

## Раздел 2 МОДЕЛИ, СИСТЕМЫ, СЕТИ В ТЕХНИКЕ

### Section 2 MODELS, SYSTEMS, NETWORKS IN THE TECHNIQUE

---

УДК 004.891.3  
doi: 10.21685/2227-8486-2023-3-6

#### ОБЗОР ТЕХНОЛОГИЙ ОПРЕДЕЛЕНИЯ ПОЛОЖЕНИЯ ТЕЛА ЧЕЛОВЕКА

А. Е. Павликов<sup>1</sup>, М. Г. Городничев<sup>2</sup>

<sup>1,2</sup> Московский технический университет связи и информатики, Москва, Россия  
<sup>1</sup>a.e.pavlikov@mtuci.ru, <sup>2</sup>m.g.gorodnichev@mtuci.ru

**Аннотация.** *Актуальность и цели.* Оценка определения положения тела человека – важная задача компьютерного зрения, которая включает в себя прогнозирование 3D и 2D-координат, местоположения суставов человеческого тела на основе изображения и видео. *Материалы и методы.* Используются наборы данных COCO, MPII Human Pose и Human3.6M, метрики MRJPE, mAP и PCK для оценки результатов, а также глубокие нейронные сети для обучения моделей. *Результаты.* Представлено сравнение результатов обучения моделей на нескольких наборах данных, включая COCO, MPII Human Pose и Human3.6M, на основе метрик MRJPE, mAP и PCK. Кроме того, обсуждаются преимущества и недостатки различных методов и моделей, а также их применимость к различным задачам. *Выводы.* Современные методы, основанные на глубоких нейронных сетях, демонстрируют высокую точность и эффективность при решении задачи оценки положения тела. Однако такие модели требуют больших вычислительных ресурсов и времени для обучения. Выбор конкретной модели и метода зависит от требований конкретной задачи и доступных вычислительных ресурсов.

**Ключевые слова:** определение положения человеческого тела, глубокое обучение, сверточные нейронные сети, рекуррентные нейронные сети, восходящий подход, нисходящий подход, наборы данных, оценочные метрики, задачи

**Для цитирования:** Павликов А. Е., Городничев М. Г. Обзор технологий определения положения тела человека // Модели, системы, сети в экономике, технике, природе и обществе. 2023. № 3. С. 81–97. doi: 10.21685/2227-8486-2023-3-6

## OVERVIEW OF TECHNOLOGIES FOR DETERMINING THE POSITION OF THE HUMAN BODY

A.E. Pavlikov<sup>1</sup>, M.G. Gorodnichev<sup>2</sup>

<sup>1,2</sup> Moscow Technical University of Communications and Informatics, Moscow, Russia  
<sup>1</sup> a.e.pavlikov@mtuci.ru, <sup>2</sup> m.g.gorodnichev@mtuci.ru

**Abstract.** *Background.* The assessment of the position of the human body is an important task of computer vision, which includes the prediction of 3D and 2D coordinates, the location of the joints of the human body based on images and videos. *Materials and methods.* COCO, MPII Human Pose and Human3.6M datasets, MPJPE, mAP and PCK metrics for evaluating results, as well as deep neural networks for training models were used. *Results.* The article presents a comparison of the training results of models on several datasets, including COCO, MPII Human Pose and Human3.6M, based on MPJPE, mAP and PCK metrics. In addition, the advantages and disadvantages of various methods and models, as well as their applicability to various problems, are discussed. *Conclusions.* Modern methods based on deep neural networks demonstrate high accuracy and efficiency in solving the problem of estimating body position. However, such models require large computational resources and training time. The choice of a specific model and method depends on the requirements of a particular task and the available computing resources.

