожидаемыми результатами могут быть такие теги, как «треугольник», «квадрат» и т.д.)

– способ измерения точности алгоритма (может быть измерено расстояние между текущим выходом алгоритма и его ожидаемым результатом), измерение используется в качестве сигнала обратной связи, например, для корректировки весовых коэффициентов нейронов в нейросетевом классификаторе, данная итерация адаптации модели и является обучением.

Глубокое обучение (a deep learning) – это особая область машинного обучения, в которой используется многоуровневое представление данных, иерархическая структура данных из последовательных слоев все более значимых представлений [10]. В глубоком обучении эти многоуровневые представления (почти всегда) изучаются с помощью многослойных нейронных сетей, в которых содержатся нейроны со своими весовыми коэффициентами. В этом случае обучение означает модификацию набора числовых значений для весов нейронов всех слоев в сети таким образом, чтобы модель глубокого обучения правильно сопоставляла примеры входных данных со связанными с ними ожидаемыми результатами. В моделях глубокого обучения в качестве способа измерения точности используется функция потерь (the loss function), которая вычисляет оценку расстояния между выходом сети и ожидаемым результатом, алгоритм оптимизации использует значение потерь для обновления весов сети [10].

Традиционные модели машинного обучения разрабатываются для решения конкретных целевых задач. Каждая модель проектируется и полностью обучается на большом массиве входных данных и ответов на них. В зависимости от типа решаемой проблемы или типа модели могут потребоваться миллионы шаблонов для обучения. Все модели для наборов данных работают изолированно друг от друга (см. рисунок 2*a*). При этом каждая модель должна быть построена и обучена с нуля, что требует большого количества данных и времени на обучение.

Трансферное обучение (a transfer learning) – это подход, используемый для передачи знаний, полученных при решении одной задачи, для решения другой проблемы или для использования нового набора данных. Исходная модель, обученная для решения близкой целевой задачи, может стать основой для точной настройки целевой модели, уменьшая парадигму изоляции разных проблем (см. рисунок 2*b*). Эта процедура может помочь, например, повысить точность модели или сократить объем данных и время на ее обучения [13].

Как видно из рисунка 2, в обычном машинном обучении каждая отдельная задача решается изолированно с помощью своей модели обучения, в то время как трансферное обучение пытается извлечь знания из исходных задач в целевую задачи, где может быть значительно меньше помеченных данных для обучения с учителем.

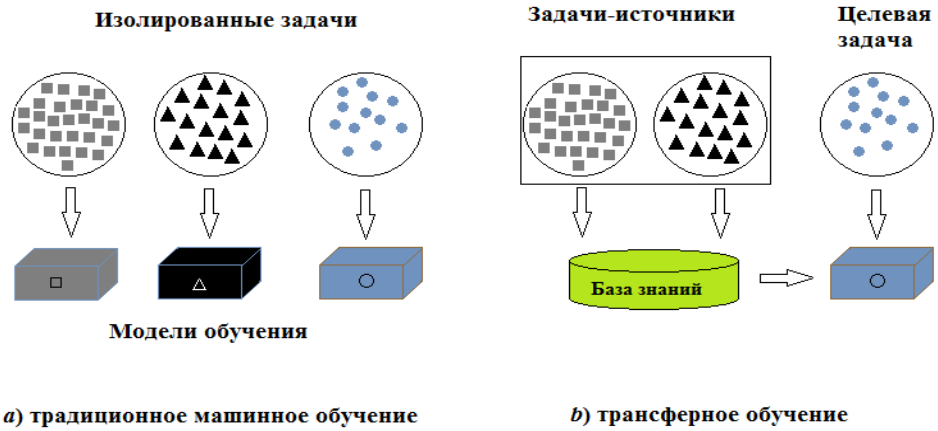

Рисунок 2. Сравнение процессов обычного машинного обучения и трансферного обучения

2 Методология исследования

2.1 Описание программной системы для распознавания жестов рук на основе предварительной обработки видео потока

Для программной реализации системы распознавания жестов на языке программирования Python использованы библиотеки сенсора захвата изображений RealSense от компании Intel, OpenCV и DL - фреймфорки с открытым исходным кодом Keras и TensorFlow. Диаграмма классов приложения для распознавания жестов рук представлена на рисунке 3.

