

DOI 10.66032/2221-2574-2024-1-3-62-72

УДК 004.93

ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ ОШИБОК ПРИ ВЫПОЛНЕНИИ ФИЗИЧЕСКИХ УПРАЖНЕНИЙ

Титаренко Дмитрий Юрьевич

магистрант кафедры физики и прикладной математики Муромского института (филиала) ФГБОУ ВО «Владимирский Государственный университет имени А.Г. и Н.Г. Столетовых».

Рыжкова Мария Николаевна

кандидат технических наук, доцент, декан факультета информационных технологий и радиоэлектроники Муромского института (филиала) ФГБОУ ВО «Владимирский Государственный университет имени А.Г. и Н.Г. Столетовых».

E-mail: masmash@mail.ru

Адрес: 602264, Российская Федерация, Владимирская обл., г. Муром, ул. Орловская, д. 23.

Аннотация: В статье приведено исследование возможности применения нейросетевых технологий для распознавания ошибок при выполнении физических упражнений. Для этого была разработана структура системы распознавания ошибок, в основу которой положены две нейросети, свёрточная нейронная сеть для извлечения пространственных признаков из видеоданных и рекуррентная нейронная сеть архитектуры LSTM для анализа временных зависимостей движений. Экспериментальный подбор параметров нейросети для анализа ошибок при выполнении физических упражнений позволил добиться точности порядка 80% при обучении и распознавании ошибок. Проведённые исследования позволяют положить выбранные технологии в основу системы распознавания ошибок при выполнении физических упражнений, которая может выполнять следующий функционал: распознавание различных движений и поз во время выполнения упражнений; отслеживание правильности выполнения упражнений, обнаружение ошибок и неправильных техник выполнения упражнений.

Ключевые слова: математическая модель, нейросеть, рекуррентная нейронная сеть, распознавание движений, распознавание точек скелета.

Введение

В последнее время наблюдается огромный рост популярности активного образа жизни. Все больше людей занимаются спортом, чтобы улучшить своё здоровье и самочувствие. Наблюдается и востребованность индивидуализированного подхода, люди по-разному реагируют на тренировки. Не существует универсальной программы тренировок, которая подходит всем. Индивидуализированный подход к тренировкам позволяет максимизировать их эффективность.

Ещё одна важная тенденция — повышение стоимости услуг персональных тренеров, однако, не у всех людей есть возможность оплачивать услуги персонального тренера. При этом количество людей, занимающихся в спортзалах, постоянно растёт. Спортзалам нужны инновационные решения, которые помогут им привлечь и удержать клиентов.

Вариантом решения данных проблем может стать система автоматического распознавания спортивных упражнений. Данная система должна распознавать спортивные упражнения и предлагать пользователям рекомендации по улучшению техники выполнения упражнений, должна быть доступна для всех, независимо от их финансового положения. В итоге, использование такой системы может привести к улучшению техники выполнения упражнений, а также к снижению риска травм. Это, в свою очередь, позволит повысить эффективность тренировок.

Целью работы является разработка и исследование возможности применения нейросетевых технологий для распознавания ошибок при выполнении физических упражнений.

Для достижения поставленной цели, требуется выполнить следующие задачи:

- изучить предметную область исследования;

- выбрать алгоритмы обработки данных;
- разработать алгоритмы распознавания упражнений и выдачи рекомендаций;
- протестировать работу разработанных алгоритмов.

Анализ предметной области

В основе системы, которая способна распознавать правильность выполнения физических упражнений и выдавать рекомендации по безопасному выполнению этих упражнений лежит распознавание действия человека в видеопотоке в режиме реального времени.

В статье [1] описаны два основных подхода к распознаванию действий:

- подход, основанный на дескрипторах, которые описывают движение с точки зрения набора функций и функциональных зависимостей. Затем функции классифицируются различными методами, например, методом опорных векторов, методами внешнего вида, а также с помощью локального двоичного шаблона, нечеткой логики или даже с помощью нейросетей. Одним из примеров может служить метод распознавания оптического потока Лукаса — Канаде, алгоритм которого рассмотрен в статье А.О. Кутергина [2]. Метод использует четыре характеристики движения: скорость, траектория, угловая скорость, ускорение. При этом классификацию действий по обнаруженным характеристикам выполняет нейросеть типа «неокогнитрон». Особенности применения таких нейросетей рассмотрены в статьях [3, 4]. В статье [5] рассмотрены нейросети, работа которых основана на методе опорных векторов, которые успешно используются для классификации изображений.