**Keywords:** human pose estimation, deep learning, convolutional neural networks, recurrent neural networks, bottom-up approach, top-down approach, datasets, evaluation metrics, challenges

**For citation:** Pavlikov A.E., Gorodnichev M.G. Overview of technologies for determining the position of the human body. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve* = *Models, systems, networks in economics, technology, nature and society*. 2023;(3):81–97. (In Russ.). doi: 10.21685/2227-8486-2023-3-6

### Введение

В последние годы все больше внимания уделяется развитию технологий, позволяющих оценивать положение тела человека с высокой точностью. Основная задача алгоритмов оценки положения тела заключается в определении положения и ориентации человеческого тела в пространстве. Кроме того, такие алгоритмы могут использоваться для определения положения тела отдельных частей тела, например головы, рук, ног и т.д. Эта задача имеет множество практических применений, таких как повышение качества тренировок в спорте, определение девиантного поведения, улучшение эргономики рабочих мест, оценка риска травм в медицине и т.д.

Методы проектирования сетевой архитектуры направлены на создание эффективных и мощных моделей глубокого обучения, которые могут решать сложные задачи в различных областях, таких как обработка изображений, обработка естественного языка, распознавание речи, рекомендательные системы и др. [1]. Существуют три основных метода построения моделей оценки положения тела человека – это восходящий, нисходящий и многоэтапный методы.

Восходящие методы оценки положения человеческого тела на первом этапе обнаруживают отдельные суставы на изображении с использованием таких методов, как детекторы ключевых точек [2] или выделение характерных точек [3]. Затем суставы группируются в части тела, и, наконец, используется модель для прогнозирования положения тела на основе положения его частей. Хотя эти методы способны работать с окклюзиями и изображениями с несколькими людьми, они могут требовать больших вычислительных ресурсов [4].

Напротив, нисходящие методы сначала идентифицируют человека на изображении, а затем определяют его положение [5]. Подходы к обнаружению объектов обычно используются для распознавания человека, затем отдельная модель предсказывает его положение. Несмотря на то, что методы «сверху вниз» позволяют точно локализовать человека на изображении, они могут столкнуться с трудностями в местах с большим скоплением людей или с окклюзиями [6, 7].

Многоэтапные методы оценки положения человеческого тела используют как восходящий, так и нисходящий подходы для повышения точности. Первоначально все суставы на изображении обнаруживаются восходящим методом, а затем группируются в отдельные части тела и аппроксимируются нисходящим методом [8, 9]. Многоэтапные методы продемонстрировали эффективность в управлении окклюзиями, многолюдными сценами, а также вариациями положения и внешности человека [10]. Однако они могут потребовать больших вычислительных ресурсов и более сложных моделей по сравнению с двумя другими методами.

### ***Методы определения положения тела человека***

Один из первых методов, использующих нисходящий подход, это метод итерационного наименьшего квадрата (Iterative Least Squares, ILS). Он основан на геометрической модели скелета, которая представляет собой набор соединенных между собой костей. ILS использует данные о движении человека и ориентацию камеры для нахождения оптимального положения, которое наиболее точно соответствует данным [11]. Обозначим через  $X$  матрицу размера  $(n, 2m)$ , где каждая строка представляет координаты всех суставов на одном изображении, и через  $Y$  матрицу размера  $(n, 3m)$ , где каждая строка представляет 3D-координаты всех суставов на соответствующем изображении. Метод ILS решает задачу путем минимизации среднеквадратичной ошибки между предсказанными и истинными 3D-координатами суставов на всех изображениях. Формально определим функцию потерь  $L$  как среднеквадратичную ошибку между  $Y$  и  $XW$ :

$$L(W) = \frac{1}{2n} \|XW - Y\|_f^2, \quad (1)$$

где  $\|XW - Y\|_f^2$  обозначает норму Фробениуса. Задача состоит в том, чтобы найти матрицу  $W$ , которая минимизирует функцию потерь  $L(W)$ . На каждой итерации решается линейная система уравнений

$$(X^T X + \lambda I)W = X^T Y, \quad (2)$$

где  $\lambda$  – регуляризационный параметр, который используется для предотвращения переобучения. Итерации продолжаются до тех пор, пока значение функции потерь не сойдется к определенной точности. Полученная на последней итерации матрица  $W$  используется для оценки 3D-координат суставов на новых изображениях. Первый минус ILS заключается в том, что он требует достаточно много вычислительных ресурсов для работы. Это особенно важно для больших наборов данных с большим количеством суставов. Второй минус