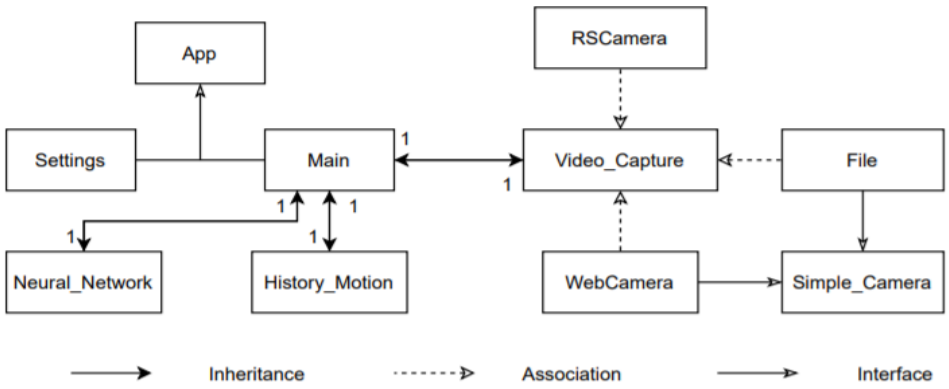

Рисунок 3. Диаграмма классов приложений для распознавания жестов рук

Методы и функции класса App используются для построения графического интерфейса, реализованного с помощью библиотеки Tkinter. Библиотека Pillow используется для рисования фреймов в графическом интерфейсе. Этот класс также содержит обработчик клавиатуры.

Класс Settings – это графическое приложение, предназначенное для выбора пользователем источника видео. Доступны следующие варианты: веб-камера, камера Intel и загрузка файла с видео. Графический интерфейс наследуется от класса App. Класс Main – основная программа, связующее звено между всеми классами. Получив выбор источника видео из класса Settings, инициализирует классы Video_Capture, History_Motion и Neural_Network для дальнейшей работы.

Классы Video_Capture, SimpleCamera, WebCamera, File и RSCamera – это классы, отвечающие за захват видеоклипов с камеры глубины RealSense или RGB-камеры. Подключение к камере глубины осуществляется с помощью библиотеки RealSense, имеющей стандартные функции инициализации камеры, настройки параметров ее работы, функции и методы чтения кадров из видеопотока, расчета расстояния от руки до камеры глубины, методы хранения RGB-изображений и карт глубины. Подключение к RGB камере осуществляется через библиотеку OpenCV.

Класс SimpleCamera отвечает за подключение к RGB-камере с помощью библиотеки OpenCV. В данной работе не использована RGB-камера для захвата изображений с жестами.

Класс RSCamera предназначен для подключения к камере глубины с помощью библиотеки RealSense. Кадр RGBD получается из видео потока и вычисляется средняя яркость кадра. Если яркость ниже порогового значения, то вместо RGB кадра в основной класс будет подаваться глубокий. Метод расчета средней яркости был разработан нами самостоятельно, идея метода заключается в переводе кадра из системы RGB в цветовую модель HSV (Hue, Saturation, Value), где V – значение яркости. HSV — это нелинейное преобразование RGB. Вычитание фона основано на сравнении пикселей глубины с порогом: если значение больше порога, то в результирующем кадре пиксель окрашивается в черный цвет, иначе - в цвет текущего RGB-кадра. Функции обнаружения рук основаны на вычислении необходимого количества пикселей в кадре без учета фона в фиксированной области.

Класс Neural_Network отвечает за подключение к нейронной сети с помощью библиотек Keras и TensorFlow. Методы класса предназначены для преобразования полученного кадра с изображением руки в формат, необходимый для подачи его на вход нейронной сети. В результате работы подключенной нейронной сети выдается вероятностная оценка сходимости с эталонными наборами жестов. Оценки отнесения образца к определенному классу патернов передаются на модуль Main для визуализации результата классификации.

Класс History_Motion предназначен для распознавания динамических жестов, в данной работе не используется.

2.2 База данных статических жестов рук

Мы подготовили базу данных, которая содержит изображения с сегментированными жестами, представленными на рисунке 4. Мы выбрали эти жесты, которые также включены в альтернативный набор данных [14], который можно использовать для сравнения или перекрестной проверки предложенной системы распознавания статических жестов рук. Камеру глубины Intel RealSense D400 разместили на штативе. 10 участников эксперимента представляли жесты в положении стоя, располагаясь перед сенсором на расстояние от 35 до 150 см.

