Оригинальная статья / Original article

https://doi.org/10.21869/2223-1560-2020-24-4-57-75

Идентификация человека по походке в видеопотоке

М.Ю. Уздяев [1], Р.Н. Яковлев [1] ✉, Д.М. Дударенко [1], А.Д. Жебрун [1]

[1] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук,
Санкт-Петербургский институт информатики и автоматизации Российской академии наук
14-я линия В.О., 39, г. Санкт-Петербург 199178, Российская Федерация

✉ e-mail: iakovlev.r@mail.ru

Резюме

Цель исследования. *Данная работа посвящена проблеме идентификации человека по походке с помощью нейросетевых моделей распознавания, ориентированных на работу с RGB изображениями. Главным преимуществом использования нейросетевых моделей перед существующими методами анализа двигательной активности является получение изображений из видеопотока без предобработки кадров, увеличивающей время анализа.*

Методы*. В данной работе был предложен подход к идентификации человека по походке, который основывается на идее многоклассовой классификации на видеопоследовательностях. Оценка качества функционирования разработанного подхода производилась на основе набора данных CASIA Gait Database, включающего в себя более 15000 видеопоследовательностей. В качестве классификаторов были апробированы 5 нейросетевых артитектур: трехмерная сверточная нейронная сеть I3D, а также 4 архитектуры, представляющие собой сверточно-рекуррентные сети, такие, как однонаправленная и двунаправленная LTSM, однонаправленная и двунаправленная GRU, скомбинированные со сверточной нейронной сетью архитектуры ResNet, используемой в данных архитектурах в качестве экстрактора визуальных признаков.*

Результаты. *Согласно результатам проведенного тестирования, разработанный подход предоставляет возможность осуществлять идентификацию человека в видеопотоке в режиме реального времени без использования специализированного оборудования. По результатам его апробации с помощью рассматриваемых нейросетевых моделей точность идентификации человека составила более 80% для сверточно-рекуррентных моделей и 79% для модели I3D.*

Заключение. *Предложенные модели на основе архитектуры I3D и сверточно-рекуррентных архитектур показали более высокую точность, чем существующие методы решения задачи идентификации человека по походке. За счет возможности покадровой обработки видео наиболее предпочтительным классификатором для разработанного подхода является использование сверточно-рекуррентных архитектур на основе однонаправленной LSTM или GRU моделей соответственно.*

Ключевые слова: *нейронные сети; компьютерное зрение; сверточные нейронные сети; рекуррентные нейронные сети; I3D; методы идентификации человека.*

Конфликт интересов: *Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.*

Для цитирования: Идентификация человека по походке в видеопотоке / М.Ю. Уздяев, Р.Н. Яковлев, Д.М. Дударенко, А.Д. Жебрун // Известия Юго-Западного государственного университета. 2020; 24(4): 57-75. https://doi.org/10.21869/2223-1560-2020-24-4-57-75.

Поступила в редакцию 16.09.2020 *Подписана в печать 20.10.2020* *Опубликована 30.12.2020*

Identification of a Person by Gait in a Video Stream

Mikhail Yu. Uzdiaev [1], Roman N. Iakovlev [1], Dmitry M. Dudarenko [1], Aleksandr D. Zhebrun [1]

[1] St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14th Line, St. Petersburg 199178, Russian Federation

✉ e-mail: iakovlev.r@mail.ru

Abstract

Purpose of research. *The given paper considers the problem of identifying a person by gait through the use of neural network recognition models focused on working with RGB images. The main advantage of using neural network models over existing methods of motor activity analysis is obtaining images from the video stream without frames preprocessing, which increases the analysis time.*

Methods. *The present paper presents an approach to identifying a person by gait. The approach is based upon the idea of multi-class classification on video sequences. The quality of the developed approach operation was evaluated on the basis of CASIA Gait Database data set, which includes more than 15,000 video sequences. As classifiers, 5 neural network architectures have been tested: the three-dimensional convolutional neural network I3D, as well as 4 architectures representing convolutional-recurrent networks, such as unidirectional and bidirectional LTSM, unidirectional and bidirectional GRU, combined with the convolutional neural network of ResNet architecture being used in these architectures as a visual feature extractor.*

Results. *According to the results of the conducted testing, the developed approach makes it possible to identify a person in a video stream in real-time mode without the use of specialized equipment. According to the results of its testing and through the use of the neural network models under consideration, the accuracy of human identification was more than 80% for convolutional-recurrent models and 79% for the I3D model.*

Conclusion. *The suggested models based on I3D architecture and convolutional-recurrent architectures have shown higher accuracy for solving the problem of identifying a person by gait than existing methods. Due to the possibility of frame-by-frame video processing, the most preferred classifier for the developed approach is the use of convolutional-recurrent architectures based on unidirectional LSTM or GRU models, respectively.*

Keywords: *neural networks; computer vision; convolutional neural networks; recurrent neural networks; I3D; human identification techniques.*

Conflict of interest. *The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.*

For citation: Uzdiaev M. Yu., Iakovlev R. N., Dudarenko D. M., Zhebrun A. D Identification of a Person by Gait in a video stream. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2020; 24(4): 57-75 (In Russ.). https://doi.org/10.21869/2223-1560-2020-24-4-57-75.

Received 16.09.2020 *Accepted 20.10.2020* *Published 30.12.2020*

***

Введение

При разработке систем охраны правопорядка, контроля доступа, а также киберфизических систем и интеллектуальных пространств, задача идентификации человека не теряет своей актуальности [1]. Походка является одним из поведенческих проявлений, по которому возможна идентификация человека на расстоянии без непосредственного с ним контакта. При этом данный биометрический параметр является стабильным и слабо подвержен изменениям, а кроме того, остается наблюдаемым в ситуациях, когда идентификацию человека невозможно выполнить на основе анализа таких распространенных типов биометрических данных, как изображение лица, голос человека, радужная оболочка глаза и отпечатки пальцев ввиду недостатка соответствующей информации [2]. Многие существующие методы анализа двигательной активности для выделения информации осуществляют сложную предобработку кадров, например, локализацию ключевых точек скелета человека [3], анализ изображений энергии походки, бинарного силуэта человека [4]. Соответствующая предобработка видеоданных замедляет процесс распознавания человека, а также требует выполнения предварительной корректировки кадров видеоряда. Альтернативным решением данной задачи является использование нейросетевых моделей распознавания, ориентированных на работу с RGB изображениями, полученными из видеопотока, без предобработки. В данной работе предложен метод распознавания человека по походке, где в качестве основы рассматриваются сверточно-рекуррентные и трехмерные нейросетевые архитектуры, которые не требуют предварительной обработки изображений или видеопотока и при этом позволяют достичь высокой точности распознавания.