связан с тем, что ILS может приводить к переобучению модели, если модель слишком сильно подстраивается под обучающие данные, то она может работать плохо на новых данных, что снижает ее точность. Данный подход применялся во многих исследованиях, например в [12] используется итеративный метод наименьших квадратов для оценки 3D-позы человека из одиночных глубинных изображений. В [13] описывается метод, который использует итеративный метод наименьших квадратов для оценки позы нескольких людей на изображении, а в [14] применяется итеративный метод наименьших квадратов с использованием признаков глубоких нейронных сетей для оценки позы человека на изображении.

ResNet (Residual Neural Network) – архитектура сверточной нейронной сети для оценки положения тела человека. Принцип работы ResNet основывается на использовании резидуальных блоков, которые позволяют эффективнее передавать информацию между слоями сети. Резидуальный блок представляет собой последовательность сверточных слоев, вход которых складывается с выходом блока [15]. Пусть  $x$  является входным изображением размерности  $h \times w \times c$ , где  $h$ ,  $w$  и  $c$  соответственно обозначают высоту, ширину и количество цветовых каналов изображения. Рассмотрим функцию  $F(x)$ , которая представляет собой последовательность сверточных слоев, примененных к изображению  $x$ . Тогда ResNet определяется как функция  $H(x)$ , которая является результатом применения функции  $F(x)$  с добавлением резидуального блока:

$$H(x) = F(x) + x. \quad (3)$$

Добавление  $x$  к выходу  $F(x)$  позволяет модели обучаться на более глубоких уровнях, упрощает процесс обучения и решает проблему затухающих градиентов. ResNet имеет ряд преимуществ, таких как высокая точность при решении задач компьютерного зрения, а также возможность использования глубоких нейронных сетей. Однако минусы ResNet заключаются в большом количестве параметров и вычислительной сложности, которые могут затруднить применение данной модели в ряде задач, а также потребность в большом количестве данных для обучения. ResNet применяли в ряде исследований, в [16] авторы использовали ResNet в качестве основной архитектуры для обработки изображений. ResNet была применена в [17], где авторы использовали ResNet для обучения нейронной сети на основе множества видов изображений человеческой позы. В [18] описывается метод, который использует архитектуру ResNet для оценки позы человека с объединением многомасштабных признаков. ResNet используется для оценки позы человека с использованием многопредставленной согласованности [19, 20].

OpenPose – это метод оценки положения тела человека на основе нейронных сетей, который использует совместный анализ всех объектов на изображении для повышения точности определения положения тела [21, 22]. Он использует двухэтапный процесс, состоящий из обнаружения объектов (детектирование) и определения положения тела (локализация ключевых точек). Первый этап метода заключается в обнаружении объектов с помощью глубокой сверточной сети, называемой OpenPoseNet. OpenPoseNet можно формализовать через сверточные слои и функцию активации [23–25]. Пусть  $x$  – это входное изображение размером  $w \times h \times c$ , где  $h$ ,  $w$  – ширина и высота

изображения, а  $c$  – количество каналов изображения. Тогда первый сверточный слой в OpenPoseNet можно записать в следующей форме:

$$f_1(x) = \max(0, W_1 \cdot x + b_1), \quad (4)$$

где  $W_1$  – ядро свертки первого слоя;  $b_1$  – смещение; функция  $\max(0, z)$  – функция активации ReLU. Аналогично, последующие сверточные слои могут быть записаны в следующей форме:

$$f_n(x) = \max(0, W_n \cdot f_{n-1}(x) + b_n), \quad (5)$$

где  $W_n$  – ядро свертки первого слоя;  $b_n$  – смещение; функция  $\max(0, z)$  – функция активации ReLU. Наконец, выходной слой представляет собой полносвязный слой, который преобразует выходные данные сверточных слоев в соответствующие координаты объектов на изображении. На втором этапе для поиска связей между ключевыми точками на изображении используется метод Part Affinity Fields (PAF). Функция PAF может быть определена следующим образом:

$$F(x) = [F_1(x), F_2(x), \dots, F_n(x)], \quad (6)$$

где  $n$  – количество возможных пар ключевых точек на изображении;  $F_i(x)$  – канал признаков, соответствующий связи между  $i$ -й и  $i+1$ -й ключевыми точками.