Рисунок 4. Образцы жестов из обучающего набора данных

Изображения RGB и пиксели глубины с сенсора RealSense D400 получены их видео потока данных в режиме реального времени с последующей предобработкой и сохранением в виде файлов. Наш набор данных имеет в общей сложности 2160 изображений, в том числе 1080 изображений RGB и 1080 карт глубины, собранных при разных фонах в нескольких комнатах с изменениями расстояния до камеры глубины. Чтобы увеличить разнообразие базы данных, мы также попытались увеличить количество изображений с помощью вариации расстояния до сенсора, углов наклона, поворота ладони и т.д. Половина собранной базы статических жестов рук использовалась для полного обучения модели сверточной нейронной сети с начальной случайной инициализацией весов, а также для реализации трансфера обучения для модифицированной модели глубокой нейронной сети. На этапах валидации и тестирования системы распознавания жестов используются, соответственно, по 25% датасета жестов.

3 Результаты исследования

Эксперименты по обучению нейронных сетей проводились на процессоре Intel® Core(TM) i3- 8100 CPU, NVIDIA GeForce GTX 1050 Ti, 16 GB RAM. Обучение глубокой нейронной сети: весовые коэффициенты новых слоев инициализируются случайными значениями, после этого начинается процесс их обучения на обучающем наборе данных жестов рук. Применена стратегия обучения из конца в конец (end-to-end), при которой предобученные весовые коэффициенты не фиксируются, а корректируются под обучающий набор данных, т.е. поддаются «тонкой настройке». На вход нейронных сетей подаются картинки, преобразованные в тензоры размером 224*224*3, на выходном слое получается предсказание класса из 5 вариантов. Количество эпох при обучении было 8. Количество картинок при обучении 2160, распределенных для train/validation/test следующим образом: 1440/360/360. Алгоритмом оптимизации был выбран оптимизатор Adam со скоростью 0,0001. Функция потерь – категоричная кросс энтропия (categorical crossentropy).

3.1 Архитектура глубоких сверточных нейронных сетей

Для реализации классификации или распознавания образов используется модель глубокого машинного обучения, которая принимает иерархическое представление входного образца и предсказывает вероятность его отношения к некоторому классу объектов. В предлагаемом подходе по распознаванию статических жестов рук использована глубокая сверточная нейронная сеть (Deep Convolutional Neural Network, DCNN) с архитектурой VGG-16, которая является одной из 6-и конфигураций DCNN, предложенной авторами работы.

Базовая модель VGG-16 показана на рисунке 5, где на вход подаются изображения размером 224×224 в цветовой модели RGB. Входные изображения проходят через стеки сверточных (convolutional) слоев, слоев подвыборки (pooling) и полносвязанных (fully-connected) слоев. На выходном слое с количеством нейронов, соответствующих числу каналов, используется функция активации softmax, которая вычисляет распределение вероятности для классов объектов. Общее число параметров для обучения в стандартной модели с архитектурой VGG-16 превышает 14 миллионов. Модель VGG-16 была обучена на 1,3 миллионе изображений в базе данных ImageNet и протестирована на 100 000 изображений, достигнутая точность классификации на 1000 классов составила 92,7%.

Для оценки производительности модели трансферного обучения в эксперименте по распознаванию 5-и классов статических жестов будут использованы две модели DCNN.

Первая модель создана на основе стандартной модели VGG-16, которую можно импортировать с помощью библиотек Keras и TensorFlow. Это предварительно обученная модель в базе данных ImageNet, в которой полносвязанные слои с 1000 выходными каналами заменены на 4 плотных (dense layers) слоя с 5-ю нейронами в выходном слое для 5-и классов жестов. В таблице 1 представлено распределение обучаемых параметров по слоям и общее количество параметров для реализации трансфера обучения модифицированной модели VGG-16. Как видно, по общему числу параметров модифицированная сеть близка к стандартной модели VGG-16. Число обучаемых параметров для реализации трансфера обучения снизилось почти в 4 раза.