Обзор методов идентификации человека по походке

В последние годы было разработано множество нейросетевых методов идентификации человека по походке, которые отличаются как технически [5]: архитектурами сетей, функциями потерь, способами обучения, так и концептуально: методами обработки данных и извлечения первичных признаков [6]. Большинство существующих методов классифицируют видео не напрямую по кадрам, а осуществляют анализ различных динамических характеристик походки, с помощью которых идентифицируется человек. Это связано с тем, что, при разной одежде, наличии различных вещей, например, сумки, а также при смене освещения, фигура и образ человека подвержены существенным изменениям, поэтому необходимо, чтобы система опиралась не на внешние признаки, а отталкивалась непосредственно от характеристик движения фигуры человека. Многие современные подходы к решению рассматриваемой задачи основаны как на анализе таких биометрических характеристик (человеческий

скелет, силуэт и их изменение при ходьбе), так и на признаках, получаемых при использовании методов машинного обучения в результате анализа биометрических данных с помощью сверточных нейронных сетей [7].

Среди существующих методов можно выделить несколько базовых подходов. К ним относятся подходы, использующие анализ человеческого скелета и признаки, сконструированные вручную. Соответствующие методы распознавания основаны на изучении осанки человека, положения суставов и основных частей тела и их движениях при ходьбе [3]. К базовым также могут быть отнесены подходы, связанные с бинарным силуэтом человека [8], такие как распознавание по изображениям энергии походки (Gait Energy Image, GEI) [4], по изображениям энтропии походки (Gait Entropy Image, GEnI) [6], по энергии разницы кадров (Frame Difference Energy Image, FDEI) [9], которые позволяют вычислять дальнейшие признаки, такие как гистограммы ориентированных градиентов (Histogram of Oriented Gradients, HOG-дескрипторы) или гистограммы оптического потока (Histogram of Optical Flow, HOF-дескрип-торы) [10]. Построение изображений энергии походки является одним из наиболее популярных методов, обеспечивающих идентификацию человека по походке. Изображения энергии походки представляют собой усредненные по одному циклу походки бинарные маски силуэта движущегося человека. Изображения энергии походки характеризуют частоты нахождения человека в той или иной позе во время движения. Данный базовый подход получил широкое распространение и лег в основу многих других методов распознавания человека по походке. Множество существующих подходов также основаны на схожей агрегации других базовых признаков, однако общим недостатком методов, использующих GEI для многоракурсного распознавания, является необходимость вычислять изображения энергии для каждого ракурса, присутствующего в выборке. Поэтому для каждого кадра видеоряда нужно знать, под каким углом он был снят, что в случае реальных данных возможно далеко не всегда. Для более качественного извлечения признаков можно использовать гистограммы оптического потока (HOF) и гистограммы ориентированных градиентов (HOG). Такие дескрипторы хорошо подходят для распознавания жестов [11], но требуют больших вычислительных ресурсов [12].

Более эффективный метод распознавания с использованием архитектуры LTSM и изображений энергии походки (GEI) предложен в работе [13]. Хотя пространственная информация в одной последовательности походки может быть хорошо представлена GEI, временная информация теряется. Чтобы решить эту проблему, авторы предлагают новый метод обучения для распознавания походки. Особенность метода заключается в использовании рекуррентной нейросетевой модели длительной краткосрочной памяти (Long Term Short Memory, LSTM) [14], соответствующая система может сохранять

временную информацию, повышая качество распознавания походки.

Стоит отметить, что нейросетевые методы, построенные на трехмерных сверточных архитектурах (3D CNN) [15-18], демонстрируют лучшие результаты в задачах распознавания большого количества классов действий на видео. При этом, особо стоит выделить архитектуру 3D CNN Inception 3D (I3D) [16], которая обладает широкими возможностями по обработке пространственно-временной динамики объектов на кадрах видео и показывает высокие результаты распознавания действий человека [16, 19]. Нейросетевые методы, основанные на применении сверточно-рекуррентных архитектур LTSM и Gated Recurrent Units (GRU) [20], также демонстрируют высокую точность в задачах распознавания действий человека на видеопоследовательностях [21]. При этом, модель GRU является более простой по сравнению с LTSM, обладая аналогичными LSTM достоинствами. Также стоит особо выделить, что 3D CNN и сверточно-рекуррентные архитектуры способны с высокой точностью выполнять распознавание действий на видео на основе анализа кадров без сложной предварительной их обработки, используя признаки, получаемые во время обучения нейронных сетей. Учитывая, что решения на основе таких признаков являются более универсальными и обычно не требуют сложной предварительной обработки изображений, а модели на основе 3D CNN и сверточно-рекуррентных архитектур являются при этом наиболее точными, в данной работе в качестве основы для разрабатываемого подхода к решению задачи идентификации человека по походке было принято решение апробировать 5 нейросетевых архитектур. Одной из них является трехмерная сверточная нейронная сеть I3D. Другие 4 архитектуры представляют собой сверточно-рекуррентные сети: однонаправленная LTSM (LTSM-1), двунаправленная LTSM (LTSM-2), однонаправленная GRU (GRU-1) и двунаправленная GRU (GRU-2), скомбинированные со сверточной нейронной сетью архитектуры ResNet [22], используемой в данных архитектурах в качестве экстрактора визуальных признаков. Несмотря на то, что применение связки архитектур является довольно распространенным приемом, однако на данный момент такой подход ранее не применялся для решения задачи распознавания человека по походке, и соответственно, сравнение производительности различных вариантов архитектур также ранее не проводилось. Таким образом, далее в работе будет представлен разработанный подход к идентификации человека по походке, где в качестве классификаторов будут апробированы представленные выше нейросетевые модели.