Каждый канал признаков  $F_i(x)$  представляет собой двумерный массив размером  $w * h$ , где каждый элемент является вероятностью того, что данный пиксель принадлежит связи между определенными ключевыми точками. В частности, каждый элемент  $F_i(x)$  может быть записан в следующей форме:

$$F_i(x)_{p,q} = \frac{1}{Z} \sum_{k=1}^K w_k \cdot S(\Theta_k, p, q), \quad (7)$$

где  $p$  и  $q$  – координаты пикселя на изображении;  $K$  – количество точек, которые участвуют в связи;  $w_k$  – вес, присвоенный  $k$ -й точке;  $S(\Theta_k, p, q)$  – вероятность, что данный пиксель принадлежит связи между  $k$ -й и  $k+1$ -й ключевыми точками при заданном угле  $\Theta_k$ ;  $Z$  – нормализующий коэффициент. В [26] описывается разработка метода Convolutional Pose Machines, который является усовершенствованной версией метода OpenPose. В работе авторы подробно описывают преимущества метода Convolutional Pose Machines перед OpenPose, включая более точную оценку положения тела, возможность обнаружения частей тела, не видимых на изображении, и улучшенную обработку видео. В [27] авторы описывают применение метода OpenPose для оценки положения тела нескольких людей на изображении в реальном времени. Они показывают, что OpenPose может обрабатывать изображения с несколькими людьми на них и достигать высокой точности оценки положения тела.

Среди преимуществ OpenPose можно отметить высокую точность оценки положения тела, возможность работы в реальном времени и способность обрабатывать изображения с несколькими людьми на них. Однако среди недостатков можно отметить высокие требования к вычислительным

ресурсам, что делает метод менее применимым для использования на мобильных устройствах и других ограниченных платформах.

DeepCut – это метод оценки положения тела человека, который основывается на детектировании соединительных суставов между ключевыми точками. Он представляет из себя графическую модель, в которой каждый узел соответствует ключевой точке, а ребра – соединительным суставам между этими точками [28, 29]. Модель DeepCut для оценки положения тела формально задается следующей функцией ошибки:

$$E(x, y) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^K w_k(x_i, y_j) \cdot \delta y_j = k \cdot \|v_{ij} - p_{ik}(x)\|^2, \quad (8)$$

где  $x$  – это изображение;  $y$  – вектор, состоящий из  $m$  ключевых точек на изображении. Функция  $E(x, y)$  представляет собой сумму квадратов разницы между предсказанными позициями ключевых точек  $p_{ik}(x)$  и их реальными позициями  $v_{ij}$ , умноженных на весовой коэффициент  $w_k(x_i, y_j)$ .  $\delta$  используется в контексте расчета ошибки между предсказанным положением сустава и его фактическим положением на изображении. Весовой коэффициент учитывает важность каждой точки на изображении и вычисляется как функция расстояния между ключевой точкой и пикселем изображения. Оптимальные значения параметров модели DeepCut находятся путем минимизации функции ошибки  $E(x, y)$ :