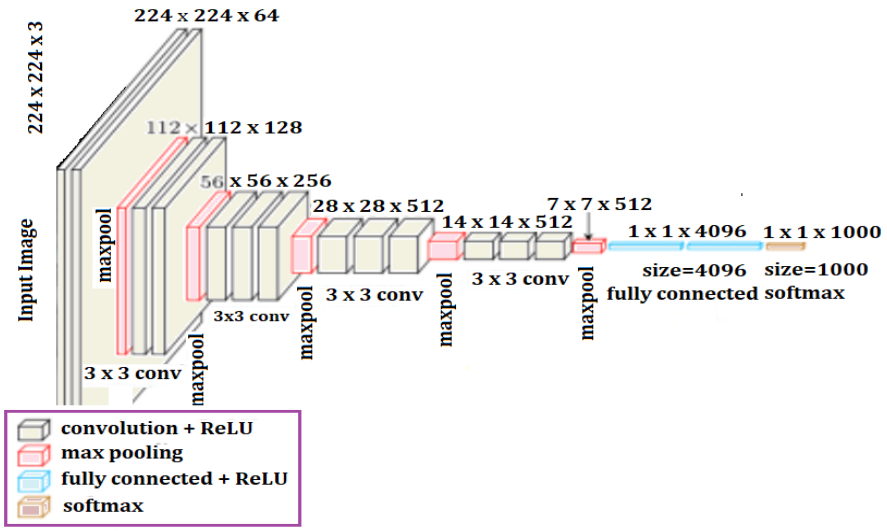

Рисунок 5. Архитектура глубокой сверточной сети VGG-16

Таблица 1. Распределение обучаемых параметров по слоям для реализации обучения сверточной нейронной сети с трансфером обучения

Тип слоя	*Выходная форма*	*Число параметров*
input_2 (InputLayer)	*(None, 224, 224, 3)*	*0*
block1_conv1 (Conv2D))	*(None, 224, 224, 64)*	*1792*
activation (Activation)	*(None, 224, 224, 64)*	*0*
block1_conv2 (Conv2D)	*(None, 224, 224, 64)*	*36928*
block1_pool(MaxPooling2D)	*(None, 112, 112, 64)*	*0*
block2_conv1 (Conv2D)	*(None, 112, 112, 128)*	*73856*
activation (Activation)	*(None, 112,11 2, 128)*	*0*
block2_conv2 (Conv2D)	*(None, 112, 112, 128)*	*147584*
block2_pool (MaxPooling2D)	*(None, 56, 56, 128)*	*0*
block3_conv1 (Conv2D)	*(None, 56, 56, 256)*	*295168*
block3_conv2 (Conv2D)	*(None, 56, 56, 256)*	*590080*
block3_conv3 (Conv2D)	*(None, 56, 56, 256)*	*65792*
block3_pool (MaxPooling2D)	*(None, 28, 28, 256)*	*0*
block4_conv1 (Conv2D)	*(None, 28, 28, 512)*	*1180160*
block4_conv2 (Conv2D)	*(None, 28, 28, 512)*	*2359808*
block4_conv3 (Conv2D)	*(None, 28, 28, 512)*	*262656*
block4_pool (MaxPooling2D)	*(None, 14, 14, 512)*	*0*
block5_conv1 (Conv2D)	*(None, 14, 14, 512)*	*2359808*
block5_conv2 (Conv2D)	*(None, 14, 14, 512)*	*2359808*
block5_conv3 (Conv2D)	*(None, 14, 14, 512)*	*262656*
block5_pool (MaxPooling2D)	*(None, 7, 7, 512)*	*0*
flatten_2 (Flatten)	*(None, 25088)*	*0*
fc1 (Dense)	*(None, 128)*	*3211392*
fc2 (Dense)	*(None, 128)*	*16512*
Fc3 (Dense)	*(None, 128)*	*16512*
dropout_2 (Dropout	*(None, 128)*	*0*
fc4 (Dense)	*(None, 64)*	*8256*
dense_2 (Dense)	*(None, 5)*	*325*
	Общее число параметров:	*13 249 093*
	Число обучаемых параметров:	*3 252 997*
	Число необучаемых параметров:	*9 996 096*

Вторая модель – это собственная модель нейронной сверточной сети, весовые коэффициенты новых слоев инициализируются случайными значениями, модель обучается с нуля. Порядок слоев и количество обучаемых параметров представлено в таблице 2. Как видно из сравнения данных в таблицах 1 и 2, упрощение архитектуры глубокой нейронной сети в модели 2 приводит к значительному уменьшению числа обучаемых параметров (более чем в 4 раза по сравнению со стандартной моделью VGG-16).