Материалы и методы

Описание разработанного подхода к идентификации человека по походке

В соответствии с результатами проведенного анализа связанных методов и подходов, для обеспечения идентификации человека по походке в рамках ис-

следования предлагается авторский подход к решению данной задачи. Согласно предложенному подходу, задача идентификации человека формулируется как задача многоклассовой классификации на видеопоследовательностях, решение которой обеспечивается за счет применения специфичной нейросетевой модели в качестве классификатора. Итоговая метка класса человека, получаемая соответствующей моделью по результатам обработки исследуемой видеопоследовательности, и представляется в качестве конечного решения задачи.

Четыре из пяти нейросетевых моделей, апробированных в качестве классификатора в настоящем исследовании, основаны на сверточно-рекуррентном принципе работы. Рекуррентные нейронные сети обычно испытывают трудности с обработкой долгосрочных зависимостей из-за затухания или резкого увеличения градиента. Для решения данной проблемы были разработаны специальные архитектуры нейронных сетей, одной из которых является LSTM-сеть. Еще одной вариацией LSTM-сетей является архитектура на основе GRU, представленная в работе [20]. В этом варианте вентиль забывания и входной вентиль объединены в один – вентиль обновления. Кроме того, объединены вместе состояние ячейки и скрытое состояние, а также присутствует ряд других второстепенных изменений. Полученная в результате модель является более простой, чем стандартные LSTM-модели, и, как следствие, в последнее время набирает все большую популярность. С учетом вышеописанных преимуществ для сравнения были выбраны оба варианта архитектуры нейронной сети и LTSM и более простая – GRU.

Обобщенная архитектура четырех сверточно-рекуррентных моделей (LSTM-1, LSTM-2, GRU-1, GRU-2), рассматриваемых в работе, представлена ниже на рис. 1.

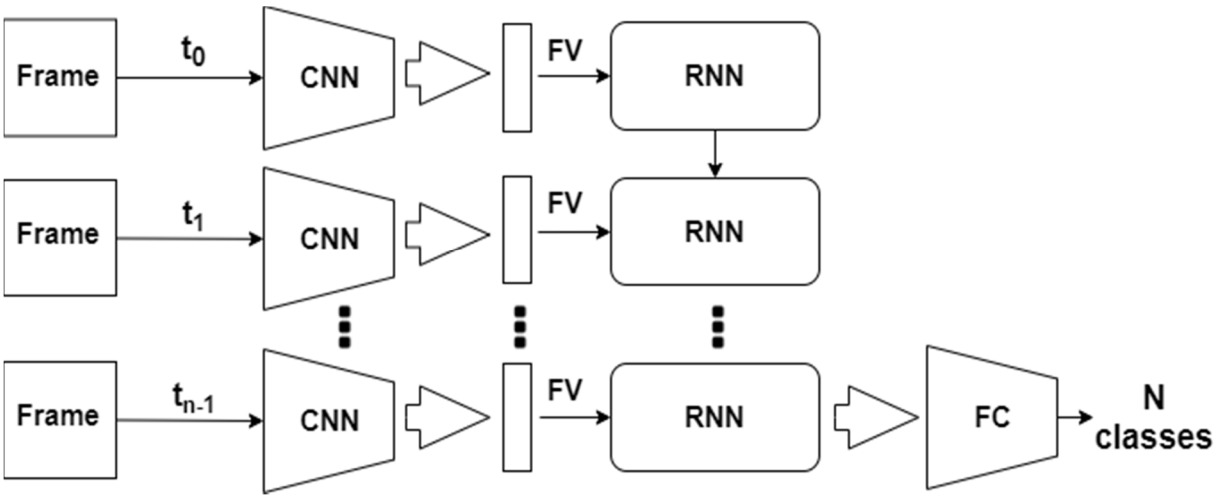

Рис. 1. Обобщенная схема комбинированной сверточно-рекуррентной архитектуры

Fig. 1. High-level diagram of a composite convolutional-recurrent architecture

В соответствии с данной архитектурой каждый кадр исследуемого видеоряда t_i поступает на вход экстрактору признаков CNN, представленному предварительно обученной на большом репрезентативном наборе изображений моделью глубокой сверточной нейронной сети. Данный экстрактор признаков осуществляет генерацию набора искусственных признаков FV_i. Полученные вектора признаков FV_i поступают на вход соответствующим блокам классификатора RNN, представленному рекуррентной нейронной сетью, в результате чего формируется результирующий набор прогнозов блоков классификации. FC – полносвязный выходной слой, обрабатывающий результаты работы выходного блока RNN. Данный слой выполняет итоговую классификацию обработанной последовательности кадров и прогнозирует метку идентифицированного человека на кадрах видеопоследовательности.

Для эффективного распознавания большого количества классов на изображениях, в качестве экстрактора признаков в моделях LSTM-1, LSTM-2, GRU-1, GRU-2 было принято решение воспользоваться сверточной нейронной сетью архитектуры resnet18 [22]. С помощью данной сети осуществляется выделение признаков визуальных объектов, которые впоследствии передаются в рекуррентную нейронную сеть. Выбор нейронной сети resnet18 в качестве блока CNN обусловлен ее широкой распространённостью и возможностью легко применить подход переноса обучения (тонкой настройки). При таком подходе в предварительно обученной модели удаляются или заново обучаются выходные слои для адаптации к новой, схожей задаче. Данный подход позволяет намного быстрее обучать нейросетевые модели специфическим задачам и требует гораздо меньшего объема входных данных.