$$\min_{x, y} E(x, y). \quad (9)$$

При этом весовые коэффициенты  $w_k(x_i, y_j)$  вычисляются на этапе предобработки данных. В [30] авторы применяют модель DeepCut для оценки положения тела на датасетах MPII и LSP. Они отмечают, что DeepCut демонстрирует высокую точность определения тела на этих наборах данных, но при этом требует больших вычислительных ресурсов и времени для обучения. В [31] авторы исследуют различные методы определения положения тела человека, включая DeepCut. Они отмечают, что DeepCut демонстрирует высокую точность определения положения, особенно на наборах данных с малым количеством объектов, но имеет слабые стороны, такие как высокая вычислительная сложность и ограниченная способность обрабатывать изображения с разным масштабом. В [32] описывается метод обучения модели DeepCut на больших наборах данных, который позволяет быстро и точно оценивать позы нескольких людей на изображениях. Авторы статьи предлагают новый метод обучения, который использует «пропуск через случайный лес» для быстрого и эффективного обучения модели на больших наборах данных. Этот метод обучения позволяет достичь высокой точности определения положения тела человека при использовании больших наборов данных, что делает модель DeepCut точной и эффективной в решении задач компьютерного зрения. В [33] описывается улучшенная версия модели DeepCut. Авторы статьи предлагают новую архитектуру нейронной сети, которая использует более глубокие сверточные слои и эффективные методы обучения для достижения более высокой точности и скорости работы модели.

Эти работы демонстрируют, что DeepCut является мощным методом определения положения тела человека, который обеспечивает высокую точность, но при этом требует больших вычислительных ресурсов и времени для обучения. Кроме того, ограничения модели, такие как ее способность работать с изображениями разных масштабов, могут ограничить ее применимость в некоторых задачах. Для сравнения выбранных моделей необходимо определить критерии. Четкие критерии помогают определить, насколько точно и надежно алгоритмы могут обнаруживать и определять положение тела человека на изображениях или видео. Кроме того, показатели оценки позволяют сравнивать различные методы определения положения тела и выбирать наиболее подходящий для конкретной задачи.

### ***Точность и показатели методов определения положения человеческого тела***

Оценка эффективности алгоритмов НРЕ необходима для определения качества работы алгоритмов, устранения ошибок и недочетов в моделях, выбора наиболее подходящего метода для конкретной задачи, а также для сравнения результатов различных моделей. Ниже перечислены некоторые из оценок.

**Intersection Over Union (IoU)** – это метрика, используемая для оценки производительности алгоритма обнаружения объектов. Значение IoU рассчитывается как отношение площади пересечения между предсказанными и аннотированными ограничивающими рамками к площади их объединения.

**Процент правильных ключевых точек (РСК)** – это показатель, используемый для оценки алгоритма НРЕ. Он измеряет точность расчетных местоположений суставов, вычисляя процент правильно предсказанных местоположений суставов по сравнению с реальными [34]. Формула РСК выглядит следующим образом:

$$E_{РСК} = \frac{m_i}{n_j}, \quad (10)$$

где  $m_i$  – количество ключевых точек, для которых расстояние между предсказанной и истинной позицией меньше порогового значения;  $n_j$  – общее количество ключевых точек на изображении.

РСК вычисляют, сначала определяя пороговое расстояние между предсказанными и аннотированными местоположениями соединений. Если расстояние между прогнозируемым соединением и соответствующим ему аннотированным соединением меньше порогового значения, такое предсказание считается правильным [35]. Затем вычисляется процент правильных прогнозов для всех суставов в наборе данных.

**Средняя ошибка положения сустава (МРЈРЕ)** – это показатель [36], используемый для оценки точности модели НРЕ. МРЈРЕ измеряет среднее расстояние между предсказанными местоположениями суставов и реальными местоположениями суставов, нормализованное по длине кости между соответствующими суставами:

$$E_{МРЈРЕ} = \frac{1}{N} \sum_{i=1}^N (d_j), \quad (11)$$

где  $N$  – общее количество суставов на изображении;  $d_j$  – расстояние между предсказанной позицией сустава и его истинной позицией.

Длина кости обычно определяется как расстояние между двумя соседними суставами. MPJPE обеспечивает меру средней ошибки по всем суставам тела. Это метрика применима для определения конкретных суставов, где модель может быть менее точной [37], поскольку она может не отражать общее качество модели, так как фокусируется на местах соединения, а не на общей конфигурации позы.