Для получения количественных показателей эффективности предложенной системы распознавания статических жестов рук используется матрица ошибок (a confusion matrix, CM), в которой каждый столбец представляет процент вероятности отнесения образца жестов к одному из 5-и классов. Вдоль главной диагонали указаны максимальные значения вероятности правильной предсказания, т.е. отнесения наблюдаемого объекта к некоторому классу жестов.

Недиагональные, элементы матрицы могут содержать значение вероятности классификации, сравнимое с некоторым пороговым значением. Если вычисленная вероятность классификации больше этого порога или равно ему, то предсказание считается правильным, в противном случае - неверным. Данный вариант матрицы ошибок соответствует формату выходных данных со сверточной нейронной сети, в финальном слое которой используется функция softmax, вычисляющая результат прогнозирования жеста в виде вероятностной величины от 0 до 1 (или от 0 до 100%).

Таблица 2. Распределение обучаемых параметров по слоям для реализации обучения сверточной нейронной сети без трансфера обучения

Тип слоя	*Выходная форма*	*Число параметров*
input_2 (InputLayer)	*(None, 224, 224, 3)*	*0*
conv2d_1 (Conv2D)	*(None, 222, 222, 32)*	*896*
activation_1 (Activation)	*(None, 222, 222, 32)*	*0*
max_pooling2d_1	*(MaxPooling2 (None, 111, 111, 32)*	*0*
conv2d_2 (Conv2D)	*(None, 109, 109, 32)*	*9248*
activation_2 (Activation)	*(None, 109, 109, 32)*	*0*
max_pooling2d_2	*(MaxPooling2 (None, 54, 54, 32)*	*0*
conv2d_3 (Conv2D)	*(None, 52, 52, 64)*	*18496*
activation_3 (Activation)	*(None, 52, 52, 64)*	*0*
max_pooling2d_3	*(MaxPooling2 (None, 26, 26, 64)*	*0*
flatten_1 (Flatten)	*(None, 43264)*	*0*
dense_1 (Dense)	*(None, 64)*	*2768960*
activation_4 (Activation)	*(None, 64)*	*0*
dropout_1 (Dropout)	*(None, 64)*	*0*
dense_2 (Dense)	*(None, 5)*	*325*
activation_5 (Activation)	*(None, 5)*	*0*
	Общее число параметров:	*2 797 925*
	Число обучаемых параметров:	*2 797 925*
	Число необучаемых параметров:	*0*

4 Дискуссия

В таблице 3 представлены матрицы средних вероятностей предсказаний жестов рук, полученные с использованием классификаторов на основе двух моделей глубокого обучения, в том числе с трансфером обучения.

Как видно из таблицы 3, для модели 1 с трансфером обучения точность классификации жестов составляет более 98%, и лишь небольшой процент выборок определяется как принадлежащий другим жестам, составляет примерно 1%. Это указывает на то, что предложенный подход имеет высокую производительность в двух измерениях: точность классификации и полнота. Исключением является распознавание жеста «Palm», где ошибка классификации превышает 2,5%, что в любом случае ниже порогового значения (50,75). Это можно объяснить определенным сходством этого жеста с жестом «Peace». Средняя точность распознавания жестов для модели 1 (по главной диагонали CM) достигает 98,75%. Для модели 2 с обучением всех параметров глубокой нейронной сети оценки точности классификации жестов ниже, чем для предобученной модели с архитектурой VGG-16.

Таблица 3. Матрицы ошибок для задачи распознавания статических жестов рук на расстоянии 50 см до камеры

Номер и тип модели	Входной образец жеста	Предсказанный жест				
		Fist	L	Okay	Palm	Peace
Модель 1 с использованием трансфера обучения	Fist	99,9567	0,0078	0,0353	0,0002	0
	L	0	99,9999	0,0001	0	0
	Okay	0	0,0052	99,9948	0	0
	Palm	0,1486	1,1342	0,875	95,2298	2,6124
	Peace	0	0,0804	0,3306	1,0001	98,5889
Модель 2 без использования трансфера обучения	Fist	88,2108	10,1923	1,3521	0,1149	0,1299
	L	0	99,9998	0,0002	0	0
	Okay	0,0008	6,7932	93,1919	0	0,0141
	Palm	0,9463	0,8103	0,7195	94,1164	3,4075
	Peace	0,0007	0,7128	4,8683	0,0013	94,4169

Это указывает на то, что модель с трансфером обучения дает лучшие оценки производительности системы распознавания жестов рук, чем использование в модели 2 архитектуры с меньшей глубиной слоев и почти сравнимым числом обучаемых параметров. Средняя точность распознавания жестов для модели 2 (по главной диагонали CM) достигает 93,98%.