В качестве последней апробируемой архитектуры в рамках настоящего исследования была выбрана модель трехмерной сверточной нейронной сети I3D. Трехмерные сети на данный момент используют не так широко, однако такие сети способны самостоятельно обрабатывать временные последовательности кадров в видеопотоке и в отличие от рекуррентных нейронных сетей не требуют предварительного выделения признаков для визуальных объектов. В то время, как обычные двумерные сверточные нейросети способны обрабатывать многоканальные изображения по отдельности, трехмерные сверточные нейронные сети способны выполнять обработку последовательностей многоканальных изображений или кадров.

Модель трехмерной сверточной нейронной сети I3D содержит в архитектуре блоки 3D Inception [16], необходимые для анализа пространственно-временных характеристик движения визуальных объектов в видео. Такие блоки схожи с блоками Inception в нейрон-

ной сети Inception-V1 [23]: пять первых слоев данной модели представляют собой низкоуровневые слои трехмерной свертки и слои пространственно-временных экстракторов признаков. Карта признаков, полученная из этих слоев, обрабатывается последовательными Inception-блоками и трехмерными слоями. Inception-блок, в свою очередь, распараллеливает обработку карты признаков, полученной на предыдущем слое, с помощью четырех различных ветвей. Результатом работы блока является конкатенация карт признаков, полученных с помощью этих четырех ветвей. Выход нейросетевой модели I3D реализует классификатор и представляет собой трехмерный сверточный слой размером (1x1x1), с активационной функцией softmax.

Для всех рассматриваемых архитектур, в качестве функции ошибок, определяющей качество работы нейронных сетей во время обучения, была выбрана логарифмическая функция потерь или перекрестная энтропия:

$$\text{logloss} = -\frac{1}{n}\sum_{i=1}^{N} y_i \log \hat{y}_i ,$$

где N – количество классов; y_i – эталонное значение класса; $\hat{y}_i$ – актуальное значение класса, сгенерированное нейронной сетью. Выбор функции обусловлен тем, что перекрестная энтропия в задачах классификации обеспечивает более быструю сходимость алгоритмов обучения по сравнению с другими функциями потерь. Более низкое значение потерь означает лучшие прогнозы [24]. В качестве оптимизатора, который управляет обратным распространением ошибки, был использован алгоритм Adam (Adaptive Momentum Estimation), который зарекомендовал себя при использовании в больших моделях и при работе с большими наборами данных [25].

В данной работе в качестве функции активации была выбрана логарифмическая нормализованная экспоненциальная функция (logsoftmax). Данная функция является расширением функции softmax и является более стабильной, чем стандартная softmax, с точки зрения выполнения вычислительных операций с плавающей точкой [26].

Далее перейдем к оценке результатов обучения представленных выше моделей с точки зрения их применимости в рамках предложенного подхода к решению задачи идентификации человека по походке на видеопоследовательностях.

Результаты и их обсуждение

Апробация и оценка качества функционирования разработанного под-хода к идентификации человека по походке на видеопоследовательностях производилась на основе набора данных CASIA Gait Database [27], поскольку он является одним из самых больших наборов данных для анализа движения и походки. Данный набор данных был составлен в исследовательских целях и содержит в себе большой объем данных

различного типа. Содержащиеся в наборе видеофайлы были записаны следующим образом: каждый из 124 людей прошел со своей привычной походкой перед 11 видеокамерами, таким образом получилось 11 разных углов прохода от 0 до 180 градусов. Кроме того, те же самые персоны повторили свою походку, сменяя одежду и вещи (сумка, портфель), влияющие на их походку. Таким образом, было сформировано 15004 видеофайла разрешением 320x240 пикселей.

Каждой видеопоследовательности (набору кадров) в выборке ставится в соответствие своя метка (label), указывающая идентификатор субъекта: от 0 до 123. Данные метки необходимы для обеспечения процесса обучения с учителем. Для обучения предложенных в настоящем исследовании нейросетевых моделей используется обучающая выборка, которая содержит 80% от общего количества всех видеофайлов в наборе данных CASIA Gait Database. В процессе обучения на вход каждой нейросетевой модели в зависимости от конкретной архитектуры поступали 16 или 32 подряд идущих кадра видеоряда, при этом начальный кадр определялся случайным образом – таким образом выполнялась временная аугментация данных. В качестве пространственной аугментации кадров видеопотока выполнялись следующие процедуры: случайный поворот изображения на $\pm 15°$, зеркальное отражение кадров и вырезание случайного фрагмента на кадре. При этом, выбранные процедуры выполнялись единым образом для всех кадров обрабатываемого нейронной сетью фрагмента. Обучение всех реализованных моделей производилось на полном наборе данных из 124 меток. Примеры работы, реализованной идентификации человека по походке в видеопотоке с помощью апробируемых нейросетевых моделей, представлены на рис. 2.

На представленных выше иллюстрациях работы нейросетевой модели GRU-1 в левом нижнем углу отображаются: предсказанный нейронной сетью индекс (метка) человека на видео (prediction), точность, с которой была предсказана метка (valid label), а также действительное значение метки для данной видеопоследовательности (probability).

По результатам ряда экспериментов по обучению рассматриваемых нейросетевых моделей, эвристически были установлены оптимальные значения числа эпох для обучения: 200 эпох для обучения I3D модели и 150 для сверточно-рекуррентных моделей (рис. 3). При обучении после каждой эпохи проводилась процедура валидации на тестовых данных, с помощью которой, во-первых, выполнялась проверка качества работы моделей, и, во-вторых, выполнялся контроль значений функции потерь для борьбы с переобучением.

Рис. 2. Примеры идентификации человека по походке в видеопотоке с помощью нейросетевой модели GRU-1

Fig. 2. Examples of human identification by gait in video stream, based on a neural network model GRU-1

Выбор указанного выше количества эпох обусловлен тем, что в ходе обучения рассматриваемых нейронных сетей при увеличении количества эпох значение функции потерь снижается и на этапе обучения, и на этапе валидации, что свидетельствует об отсутствии переобучения. Кроме того, значение функции потерь с увеличением количества эпох стабилизируется. Таким образом, выбор такого количества эпох позволяет соблюсти баланс между общим временем, затрачиваемым на обучение моделей нейронных сетей, и получаемой точностью.