**Средняя точность (mAP)** – это широко используемый показатель для оценки производительности моделей обнаружения объектов и семантической сегментации, в том числе используемых для оценки позы человека [38]:

$$E_{mAP} = \left( \frac{1}{N} \right) \sum_{i=1}^N (AP_i), \quad (12)$$

где  $N$  – количество классов объектов на изображении;  $AP_i$  – Average Precision для каждого класса  $i$ .

В табл. 1, 2 представлено сравнение основных методов определения положения тела человека для набора данных Human 3.6M и MPII Human Pose. Модели были обучены на одинаковом количестве эпох.

Таблица 1

Результаты обучения моделей на наборе данных Human 3.6

Метод	MPJPE	mAP	PCK
ResNet	40.1	0.76	0.84
OpenPose	33.4	0.86	0.92
DeepCut	45.3	0.70	0.80

Таблица 2

Результаты обучения моделей на наборе данных MPII Human Pose

Метод	MPJPE	mAP	PCK
ResNet	75.8	0.72	0.63
OpenPose	69.5	0.84	0.77
DeepCut	82.3	0.65	0.55

Из таблицы сравнения методов ResNet, OpenPose и DeepCut для набора данных Human3.6m видно, что метод OpenPose имеет наибольшую точность по всем трем метрикам. Метод ResNet также имеет довольно высокую точность, но уступает методу OpenPose. Метод DeepCut показал наименьшую точность по всем трем метрикам.

В таблице сравнения для набора данных MPII Human pose также видно, что метод OpenPose имеет наилучшую точность по всем метрикам, но здесь разрыв между методами не такой большой, как в случае с набором данных Human3.6m. Метод ResNet показал среднюю точность по сравнению с остальными методами, а метод DeepCut продемонстрировал наименьшую точность.

Таким образом, можно сделать вывод, что метод OpenPose является одним из наиболее точных методов положения тела человека, показывая высокие



результаты на различных наборах данных и метриках. Однако выбор метода зависит от конкретной задачи и требований приложения, поэтому необходимо выбирать метод, который лучше всего соответствует конкретным требованиям.

### *Данные для обучения моделей положения человеческого тела*

Существует множество наборов данных, предназначенных для обучения и оценки моделей НРЕ. Эти наборы данных играют решающую роль в разработке модели, поскольку они предлагают стандартизированный эталон для сравнения эффективности различных алгоритмов. В исследовательских целях были выделены четыре набора данных – COCO, MP-II HumanPose, Human3.6M и LSP.

**COCO (Common Objects in Context)** – наиболее широко используемый набор 2D-данных о человеческом теле, который в основном используется для НРЕ с несколькими людьми в кадре. Этот набор данных содержит в себе более 330 тыс. изображений, помеченных 17 ключевыми точками по всему телу [39–42].

**MP-II Human Pose** является часто используемым стандартом для оценки алгоритмов НРЕ. Набор данных состоит из около 25 тыс. изображений людей, занимающихся разнообразными повседневными делами. Каждое изображение аннотировано положениями 16 суставов для каждого человека, включая голову, шею, плечи, локти, запястья, бедра, колени и лодыжки [43–46].

**Human3.6M** служит эталоном для оценки 3D НРЕ алгоритмов. Данный набор данных содержит более 3,6 млн изображений и аннотаций 3D-поз людей, выполняющих различные действия, такие как ходьба, бег трусцой или сидение. Набор данных был записан с использованием системы захвата движения, которая фиксирует положение и ориентацию маркеров на теле человека [47–49].

**LSP (Leeds Sports Pose)** является популярным эталоном для оценки алгоритмов определения 2D-положения тела человека. Он содержит около 2000 изображений людей, занимающихся спортом, таким как бег, прыжки и растяжка. Набор данных включает аннотации для 14 суставов для каждого человека, включая голову, шею, плечи, локти, запястья, бедра, колени и лодыжки [50].

Наборы данных COCO, MP-II Human Pose, Human3.6M и LSP широко используются для оценки алгоритмов НРЕ. Приведем ключевые различия между этими наборами данных.