Проведены экспериментальные исследования по распознаванию жестов рук при их демонстрации перед камерой глубины на расстоянии 75 см, 100 см и 125 см. На рисунках 6,7 представлены результаты распознавания статических жестов рук с использованием 2-х моделей глубокого обучения при варьировании расстояния до сенсора захвата изображений.

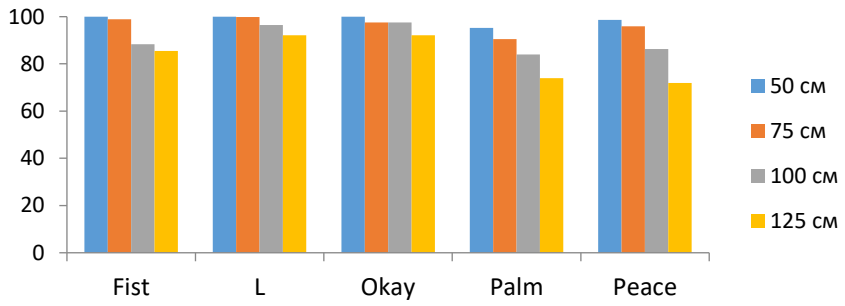

Рисунок 6. Средняя точность распознавания статических жестов рук с использованием предобученной модели глубокого обучения

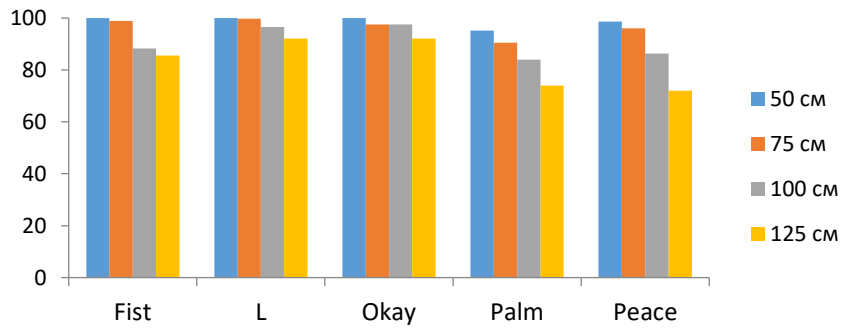

Рисунок 7. Средняя точность распознавания статических жестов рук с использованием модели сверточной сети без трансфера обучения

Анализ результатов классификации статических жестов рук при увеличении расстояния до сенсора захвата изображений показывает, что ошибки распознавания увеличиваются для обеих моделей глубокого обучения. Однако снижение точности классификации жестов в случае использования глубокой сверточной сети без трансфера более значительное. Причиной данного результата может быть уменьшение количества слоев в сетевых архитектурах с 16 до 10 от модели 1 к модели 2.

Нейросетевой классификатор, построенный на основе модифицированной архитектуры VGG-16, предобученной на 1,3 миллионе изображений в базе данных ImageNet, сохраняет высокую производительность при расстояниях больше 1 метра.

5 Заключение

В работе рассмотрена концепция трансферного обучения глубоких сверточных сетей, которая позволяет заимствовать помеченные данные или знания, извлеченные из некоторых связанных областей, чтобы помочь алгоритму машинного обучения достичь большей производительности в интересующей области.