На основе данных, полученных в результате применения разработанного подхода к набору тестовых видеопоследовательностей, были сформированы количественные оценки точности работы (accuracy) апробированных архитектур нейросетевых моделей. В качестве тестового набора видеопоследовательностей выступала валидационная часть набора данных CASIA Gait Database. Полученные результаты представлены ниже в табл. 1.

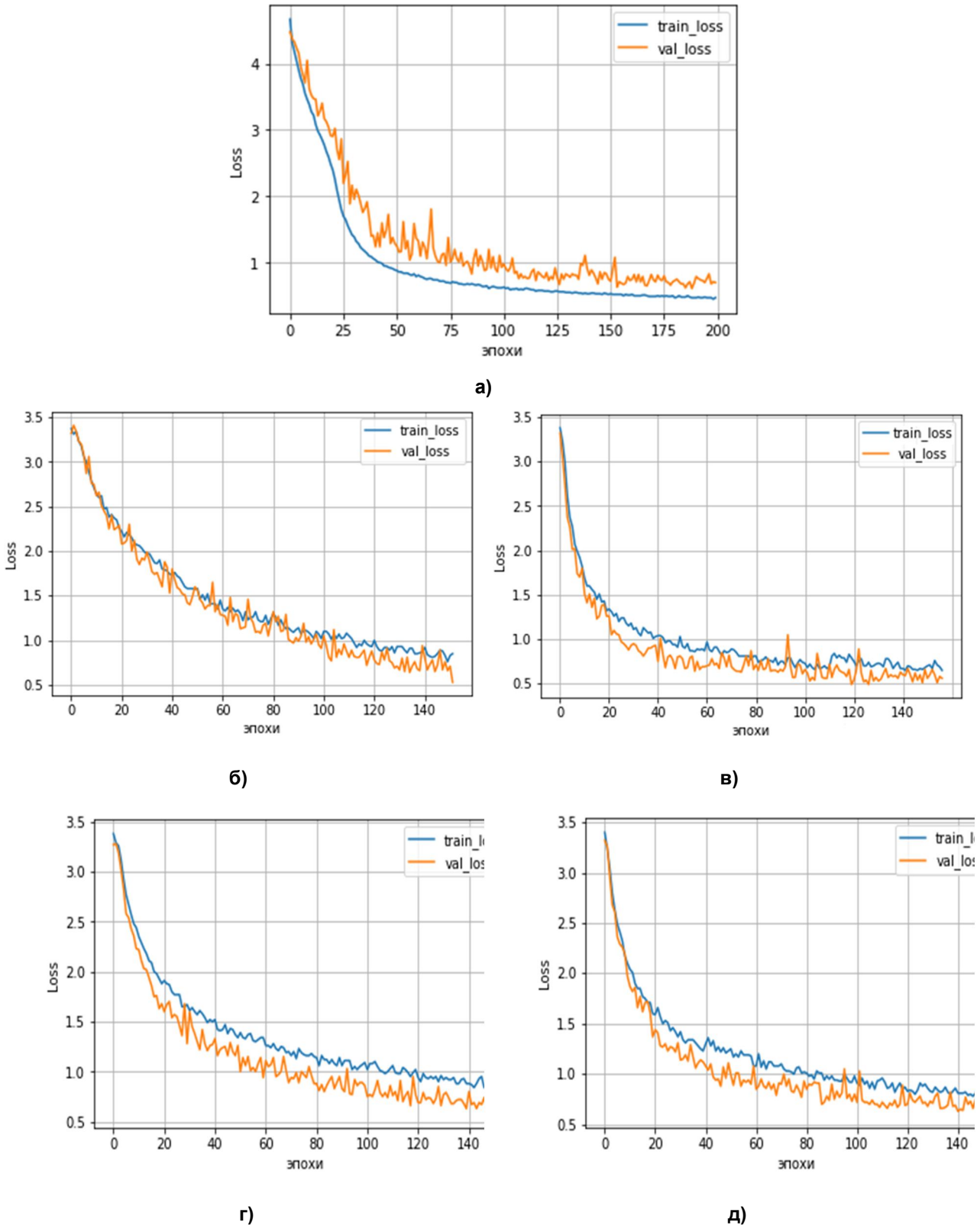

Рис. 3. График зависимости величины ошибки от количества пройденных эпох обучения: модели I3D (**а**), модели LSTM-1 (**б**), модели GRU-1 (**в**), модели LSTM-2 (**г**), модели GRU-2 (**д**)

Fig. 3. Curve of error value depending on the number of model learning epochs passed: model I3D (**а**), model LSTM-1 (**б**), model GRU-1 (**в**), model LSTM-2 (**г**), model GRU-2 (**д**)

Таблица 1. Оценки точности идентификации человека по походке с использованием различных нейросетевых моделей

Table 1. Estimation of accuracy of human identification by gait, using various neural network models

Нейросетевая модель / Neural network model	Accuracy
GRU-1	85%
LSTM-1	83%
LSTM-2	82%
GRU-2	81%
I3D	79%
Sokolova [2]	74%
Wu [27]	73%
Yu, SPAE [28]	64%
Yu, GaitGAN [29]	63%
Feng [13]	58%

Как можно заметить из таблицы, предложенные модели на основе архитектуры I3D и сверточно-рекуррентных архитектур показали более высокую точность, чем существующие методы решения задачи идентификации человека по походке. В частности, в рамках экспериментальной оценки средняя точность определения человека по походке в видеопотоке с использованием трехмерной сверточной модели I3D составила 79%. Для сверточно-рекур-рентных архитектур наилучшие показатели точности составили 83% и 85% для LSTM-1 и GRU-1 моделей соответственно. По результатам моделирования следует также отметить, что рекуррентные модели способны покадрово обрабатывать информацию и выводить результаты в реальном масштабе времени. Среднее время обработки одного кадра составляет 0,2067 с для LSTM и 0,2097 с для GRU на CPU AMD RYZEN 7 2700x.