- методы с применением глубокого обучения, к которым авторы [1] относят:

1) генеративные/неконтролируемые модели [6, 7] (например, Deep Belief Networks (DBNs), машины Deep Boltzmann (DBM), Restricted Boltzmann Машины (RBM) и регуляризованные автокодеры). DBN используются для распознавания рукописного текста и анализа ге-

номных данных. DBM применяются в задачах моделирования изображений и анализа естественного языка. RBM используются для рекомендательных систем и восстановления пропущенных данных;

2) дискриминационные/контролируемые модели (например, глубокие нейронные сети (DNN) [8], рекуррентные нейронные сети (RNN) [9] и свёрточные нейронные сети (CNN) [10]). DNN применяются в задачах классификации изображений и речи, а также в играх (например, в AlphaGo). RNN используются для машинного перевода, распознавания речи и генерации текста. CNN применяются в задачах классификации изображений, распознавания объектов, сегментации изображений и классификации видео;

3) гибридные модели.

Первые методы, хотя и менее распространены в задачах обработки изображений и видео по сравнению с нейронными сетями, также имеют свои преимущества и могут быть эффективны в различных сценариях. Однако следует отметить, что нейронные сети часто демонстрируют лучшие результаты во многих приложениях из-за их способности извлекать сложные признаки из данных. В задачах распознавания действий зачастую используются несколько нейронных сетей, образуя так называемую гибридную нейронную сеть.

В задачах распознавания чаще всего используют несколько типов нейронных сетей:

1) Свёрточные нейронные сети [8]. Свёрточные нейронные сети (Convolutional Neural Networks, CNN) широко используются в задачах компьютерного зрения. Они используют свёрточные слои для выделения пространственных особенностей данных. Применения:

- классификация изображений и распознавание объектов,
- сегментация изображений,
- генеративные задачи (например, синтез изображений),
- классификация видео.

2) Рекуррентные нейронные сети [11]. Рекуррентные нейронные сети (Recurrent Neural

Networks, RNN) имеют циклические связи, что позволяет моделировать последовательные и временные зависимости. Их усовершенствованные варианты, такие, как LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit), решают проблему затухания градиентов.

Основные применения:

- обработка и генерация последовательных данных (тексты, временные ряды, аудио, видео),
- машинный перевод,
- речевое распознавание,
- прогнозирование последовательных данных.

3) Графовые нейронные сети [12]. Графовые нейронные сети (Graph Neural Networks, GNN) используют графы для представления данных и обрабатывают информацию о связях между узлами. Применяются там, где данные имеют сетевую или связанную структуру:

- социальные сети и анализ связей,
- молекулярная биология и предсказание свойств молекул,
- рекомендательные системы.

4) Трансформеры [13]. Трансформеры (Transformers) — архитектура, основанная на механизме внимания (attention mechanism). Они позволяют обрабатывать большие последовательности данных без использования рекуррентных связей. Основные сферы применения:

- обработка естественного языка (NLP), включая машинный перевод, обобщение текста, генерацию и диалоговые системы,
- компьютерное зрение (Vision Transformers),
- аудиообработка и синтез речи.

Каждая из этих сетей имеет свои сильные стороны и подходит для определённых типов задач, которые зависят от структуры и характера данных. На основе этих пяти классов также создаются многочисленные гибриды, и другие малозначительно изменённые архитектуры.

Среди подходов к распознаванию движений человека выделяют ещё два подхода:

– распознавание пятна движения [14, 15]. Анализируют изменения интенсивности пикселей в видеопотоке для обнаружения движений и их классификации. Эти методы используются в системах видеонаблюдения и безопасности,

– распознавание точек скелета для идентификации ключевых точек человеческого тела и их движений [10, 16–18]. Такие системы широко применяются в спортивных аналитиках и медицинской реабилитации.

Для задач распознавания ошибок в физических упражнениях была выбрана следующая комбинация нейронных сетей:

1) свёрточная нейронная сеть (CNN) для извлечения пространственных признаков из видеоданных,

2) рекуррентная нейронная сеть специальной архитектуры LSTM (RNN-LSTM) для анализа временных зависимостей движений.

Это позволяет реализовать определённый функционал системы:

1) распознавание упражнений: система должна быть способна распознавать спортивные упражнения на основе видео данных; система должна обеспечивать оптимальную точность распознавания различных движений и поз во время упражнений;

2) контроль выполнения упражнений: система должна отслеживать правильность выполнения упражнений и предоставлять обратную связь пользователю в реальном времени; должна обнаруживать ошибки и неправильные техники выполнения упражнений и указывать пользователю на них.

Разработка алгоритмов распознавания ошибок

Для успешной реализации системы и проверки её работоспособности необходимо разработать математическую модель системы и её модулей, которая может быть описана алгоритмически. На основе данного вывода была разработана структурная схема системы, приведённая на рис. 1.