Представлены результаты экспериментальных исследований производительности системы распознавания статических жестов рук, которая использует комбинацию глубинных представлений входных изображений и глубоких сверточных нейронных сетей. Предложенный подход реализован как законченный программный продукт на языке Python с использованием фреймворков глубокого обучения Keras и TensorFlow. Набор жестов рук собран вручную с использованием RGB-изображений и карт глубины от камеры глубины. База данных из 2160 образцов, выполненных 10 участниками эксперимента, использована для обучения, валидации и тестирования системы в пропорциях 50%/25%/25%, соответственно. В работе использована модифицированная модель глубокой сверточной сети с архитектурой VGG-16, предобученная на открытой базе изображений ImageNet. Использование техники трансфера обучения для данной модели с выходным слоем на 5 каналов позволяет снизить в несколько раз число обучаемых параметров, время на обучение и, в конечном счете, обеспечивает высокую точность на тренировочном наборе изображений, а также при тестировании на разных расстояниях демонстрации входных образцов до сенсора захвата изображений. Уменьшение числа слоев в архитектуре глубокой сверточной сети с целью снижения объема обучаемых параметров и времени обучения по сравнению со стандартной архитектурой VGG-16 не приводит к большому выигрышу в производительности системы распознавания жестов.

Полученные оценки классификации жестов подтверждают эффективность глубокого трансферного обучения и указывают на потенциальные возможности использования предложенной модели для будущих приложений на основе человеко-компьютерного взаимодействия.

Данная работа выполнена при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан (грант AP14872171).

Список использованной литературы:

1 *Xu J. et al. Robust hand gesture recognition based on RGB-D Data for natural human–computer interaction //IEEE Access. – 2022. – V. 10. – P. 54549-54562. https://doi.org/10.1109/ACCESS.2022.3176717*

2 *Goli A., Teymournia F., Naemabadi M. and Garmaroodi A.A. Architectural design game: A serious game approach to promote teaching and learning using multimodal interfaces //Education and Information Technologies. – 2022. – V. 27. – №. 8. – P. 11467-11498. https://doi.org/10.1007/s10639-022-11062-z*

3 *Rehman I. U. et al. Gesture-based guidance for navigation in virtual environments //Journal on Multimodal User Interfaces. – 2022. – V. 16. – №. 4. – P. 371-383. https://doi.org/10.1007/s12193-022-00395-1*

4 *Chen X. et al. An IoT and Wearables-Based Smart Home for ALS Patients //IEEE Internet of Things Journal. – 2022. – V. 9. – №. 21. – P. 20945-20956. https://doi.org/10.1109/JIOT.2022.3176202*

5 *Moysiadis V. et al. An Integrated Real-Time Hand Gesture Recognition Framework for Human–Robot Interaction in Agriculture //Applied Sciences. – 2022. – V. 12. – №. 16. – Article N.8160. https://doi.org/10.3390/app12168160*

6 *Сатыбалдина Д.Ж., Овечкин Г.В., Калымова К.А Система распознавания статических жестов рук с использованием камер глубины // Вестник РГРТУ - 2020. – № 72. – стр. 93-105. https://doi.org/10.21667/1995-4565-2020-72-93-105*

7 *Satybaldina D., Kalymova G.; Glazyrina, N. Application development for hand gestures recognition with using a depth camera // Communications in Computer and Information Science. – 2020, 1243 CCIS. – P. 55–67. https://doi.org/10.1007/978-3-030-57672-1_5*

8 *Satybaldina, D., Kalymova, G. Deep learning based static hand gesture recognition // Indonesian Journal of Electrical Engineering and Computer Science. 2021. 21(1). P. 398-405.* http://doi.org/10.11591/ijeecs.v21.i1.pp398-405

9 *Сатыбалдина Д.Ж., Глазырина Н.С., Степанов В.С., Калымова К.А. Разработка Python приложения для распознавания жестов рук из видеопотока RGB и RGBD камер// Вестник ЕНУ им. Л.Н. Гумилева. Математика. Компьютерные науки. Механика. 2021. Том 136, №3. С.6-17.* https://bulmathmc.enu.kz/index.php/main/article/view/93

10 *Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с анг. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.*

11 *Ali M. S. et al. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models //Machine learning with Applications. – 2021. – V. 5. – C. 100036.* https://doi.org/10.1016/j.mlwa.2021.100036

12 *Aljuaid H. et al. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning //Computer Methods and Programs in Biomedicine. – 2022. – V. 223. – Article N. 106951. https://doi.org/10.1016/j.cmpb.2022.106951*

13 *Ribani R., Marengoni M. A survey of transfer learning for convolutional neural networks //2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T). – IEEE, 2019. – P. 47-57.* https://doi.org/10.1109/SIBGRAPI-T.2019.00010