Полученные экспериментальные результаты позволяют сделать вывод, что среди всех апробированных нейросетевых архитектур лучшие показатели качества работы характерны для сверточно-рекуррентных нейросетевых моделей, среди которых, в свою очередь, наилучшие показатели продемонстрировали однонаправленные рекуррентные модели. Модель на основе трехмерной сверточной нейронной сети I3D с точки зрения точности работы показала результат, сопоставимый с другими апробированными архитектурами, однако данная модель является более ресурсоемкой с точки зрения выполняемых операций с плавающей точкой. Кроме того, нейросетевая модель I3D не способна выполнять классификацию в порядке поступления кадров видеопотока. Для выполнения классификации трехмерные сверточные нейросети должны обрабатывать цельные последовательности кадров заданной длины. В то время как модели, основанные на применении сверточно-рекуррентных архитектур, способны покадрово обрабатывать видеопоток и выполнять классификацию после обработки каждого кадра в порядке их поступления. В табл. 2 приведены результаты времени обработки последовательности из 32 кадров разрешением 224x224 пикселя, что примерно соответствует секунде воспроизведения видеоряда с частотой, равной 30 кадров/с, на вычислительном оборудовании следующей конфигурации: ЦПУ Intel i7 6700k 4000 МГц, ОЗУ 32 ГБ DDR4 2133 МГц, графический процессов общего назначения Nvidia GTX 1080.

Таблица 2. Полученные результаты в отношении времени обработки 32 кадров видеопоследовательности рассматриваемыми нейросетевыми моделями

Table 2. Processing time values, obtained on video stream of 32 frames using the neural network models considered

Модель / Model	Время обработки на ЦПУ, с / The processing time on CPU, c	Время обработки на GPU, с / The processing time on the GPU, c
I3D	1.432	0.041
LSTM-1	0.809	0.065
LSTM-2	0.804	0.065
GRU-1	0.847	0.070
GRU-2	0.826	0.069

Из табл. 2 видно, что время обработки последовательности из 32 кадров моделью I3D на графическом ускорителе ниже, чем сверточно-рекуррентными нейросетевыми архитектурами. В то же время обработки последовательности из 32 кадров моделью I3D на центральном процессоре существенно превышает время работы на графическом ускорителе ввиду большой вычислительной сложности операций трехмерных сверток, которые лежат в основе архитектуры I3D. Таким образом, в качестве классификатора для разработанного подхода к идентификации человека по походке на видеопоследовательности с точки зрения точности и скорости работы моделей, а также возможности покадровой обработки видеопотока предпочтительным является использование сверточно-рекуррентных архитектур на основе LSTM-1 или GRU-1 моделей соответственно.

Выводы

По результатам апробации разработанного подхода к идентификации человека по походке на наборе данных CASIA Gait Database, предложенное решение продемонстрировало высокое качество работы для всех рассмотренных нейросетевых моделей. В частности, точность идентификации человека с использованием в качестве классификатора трехмерной сверточной нейронной сети I3D составила 79%. Другие 4 рассмотренные сверточно-рекуррентные архитектуры продемонстрировали еще более высокие показатели: двунаправленная GRU – 81%, двунаправленная LTSM – 82%, однонаправленная LTSM – 83% и однонаправленная GRU – 85%. Результаты, продемонстрированные разработанным методом с использованием данных архитектур, существенно превосходят известные подходы, основанные как на использовании динамических характеристик походки, анализе человеческого скелета, сконструированных вручную признаков, так и подходы, в основе которых лежит применение других нейросетевых моделей. Кроме того, при использовании предложенного подхода не требуется осуществлять предварительную обработку кадров видеоряда, а также выполнять извлечение признаков вручную.

Несмотря на близкие по точности результаты, рассмотренные сверточно-рекуррентные модели обладают возможностью покадровой обработки видео, показывают более высокую точность идентификации, меньшее время обработки данных на центральном процессоре и более высокую скорость обучения в сравнении с моделью I3D. Таким образом, в качестве классификатора для разработанного подхода к идентификации человека по походке на видеопоследовательности предпочтительным является использование сверточно-рекуррентных архитектур на основе однонаправленной LSTM или GRU моделей соответственно.

В дальнейшем предполагается исследовать возможность модернизации разработанного подхода для повышения эффективности его работы в условиях наличия нескольких людей на кадрах анализируемой видеопоследовательности.

Список литературы

1. Распознавание лиц на групповых фотографиях с использованием алгоритмов сегментации / А.И. Шерстобитов, В.П. Федосов, В.А. Приходченко, М.В. Тимофеев // Известия Южного федерального университета. Технические науки. 2013. 11(148). URL: https://cyberleninka.ru/article/n/raspoznavanie-lits-na-gruppovyh-fotografiyah-s-ispolzovaniem-algoritmov-segmentatsii

2. Sokolova A., Konushin A. Gait recognition based on convolutional neural networks // International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. 2017. XLII-2/W4. P. 207-212. https://doi.org/isprs-archives-XLII-2-W4-207-2017

3. Sokolova A., Konushin A. Pose-based deep gait recognition // IET Biometrics. 2018. 8(2). P. 134-143. https://doi.org/10.1049/iet-bmt.2018.5046

4. Han J., Bhanu B. Individual recognition using gait energy image // IEEE transactions on pattern analysis and machine intelligence. 2005. 28(2). P. 316-322. https://doi.org/10.1109/TPAMI.2006.38

5. Лютов В.С., Конушин А.С., Арсеев С.П. Распознавание человека по походке и внешности // Программирование. 2018. № 4. С. 97-106. https://doi.org/10.31857/S000523100000515-0

6. Соколова А.И., Конушин А.С. Методы идентификации человека по походке в видео // Труды Института системного программирования РАН. 2019. №31(1). С. 69-82. https://doi.org/10.15514/ISPRAS-2019-31(1)-5

7. Alotaibi M., Mahmood A. Improved gait recognition based on specialized deep convolutional neural network // Computer Vision and Image Understanding. 2017. № 164. P. 103-110. https://doi.org/10.1016/j.cviu.2017.10.004

8. Малашин Р.О., Луцив В.Р. Восстановление силуэта руки в задаче распознавания жестов с помощью адаптивной морфологической фильтрации бинарного изображения // Оптический журнал. 2013. № 80(11). С. 54-61.