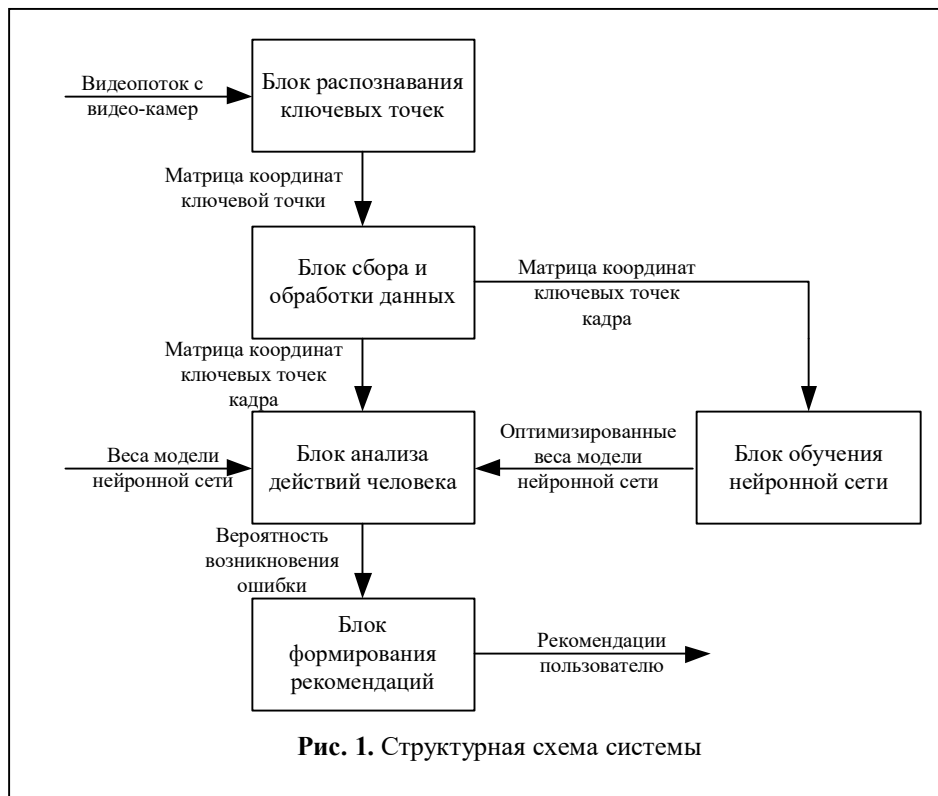


Рис. 1. Структурная схема системы

держашую встроенную и заранее обученную свёрточную нейронную сеть, обладающую высокой точностью и скоростью. Сеть анализирует каждый кадр для выявления характерных точек тела. Обычно выделяются 33 ключевые точки на теле человека, которые охватывают основные суставы и части тела, такие как плечи, локти, запястья, бедра, колени и лодыжки. Эти точки используются для различных задач в

Для дальнейшего описания математической модели системы введём обозначения: X_1 — видеопоток неопределённого разрешения, частотой от 30 кадров в секунду; X_2 — предварительные настройки нейронной сети распознавания действий; Z_1 — неупорядоченный набор данных, состоящий из значений координат по трём осям для каждой ключевой точки скелета; Z_2 — матрица обработанных и упорядоченных данных о ключевых точках; Z_3 — вероятности нахождения ошибки в упражнении; Z_4 — полученные в ходе обучения веса нейронной сети; Y — рекомендации исходя из вероятности нахождения ошибки в упражнении.

Для корректной работы системы были разработаны несколько алгоритмов:

- 1) алгоритм локализации ключевых точек скелета,
- 2) алгоритм формирования рекомендаций.

Для распознавания ключевых точек скелета человека из видеопотока было принято решение использовать библиотеку MediaPipe, со-

компьютерном зрении и анализе движений, включая отслеживание позы и анализ кинематики. Однако для более точного и детализированного анализа рекомендуется использовать дополнительные точки распознавания. Эти точки охватывают не только контуры лица, но и мелкие детали, такие как губы, глаза и нос. Такой детализированный подход позволяет моделям более точно анализировать выражения лица, мимику и даже небольшие движения. Координаты каждой точки представляют собой тройку (x, y, z) для трёхмерного анализа.

Алгоритм локализации ключевых точек скелета:

1) нейронная сеть получает кадр X_1 в качестве входных данных и производит оценку координат ключевых точек,

2) кадр X_1 проходит через свёрточную нейронную сеть, встроенную в библиотеку MediaPipe, обученную обнаруживать ключевые точки тела,

3) для каждой из n ключевых точек p_k сеть прогнозирует координаты:

$$p_k = (x_k, y_k, z_k), k \in [0, n], \quad (1)$$

где (x_k, y_k, z_k) — нормализованные координаты точки относительно размеров кадра (в диапазоне от 0 до 1); p_k — ключевая точка скелета человека; n — общее количество ключевых точек.