14 *Mantecón T., del Blanco C.R., Jaureguizar F., García N. Hand Gesture Recognition using Infrared Imagery Provided by Leap Motion Controller //International Conference on Advanced Concepts for Intelligent Vision Systems. – Springer, Cham, 2016. – LNCS 10016. – Pp. 47–57.* https://doi.org/10.1007/978-3-319-48680-2_5

References:

1 Xu J. et al. Robust hand gesture recognition based on RGB-D Data for natural human–computer interaction //IEEE Access. – 2022. – V. 10. – P. 54549-54562. https://doi.org/10.1109/ACCESS.2022.3176717

2 Goli A., Teymournia F., Naemabadi M. and Garmaroodi A.A. Architectural design game: A serious game approach to promote teaching and learning using multimodal interfaces //Education and Information Technologies. – 2022. – V. 27. – №. 8. – P. 11467-11498. https://doi.org/10.1007/s10639-022-11062-z

3 Rehman I. U. et al. Gesture-based guidance for navigation in virtual environments //Journal on Multimodal User Interfaces. – 2022. – V. 16. – №. 4. – P. 371-383. https://doi.org/10.1007/s12193-022-00395-1

4 Chen X. et al. An IoT and Wearables-Based Smart Home for ALS Patients //IEEE Internet of Things Journal. – 2022. – V. 9. – №. 21. –P. 20945-20956. https://doi.org/10.1109/JIOT.2022.3176202

5 Moysiadis V. et al. An Integrated Real-Time Hand Gesture Recognition Framework for Human–Robot Interaction in Agriculture //Applied Sciences. – 2022. – V. 12. – №. 16. – Article N.8160. https://doi.org/10.3390/app12168160

6 Satybaldina D.Zh., Ovechkin G.V., Kalymova K.A (2020) Sistema raspoznavaniya staticheskikh zhestov ruk s ispol`zovaniem kamer glubiny` [Static hand gesture recognition system using depth cameras]. Vestnik RGRTU 2020. № 72. 93-105 (In Russian) https://doi.org/10.21667/1995-4565-2020-72-93-105

7 Satybaldina D., Kalymova G.; Glazyrina, N. Application development for hand gestures recognition with using a depth camera // Communications in Computer and Information Science.- – 2020, 1243 CCIS. – P. 55–67. https://doi.org/10.1007/978-3-030-57672-1_5

8 Satybaldina, D., Kalymova, G. Deep learning based static hand gesture recognition // Indonesian Journal of Electrical Engineering and Computer Science. 2021. 21(1). P. 398–405. http://doi.org/10.11591/ijeecs.v21.i1.pp398-405

9 Satybaldina D.Zh., Glazyrina N.S., Stepanov V.S., Kalymova K.A. Razrabotka Python prilozheniya dlya raspoznavaniya zhestov ruk iz videopotoka RGB i RGBD kamer [Development of a Python application for recognizing gestures from a video stream of RGB and RGBD cameras]// Vestnik ENU im. L.N. Gumileva. Matematika. Komp`yuterny`e nauki. Mekhanika. – 2021. – Tom 136, №3. –S.6-17. https://bulmathmc.enu.kz/index.php/main/article/view/93 *(In Russian)*

10 Gudfellou Ya., Bendzhio I., Kurvill` A. (2018) Glubokoe obuchenie [Deep Learning]. per. s ang. A. A. Slinkina. 2-e izd., ispr. M.: DMK Press, 652. (In Russian)

11 Ali M. S. et al. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models //Machine learning with Applications. – 2021. – V. 5. – C. 100036. https://doi.org/10.1016/j.mlwa.2021.100036

12 Aljuaid H. et al. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning //Computer Methods and Programs in Biomedicine. – 2022. – V. 223. – Article N. 106951.https://doi.org/10.1016/j.cmpb.2022.106951

13 Ribani R., Marengoni M. A survey of transfer learning for convolutional neural networks //2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T). – IEEE, 2019. – P. 47-57. https://doi.org/10.1109/SIBGRAPI-T.2019.00010

14 Mantecón T., del Blanco C.R., Jaureguizar F., García N. Hand Gesture Recognition using Infrared Imagery Provided by Leap Motion Controller //International Conference on Advanced Concepts for Intelligent Vision Systems. – Springer, Cham, 2016. – LNCS 10016. – Pp. 47–57. https://doi.org/10.1007/978-3-319-48680-2_5