9. Frame difference energy image for gait recognition with incomplete silhouettes / C. Chen, J. Liang, H. Zhao, H. Hu, J. Tian // Pattern Recognition Letters. 2009. №30(11). P. 977-984. https://doi.org/10.1016/j.patrec.2009.04.012

10. Castro F.M., Marín-Jimenez M.J., Medina-Carnicer R. Pyramidal Fisher Motion for Multiview Gait Recognition // 2014 22nd International Conference on Pattern Recognition, Stockholm. 2014. P. 1692-1697. https://doi.org/doi: 10.1109/ICPR.2014.298

11. Kaaniche M.B., Bremond F. Tracking hog descriptors for gesture recognition // 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2009. P. 140-145. https://doi.org/10.1109/AVSS.2009.26

12. Realtime video classification using dense hof/hog / J.R.R. Uijlings, I.C. Duta, N. Rostamzadeh, N. Sebe // Proceedings of international conference on multimedia retrieval. 2014. P. 145-152. https://doi.org/10.1145/2578726.2578744

13. Feng Y., Li Y., Luo J. Learning effective gait features using LSTM // 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016. P. 325-330. https://doi.org/0.1109/ICPR.2016.7899654

14. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. № 9(8). P. 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

15. Learning spatiotemporal features with 3d convolutional networks / D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri // Proceedings of the IEEE international conference on computer vision. 2015. P. 4489-4497. https://doi.org/10.1109/ICCV.2015.510

16. Carreira J., Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 6299-6308. https://doi.org/10.1109/CVPR.2017.502

17. Hara K., Kataoka H., Satoh Y. Learning spatio-temporal features with 3D residual networks for action recognition // Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017. P. 3154-3160. https://doi.org/10.1109/ICCVW.2017.373

18. Hara K., Kataoka H., Satoh Y. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018. P. 6546-6555. https://doi.org/10.1109/CVPR.2018.00685

19. Saveliev A., Uzdiaev M., Dmitrii M. Aggressive Action Recognition Using 3D CNN Architectures // 2019 12th International Conference on Developments in eSystems Engineering (DeSE). IEEE, 2019. P. 890-895. https://doi.org/10.1109/10.1109/DeSE.2019.00165

20. Learning phrase representations using RNN encoder-decoder for statistical machine translation / K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio // arXiv preprint arXiv:1406.1078. 2014. URL: https://arxiv.org/abs/1406.1078

21. Beyond short snippets: Deep networks for video classification / J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. P. 4694-4702. https:// doi.org/10.1109/CVPR.2015.7299101

22. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 770-778. https://doi.org/10.1109/CVPR.2016.90

23. Going deeper with convolutions / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. P. 1-9. https://doi.org/ 10.1109/CVPR.2015.7298594

24. What is Log Loss? [Quoted May 6, 2020]. URL: https://www.kaggle.com/ dansbecker/what-is-log-loss

25. Kingma D.P., Ba J. Adam. A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014. URL: https://arxiv.org/abs/1412.6980

26. Logsoftmax vs softmax [Quoted May 6, 2020]. URL: https://discuss.pytorch.org/t/ logsoftmax-vs-softmax/21386

27. A comprehensive study on cross-view gait based human identification with deep cnns / Z. Wu, Y. Huang, L. Wang, X. Wang, T. Tan // IEEE transactions on pattern analysis and machine intelligence. 2016. № 39(2). P. 209-226. https://doi.org/ 10.1109/ TPAMI.2016.2545669

28. Invariant feature extraction for gait recognition using only one uniform model / S. Yu, H. Chen, Q. Wang, L. Shen, Y. Huang // Neurocomputing. 2017. 239. P. 81-93. https://doi.org/10.1016/j.neucom.2017.02.006

29. GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Network / S. Yu, H. Chen, E.B.G. Reyes, N. Poh // In Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. P. 532-539. https://doi.org/10.1109/CVPRW.2017.80

References

1. Sherstobitov A.I., Fedosov V.P., Prihodchenko V.A., Timofeev D.V. Raspoznavanie lits na gruppovykh fotografiyakh s ispol'zovaniem algoritmov segmentatsii [Face recognition on groups photos with using segmentation algorithms]. *Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki = Bulletin of the Southern Federal University. Technical science*, 2013, no. 11(148) (In Russ.). Available at: https://cyberleninka.ru/ article/n/raspoznavanie-lits-na-gruppovyh-fotografiyah-s-ispolzovaniem-algoritmov-segmentatsii

2. Sokolova A., Konushin A. Gait recognition based on convolutional neural networks. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2017; XLII-2/W4, pp. 207-212. https://doi.org/isprs-archives-XLII-2-W4-207-2017

3. Sokolova A., Konushin A. Pose-based deep gait recognition. *IET Biometrics*, 2018, no. 8(2), pp. 134-143. https://doi.org/10.1049/iet-bmt.2018.5046

4. Han J., Bhanu B. Individual recognition using gait energy image. *IEEE transactions on pattern analysis and machine intelligence*, 2005, no. 28(2), pp. 316-322. https://doi.org/10.1109/TPAMI.2006.38

5. Liutov V., Konushin A., Arseev S. Raspoznavanie cheloveka po pokhodke i vneshnosti [Human recognition by appearance and gait]. *Programmirovanie = Programming and Computer Software,* 2018, no. 44(4), pp. 258-265 (In Russ.). https://doi.org/ 10.31857/ S000523100000515-0