Блок сбора и обработки данных выполняет сбор ключевых точек для каждого кадра X_1 , распознанного блоком. Полученные координаты объединяются в матрицу Z_2 :

$$Z_2 = \{p_1, p_2, \dots, p_n\}, \quad (2)$$

где Z_2 — матрица с координатами всех ключевых точек данного кадра; p_1 — матрица с координатами определённой ключевой точки.

После этого, в зависимости от наличия весов для работы нейронной сети, матрица с координатами (2) направляется либо в блок обучения, либо в блок распознавания.

В случае отсутствия промежуточных весов для работы блока распознавания, активируется блок, ответственный за обучение нейронных сетей. Обучение модели нейронной сети — это процесс подбора оптимальных весов Z_4 для каждого нейрона сети. Алгоритм работы блока обучения выглядит следующим образом:

1) выбирается количество видеозаписей k и количество кадров, содержащееся в каждой видеозаписи u . После получения потока Z_2 матрицы всех кадров видеоданных совмещаются и группируются как по кадрам, так и по видео:

$$E_k = \{Z_{21}, Z_{22}, \dots, Z_{2u}\}, \quad (3)$$

где E_k — совмещённая матрица кадров, характеризующая видеозапись.

2) Наборы кадров объединяются в одну структуру, формируя набор данных:

$$D = \{E_1, E_2, \dots, E_k\}, \quad (4)$$

где D — набор данных из видеозаписей.

3) Для поиска оптимальных весов используется метод обратного распространения ошибки [19]:

$$\bar{Y} = f(D, W), \quad (5)$$

где \bar{Y} — предсказанные значения; f — одна из функций активации; W — веса обучаемого слоя.

4) Выбираются функции активации. Для LSTM был выбран гиперболический тангенс \tanh :

$$f = \tanh D = \frac{e^D - e^{-D}}{e^D + e^{-D}}, \quad (6)$$

где e^D — набор данных из видеозаписей в экспоненциальной функции.

5) Для полносвязных слоёв была использована функция ReLU, которая преобразует входное значение таким образом, что все отрицательные значения становятся нулевыми, а положительные остаются без изменений:

$$f = \text{ReLU}(D) = \max(0, D), \quad (7)$$

где $\max(0, D)$ — пропуск всех отрицательных значений набора данных.

6) На выходе полносвязной сети используется нормализующая функция, выдающая вероятность отношения к какому-либо классу для классификации в пределах $[0..1]$:

$$f = \text{SOFTMAX}(D_i) = \frac{e^{D_i}}{\sum_{j=1}^k e^{D_j}}. \quad (8)$$

7) Вычисляется градиент функции потерь и происходит обновление весов:

$$Z_4 = W_t - f \cdot \frac{dL}{dW_t}, \quad (9)$$

где Z_4 — конечные веса модели; W_t — временные веса; $\frac{dL}{dW_t}$ — градиент функции потерь.

Процесс продолжается до тех пор, пока функция потерь, которая измеряет величину ошибки при текущих значениях весов, не будет минимизирована. Функция потерь представляет собой количественное измерение расхождения между предсказанными значениями модели и истинными значениями:

$$L(Y, \bar{Y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C Y_{i,c} \log(\bar{Y}_{i,c}), \quad (10)$$

где Y — истинные значения; \bar{Y} — предсказанные значения; N — количество примеров; C — количество классов.

8) Во время обучения, для оценки, используется стандартная метрика точности:

$$A = \frac{a'}{a}, \quad (11)$$

где a' — количество правильных предсказаний; a — количество всех ответов.

Итогом работы блока станут веса Z_4 для работы модели.

Блок анализа действий человека является модулем распознавания ошибок при выполнении упражнений в реальном времени. После получения настроек Z_4 либо X_2 и данных Z_2 , начинается процесс анализа, с применением рекуррентной, а после полносвязной нейронной сети.

В ходе исследования было проведено обучение модели нейронной сети, которое заклю-

чается в подборе метрик точности и качества обучения, функций активации, оптимизатора, оптимальной архитектуры модели нейронной сети и экспериментов с разными комбинациями данных. В таблице 1 представлены комбинации обучения, показывающие лучшую точность и стабильность работы. Используемые модули для обучения: оптимизаторы — Adam, SGD; функции активации — RELU, Softmax, tanh. Слои архитектуры — LSTM, полносвязные, GRU. Метрики точности — Categorical Accuracy (CA).

Для обучения нейросети был использован набор из 1800 видеозаписей, каждая видеозапись длиной 30 кадров с разрешением от 240 пикселей.