6. Sokolova A.I., Konushin A.S. Metody identifikatsii cheloveka po pokhodke v video [Methods of gait recognition in video]. *Trudy Instituta sistemnogo programmirovaniya RAN = Proceedings of the Institute for System Programming of the RAS* (*Proceedings of ISP RAS*), 2019, no. 31(1), pp. 69-82 (In Russ.). https://doi.org/10.15514/ISPRAS-2019-31(1)-5

7. Alotaibi M., Mahmood A. Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 2017, no. 164, pp. 103-110. https://doi.org/10.1016/j.cviu.2017.10.004

8. Malashin R.O., Lutsiv V.R. Vosstanovlenie silueta ruki v zadache raspoznavaniya zhestov s pomoshch'yu adaptivnoi morfologicheskoi fil'tratsii binarnogo izobrazheniya [Restoring a silhouette of the hand in the problem of recognizing gestures by adaptive morphological filtering of a binary image]. *Opticheskii zhurnal = Journal of Optical*, 2013, no. 80(11), pp. 54-61 (In Russ.). https://doi.org/10.1364/JOT.80.000685

9. Chen C., Liang J., Zhao H., Hu H., Tian J. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 2009, no. 30(11), pp. 977-984. https://doi.org/10.1016/j.patrec.2009.04.012

10. Castro F.M., Marín-Jimenez M.J., Medina-Carnicer R. Pyramidal Fisher Motion for Multiview Gait Recognition. *2014 22nd International Conference on Pattern Recognition*, Stockholm, 2014, pp. 1692-1697. https://doi.org/doi: 10.1109/ICPR.2014.298

11. Kaaniche M.B., Bremond F. Tracking hog descriptors for gesture recognition. *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2009, pp. 140-145. https://doi.org/10.1109/AVSS.2009.26

12. Uijlings J.R.R., Duta I.C., Rostamzadeh N., Sebe N. Realtime video classification using dense hof/hog. *Proceedings of international conference on multimedia retrieval*, 2014, pp. 145-152. https://doi.org/10.1145/2578726.2578744

13. Feng Y., Li Y., Luo J. Learning effective gait features using LSTM. *2016 23rd International Conference on Pattern Recognition* (*ICPR*). IEEE, 2016, pp. 325-330. https://doi.org/0.1109/ICPR.2016.7899654

14. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural computation*, 1997, no, 9(8), pp.1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

15. Tran D., Bourdev L., Fergus R., Torresani L., Paluri M. Learning spatiotemporal features with 3d convolutional networks. *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489-4497. https://doi.org/10.1109/ICCV.2015.510

16. Carreira J., Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2017, pp. 6299-6308. https://doi.org/10.1109/CVPR.2017.502K

17. Hara K., Kataoka H., Satoh Y. Learning spatio-temporal features with 3D residual networks for action recognition. *Proceedings of the IEEE International Conference on Computer Vision Workshops,* 2017, pp. 3154-3160. https://doi.org/10.1109/ICCVW.2017.373

18. Hara K., Kataoka H., Satoh Y. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition,* 2018, pp. 6546-6555. https://doi.org/10.1109/CVPR.2018.00685

19. Saveliev A., Uzdiaev M., Dmitrii M. Aggressive Action Recognition Using 3D CNN Architectures. *2019 12th International Conference on Developments in eSystems Engineering* (*DeSE*). IEEE, 2019, pp. 890-895. https://doi.org/10.1109/10.1109/DeSE.2019.00165

20. Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. 2014. Available at: https://arxiv.org/abs/1406.1078

21. Yue-Hei Ng J., Hausknecht M., Vijayanarasimhan S., Vinyals O., Monga R., Toderici G. Beyond short snippets: Deep networks for video classification. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp.4694-4702. https://doi.org/10.1109/CVPR.2015.7299101

22. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

23. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9. https://doi.org/10.1109/CVPR.2015.7298594

24. What is Log Loss? [Quoted May 6, 2020]. Available at: https://www.kaggle.com/dansbecker/what-is-log-loss

25. Kingma D.P., Ba J. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014. Available at: https://arxiv.org/abs/1412.6980

26. Logsoftmax vs softmax [Quoted May 6, 2020]. Available at: https:// discuss.pytorch.org/t/logsoftmax-vs-softmax/21386

27. Wu Z., Huang Y., Wang L., Wang X., Tan T. A comprehensive study on cross-view gait based human identification with deep CNNS. *IEEE transactions on pattern analysis and machine intelligence*, 2016, no. 39(2), pp. 209-226. https://doi.org/10.1109/TPAMI.2016.2545669

28. Yu S., Chen H., Wang Q., Shen L., Huang Y. Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing*. 2017, no. 239, pp. 81-93. https://doi.org/10.1016/j.neucom.2017.02.006

29. Yu S., Chen H., Reyes E. B. G., Poh, N. GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Network. *In Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). 2017, pp. 532-539. https://doi.org/10.1109/CVPRW.2017.80

Информация об авторах / Information about the Authors

Уздяев Михаил Юрьевич, младший научный сотрудник лаборатории технологий больших данных социокиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: m.y.uzdiaev@gmail.com

Mikhail Yu. Uzdiaev, Junior Researcher of Laboratory of Big Data in Socio-Cyberphysical Systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: m.y.uzdiaev@gmail.com

Яковлев Роман Никитич, младший научный сотрудник лаборатории технологий больших данных социокиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: iakovlev.r@mail.ru

Roman N. Iakovlev, Junior Researcher of Laboratory of Big Data in Socio-Cyberphysical Systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: iakovlev.r@mail.ru

Дударенко Дмитрий Михайлович, младший научный сотрудник лаборатории технологий больших данных социокиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: dmitry@dudarenko.net

Dmitry M. Dudarenko, Junior Researcher of Laboratory of Big Data in Socio-Cyberphysical Systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: dmitry@dudarenko.net

Жебрун Александр Дмитриевич, программист лаборатории технологий больших данных социокиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: sashakotovich@gmail.com

Aleksandr D. Zhebrun, Programmer of Laboratory of Big Data in Socio-Cyberphysical Systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: sashakotovich@gmail.com