Наибольшая точность и стабильность работы были достигнуты с использованием комбинации №2: функция активации Tanh и RELU, оптимизатор обучения ADAM, архитектура модели с 3 LSTM слоями, 3 полносвязными

Таблица 1. Варианты обучения нейронной сети

Номер	Метрика	Функция активации	Оптимизатор	Архитектура	Количество нейронов на слоях	Кадры	Точность (CA)
1	CA	RELU, SOFTMAX	ADAM	3x LSTM 3x DENSE	64,128,64 64,32,2	30	69%
2	CA	Tanh, RELU, SOFTMAX	ADAM	3x LSTM 3x DENSE	64,128,64 64,32,2	30	89%
3	CA	RELU, SOFTMAX	ADAM	3x LSTM 3x DENSE	64,128,64 64,32,2	60	—
4	CA	RELU, SOFTMAX	ADAM, SGD	3x LSTM 3x DENSE	256, 512, 256 256, 128, 64, 32, 2	60	—
5	CA	Tanh, RELU, SOFTMAX	SGD 0.9	6x GRU 3x DENSE	256, 512, 256 256, 128, 64, 32, 2	60	50%
6	CA	Tanh, RELU, SOFTMAX	SDG 0.9	6x LSTM 3x DENSE	256, 512, 256 256, 128, 64, 32, 2	60	47%
7	CA	Tanh, RELU, SOFTMAX	SGD 0.9	8x GRU 3x DENSE	256, 512, 256 256, 128, 64, 32, 2	60	—

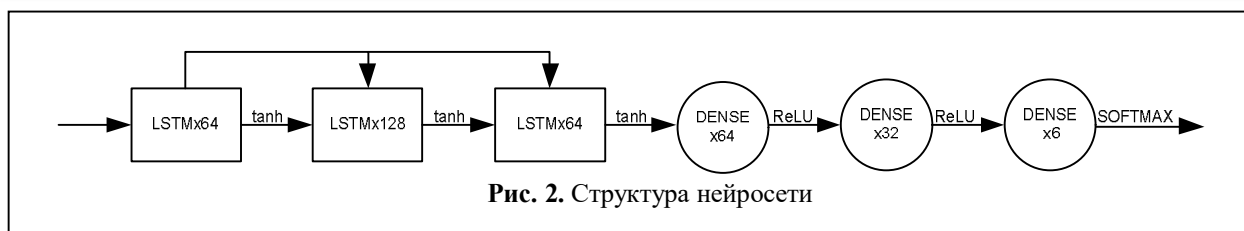


Рис. 2. Структура нейросети

слоями и количеством нейронов (64, 128, 64) в LSTM слоях и (64, 32, 2) в полносвязных слоях соответственно. Эта комбинация позволила достичь высокой точности распознавания на уровне 89%. Проведён эксперимент по увеличению длины последовательности в 2 раза, до 60 кадров, что в будущем позволит обрабатывать более сложные спортивные упражнения, однако ни один из вариантов архитектуры нейронной сети не достиг точности более 50 процентов, либо не прошёл обучения в целом. Необходимо большее количество экспериментов с архитектурой и больший набор данных.

Таким образом, в результате исследования была выбрана нейронная сеть, состоящая из трёх рекуррентных LSTM слоёв, а также трёх полносвязных слоёв DENSE. Здесь рекуррентные слои выбраны из-за своей способности эффективно анализировать последовательности данных путём сохранения контекста набора данных. В работе LSTM слоёв присутствуют специальные механизмы эффективного управления памятью — вентили, которые позволяют эффективно управлять данными на момент обработки внутри сети, что позволяет не только выполнять анализ последовательности, учитывая контекст данных во времени, но и отсеивать ненужные детали во время анализа. Это позволяет оптимизировать работу данных слоев, а также повысить точность и скорость работы [20]. Полносвязные слои DENSE являются дополнением и помогают классифицировать ошибку в спортивном упражнении. Структура нейросети приведена на рис. 2.

Итоговым значением работы последнего слоя Z_3 будет являться интерпретированная вероятность отношения полученных данных к одному из классов в области $[0; 1]$, которая поступает на вход блока формирования реко-

мендаций. Блок анализирует вероятности классификации, используя определённый порог. При значении вероятности отношения к классу больше 0,7, блок выбирает рекомендацию в текстовом виде, заранее определённую для каждого из имеющихся трёх классов. Алгоритм формирования рекомендации:

- 1) получение интерпретированной вероятности Z_3 отношения полученных данных к одному из классов в области $[0...1]$;
- 2) выбор рекомендации: рекомендация формируется, если значения вероятности отношения к классу больше THR:

$$Y = Z_3 > THR, \quad (12)$$

где Y — получаемая рекомендация; Z_3 — вероятность отношения ответа к одному из классов; THR — изменяемый порог вероятности.

В ходе исследования было принято решение использовать упражнение с большим количеством ошибок, приводящим к серьёзным травмам — жим штанги на бицепс стоя. В результате анализа техники выполнения упражнения было выделено три класса возможных ошибок, которые характеризуют уровень поднятии штанги, помощь мышц спины и коленей, а также скорость выполнения повторов упражнения.

В дальнейшем в режиме реального времени рекомендация при обнаружении ошибки выводится на экран для пользователя.

Экспериментальный анализ

Для оценки корректности работы нейросетей было проведено их тестирование на студентах-добровольцах. В рамках эксперимента все студенты были разбиты на три схожие по физическим параметрам (рост, вес, телосложение) группы. Эксперименты выполнялись при различном положении людей от камеры, что поз-

Таблица 4. Тестирование при разной освещённости помещения

Номер эксперимента	Ошибка упражнения	Освещённость, люкс	Точность распознавания ошибки для добровольцев, %		
			1	2	3
1	Непосильный вес штанги (изгиб спины)	268	88	85	86
2		357	86	84	86
3		160	77	78	75
4	Слишком медленное выполнение подхода	268	87	86	84
5		357	87	82	83
6		160	78	72	72
7	Поднятие штанги не на уровень груди (выше)	268	89	86	86
8		357	88	78	79
9		160	81	77	77

Таблица 3. Тестирование после выполнения масштабирования

Номер эксперимента	Ошибка упражнения	Расстояние от камеры, м	Точность распознавания ошибки для добровольца %		
			1	2	3
1	Непосильный вес штанги (изгиб спины)	2	84	83	81
2		3	83	80	77
3		6	81	82	78
4	Слишком медленное выполнение подхода	3	86	88	85
5		2	82	85	84
6		6	80	79	85
7	Поднятие штанги не на уровень груди (выше)	3	85	83	82
8		2	85	83	77
9		6	84	83	79

волило выявить критическую зависимость положения человека для распознавания с использованием нейросетей (таблица 2). Для устранения этой проблемы было выполнено масштабирование видеозаписей, результаты исследования после масштабирования приведены в таблице 3.

Результаты показали, что точность распознавания движений варьируется в зависимости от условий съёмки. Наиболее высокие показатели точности в 89% были получены при средней освещённости и средней дистанции от камеры. Изменение этих параметров, как в сторону увеличения, так и в сторону уменьшения, снижало точность распознавания. Так, например, при изменении положения относительно камеры максимальная точность распознавания была достигнута на расстоянии 3 метров, в то время как на расстоянии 6 метров точность существенно снижалась. При этом масштаби-

рование входных видеофайлов позволило решить данную проблему и снизить зависимость от положения человека.

В таблице 4 представлена зависимость точности распознавания ошибки в упражнении при различной освещённости помещения, в котором проводились эксперименты. Исследование показало слабую зависимость от освещённости.

При изменении освещённости наиболее высокая точность была зафиксирована при освещённости в 268 люкс, тогда как при понижении освещённости до 160 люкс точность уменьшалась.

Таким образом, на основе проведённых испытаний можно сделать вывод, что разработанное программное обеспечение обладает достаточной точностью для использования в условиях средней освещённости и средней дистанции от камеры, однако для повышения его

эффективности рекомендуется учитывать влияние внешних факторов и при необходимости проводить дополнительные калибровки и настройки.

Заключение

В результате исследований была разработана система компьютерного зрения позволяющая выполнять распознавание ошибок при выполнении спортивных упражнений в режиме реального времени. Проведённые исследования позволили сформировать структуру системы, которая основана на двух нейросетях. Для распознавания ключевых точек скелета человека из видеопотока была использована библиотека MediaPipe, содержащая встроенную и заранее обученную свёрточную нейронную сеть, обладающую высокой точностью и скоростью. Для распознавания и классификации ошибок при выполнении упражнений была выбрана и исследована нейронная сеть, состоящая из трёх рекуррентных LSTM слоев, а также трёх полносвязных слоев DENSE. Эта комбинация позволила достичь высокой точности распознавания на уровне 89%.

Проведённые исследования позволят положить выбранные технологии в основы системы распознавания ошибок при выполнении физических упражнений.

Литература

1. Азаренко К.А., Каунг М.Н., Белов Ю.С. Обзор методов для распознавания действий человека [Электронный ресурс]: E-Scio. 2019. №5 (32). URL: http://e-scio.ru/wp-content/uploads/2019/05/E-SCIO-5_2019.pdf (дата обращения 04.07.2024).
2. Кутергин А.О. Алгоритм решения задачи распознавания действий объекта на видеопотоке [Электронный ресурс]: ВК. 2013. №12. URL: <http://ipos-tmn.ru/wp-content/uploads/2015/12/12-111-117.pdf> (дата обращения 04.07.2024).
3. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position // Biological Cybernetics. 1980. Vol. 36(4). Pp. 193–202.
4. Shouno, H., & Fukushima, K. Neocognitron capable of incremental learning // Neural Networks. 2004. Vol. 12(7–8). Pp. 1171–1179.
5. Cortes C., Vapnik V. Support-vector networks // Machine learning. 1995. Vol. 20(3). Pp. 273–297.

6. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B. et al. Generative adversarial nets // Advances in neural information processing systems. 2014. Vol. 27. Pp. 1–9.
7. Rezende D.J., Mohamed S., Wierstra D. Stochastic backpropagation and approximate inference in deep generative models // International Conference on Machine Learning. 2014. Pp. 1278–1286.
8. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks // Advances in neural information processing systems. 2012. Vol. 25. DOI:10.1145/3065386
9. Буйко А.Ю., Виноградов А.Н. Выявление действий на видео с помощью рекуррентных нейронных сетей // Программные системы: теория и приложения. 2017. №4(35). С. 327–345.
10. Подопросветов А.В., Лузина М.Е., Алисейчик А.П., Павловский Е.В., Орлов И.А. Применение нейронных сетей для предсказания физических действий человека // Препринты ИПМ им. М.В.Келдыша. 2021. № 109. 16 с.
11. Hochreiter, S., Schmidhuber, J. Long short-term memory // Neural computation. 1997. Vol. 9(8). Pp. 1735–1780.
12. Kipf T. N., Welling M. Semi-supervised classification with graph convolutional networks [Электронный ресурс]: International Conference on Learning Representations. 2017. URL: <https://arxiv.org/abs/1609.02907> (дата обращения: 04.07.2024).
13. Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need [Электронный ресурс]: Advances in neural information processing systems. Vol. 30. Pp. 5998–6008. 2017. URL: <https://arxiv.org/abs/1706.03762> (дата обращения: 04.07.2024).
14. Wang H., Schmid C. Action recognition with improved trajectories // Proceedings of the IEEE international conference on computer vision. 2013. Pp. 3551–3558.
15. Simonyan K., Zisserman A. Two-stream convolutional networks for action recognition in videos // Advances in neural information processing systems. 2014. Vol. 27. Pp. 1–9.
16. Shotton J., Fitzgibbon A., Cook M., et al. Real-time human pose recognition in parts from single depth images // CVPR. 2011. Pp. 1297–1304.
17. Cao Z., Simon T., Wei S. E., et al. Realtime multi-person 2d pose estimation using part affinity fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. Pp. 7291–7299.
18. Sun X., Shang J., Liang S., et al. Compositional human pose regression // Proceedings of the IEEE International Conference on Computer Vision. 2017. Pp. 2602–2611.
19. Созыкин А.В. Обзор методов обучения глубоких нейронных сетей // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6. № 3. С. 28–59.

20. Дауб И.С. Обзор методов прогнозирования временных рядов с помощью искусственных нейронных сетей [Электронный ресурс]: StudNet. 2020. №10. URL: <https://cyberleninka.ru/article/n/>

[obzor-metodov-prognozirovaniya-vremennyh-ryadov-s-pomoschyu-iskusstvennyh-neyronnyh-setey](#) (дата обращения: 04.07.2024).

Поступила 4 июля 2024 г.

English

POSSIBLE NEURAL-NETWORK USE FOR ERROR RECOGNITION DURING PHYSICAL EXERCISING

Dmitry Yuryevich Titarenko — Master's degree Student; Department of Physics and Applied Mathematics, Murom Institute (branch) "Vladimir State University named after A.G. and N.G. Stoletovs".

Maria Nikolayevna Ryzhkova — PhD, Associate Professor, the Dean of the Faculty of Information Technologies and Radioelectronics, Murom Institute (branch) "Vladimir State University named after A.G. and N.G. Stoletovs".

E-mail: masmash@mail.ru

Address: 602264, Russian Federation, Vladimir region, Murom, Orlovskaya st., 23.

Abstract: The article examines possible neural-network use for error recognition during physical exercising. The aim of the work was to develop and study the possibility of using neural network technologies for error recognition during physical exercising. The system is based on recognizing person's activities in a real-time video stream and it can verify correct physical exercising and give recommendations for the safe performance of these exercises. The article describes the primary approaches to identifying human actions using neural networks. Based on the considered approaches there was developed system structure and were defined system units, which should not only recognize an error in the exercise, but also render recommendations when it is detected. The algorithm of skeleton key points' localization and recommendation-rendering algorithm were developed for this purpose. The certain combination was selected to solve the error recognition problem in physical exercises and it comprises convolutional neural network (CNN) to extract spatial features from video data and LSTM recurrent neural network (RNN-LSTM) for the analysis of movement time dependencies. MediaPipe library was used to recognize human skeleton keypoints from the video stream as it contains incorporated and pre-trained convolutional neural network with high accuracy and speed. The neural network consisting of three recurrent LSTM layers as well as three fully connected DENSE layers was selected and studied to identify and classify errors during exercises. This combination enabled to achieve high recognition accuracy at 89% level. Neural networks were tested on volunteer students to assess their correct operation. All students were split into three groups similar in physical parameters (height, weight, bodybuild) as part of the experiment. The results showed that movement recognition accuracy varies with video shooting situation. The highest accuracy figures of 89% were obtained at average illumination and average distance from the camera. Changing these parameters reduced the recognition accuracy. Scaling the input video files helped to solve this problem and reduce dependence on the person's position. The performed research will enable to accept the selected technologies as the basis for error recognition system during physical exercising.

Keywords: mathematical model, neural network, recurrent neural network, movement recognition, skeleton point recognition.

References

1. Azarenko K.A., Kaung M.N., Belov Yu.S. Review of methods for recognizing human actions [Electronic source]: E-Scio. 2019. No.5 (32). URL: http://e-scio.ru/wp-content/uploads/2019/05/E-SCIO-5_2019.pdf (access date 04.07.2024).
2. Kutergin A.O. Algorithm of solving a problem to recognize actions of object at the video stream [Electronic source]: Proceedings of Cybernetics. 2013. No. 12. URL: <http://ipos-tmn.ru/wp-content/uploads/2015/12/12-111-117.pdf> (access date 04.07.2024).
3. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics. 1980. Vol. 36(4). Pp. 193–202.

4. Shouno H., Fukushima K. Neocognitron capable of incremental learning. *Neural Networks*. 2004. Vol. 12(7–8). Pp. 1171–1179.
5. Cortes C., Vapnik V. Support-vector networks. *Machine learning*. 1995. Vol. 20(3). Pp. 273–297.
6. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B. et al. Generative adversarial nets. *Advances in neural information processing systems*. 2014. Vol. 27. Pp. 1–9.
7. Rezende D.J., Mohamed S., Wierstra D. Stochastic backpropagation and approximate inference in deep generative models. *International Conference on Machine Learning*. 2014. Pp. 1278–1286.
8. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012. Vol. 25. DOI:10.1145/3065386
9. Buyko A., Vinogradov A. Action recognition on video using recurrent neural networks. *Program systems: Theory and applications*. 2017. No. 4(35). Pp. 327–345.
10. Podoprosvetov A.V., Luzina M.E., Pavlovsky E.V., Aliseychik A.P., Orlov I.A.. Neural networks for predicting human actions. *KIAM Preprint*. 2021. No. 109. 16 p.
11. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural computation*. 1997. Vol. 9(8). Pp. 1735–1780.
12. Kipf T. N., Welling M. Semi-supervised classification with graph convolutional networks [Electronic source]: *International Conference on Learning Representations*. 2017. URL: <https://arxiv.org/abs/1609.02907> (access date: 04.07.2024).].
13. Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need [Electronic source]: *Advances in neural information processing systems*. Vol. 30. Pp. 5998-6008. 2017. URL: <https://arxiv.org/abs/1706.03762> (access date: 04.07.2024).
14. Wang H., Schmid C. Action recognition with improved trajectories. *Proceedings of the IEEE international conference on computer vision*. 2013. Pp. 3551–3558.
15. Simonyan K., Zisserman A. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*. 2014. Vol. 27. Pp. 1–9.
16. Shotton J., Fitzgibbon A., Cook M., et al. Real-time human pose recognition in parts from single depth images. *CVPR*. 2011. Pp. 1297–1304.
17. Cao Z., Simon T., Wei S. E., et al. Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. Pp. 7291–7299.
18. Sun X., Shang J., Liang S., et al. Compositional human pose regression. *Proceedings of the IEEE International Conference on Computer Vision*. 2017. Pp. 2602–2611.
19. Sozykin A.V. An Overview of Methods for Deep Learning in Neural Networks.. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. Vol. 6. No. 3. Pp. 28–59.
20. Daub I.S. Review of methods for forecasting time series using artificial neural networks [Electronic source]: *StudNet*. 2020. No. 10. URL: <https://cyberleninka.ru/article/n/obzor-metodov-prognozirovaniya-vremennyh-ryadov-s-pomoschyu-iskusstvennyh-neyronnyh-setey> (access date: 04.07.2024).