**Масштаб и сложность:** набор данных COCO является самым большим из четырех наборов данных и содержит более 330 тыс. изображений и аннотаций для 17 ключевых точек человеческого тела. Наборы данных MP-II Human Pose и LSP меньше, примерно 25 тыс. и 2000 изображений соответственно, а также содержат аннотации для 16 и 14 местоположений ключевых точек. Набор данных Human3.6M содержит более 3,6 млн изображений и трехмерных положений человеческого тела, выполняющих различные действия, что делает его самым большим и сложным набором данных из четырех.

**Тип аннотаций:** наборы данных COCO, MP-II Human Pose и LSP включают аннотации для двухмерных положений ключевых точек, а набор данных Human3.6M включает аннотации для трехмерных положений человеческого тела.

**Сложные сценарии:** наборы данных COCO и MPII Human Pose включают изображения с несколькими людьми, окклюзиями и различными условиями освещения. Набор данных LSP включает в себя изображения людей, занимающихся спортом. Набор данных Human3.6M включает изображения, снятые с 15 камер, что делает его сложным эталоном для оценки алгоритмов НРЕ.

**Цель:** набор данных COCO предназначен не для оценки позы человека, а скорее для обнаружения объектов и сегментации. Наборы данных MPII Human Pose, LSP и Human3.6M специально разработаны для оценки позы человека.

Таким образом, каждый из этих наборов данных имеет свои сильные стороны и ограничения, и выбор набора данных зависит от конкретной задачи и требований приложения, таких как потребность в крупномасштабных данных, аннотациях 2D или 3D-поз, сложных сценариях и фокусе на оценку позы человека или другие задачи компьютерного зрения (табл. 3).

Таблица 3

Сравнение наборов данных

Наборы данных	COCO	MP II	Human3.6M	LSP
Классы	16	17	32	15
Формат данных	JSON	MatLAB	Binary	MatLAB
Объем (шт.)	330 000	25 000	3 600 000	2000
Разделение тест./вер.	+	+	-	+

### Заключение

Определение положения человеческого тела – это одна из ключевых задач в области компьютерного зрения и машинного обучения. В обзорной статье были рассмотрены основные методы определения положения человеческого тела, включая нисходящие и многоэтапные подходы, а также приведены основные метрики оценки качества моделей, такие как MPJPE, mAP, PCK и др.

Были рассмотрены также основные наборы данных, используемые для обучения и оценки моделей НРЕ, такие как MPII Human Pose, Human3.6M, COCO и LSP. Каждый набор данных имеет свои особенности, и выбор подхода к определению положения человеческого тела зависит от конкретной задачи и требований приложения.

Были также рассмотрены некоторые конкретные модели, такие как ResNet, OpenPose и DeepCut, их принцип работы, а также преимущества и недостатки. Особое внимание было уделено модели OpenPose, которая на текущий момент является одним из самых точных методов определения положения тела и находит широкое применение в различных областях. В целом определение положения человеческого тела – это сложная задача, которая требует глубоких знаний в области компьютерного зрения и машинного обучения. Однако благодаря появлению новых методов и технологий мы можем ожидать дальнейшего улучшения точности и эффективности алгоритмов определения человеческого тела.

*Список литературы*

1. Andriluka M., Pishchulin L., Gehler P., Schiele B. 2d human pose estimation: New benchmark and state of the art analysis // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014. June.
2. Zhang X., Zhou F., Lin Y. [et al.]. A Hierarchical Deep Learning Architecture for Human Pose Estimation // Computer Vision and Image Understanding. 2017. Vol. 161. P. 1–10. doi: 10.1016/j.cviu.2017.05.011
3. Almeida T., Silva J., Oliveira L. [et al.]. Real-time Human Pose Estimation using Deep Learning in Video Streams for Smart Environments // Sensors. 2019. Vol. 19 (7). P. 1576. doi: 10.3390/s19071576
4. Sun Y., Wang L., Yang Y. [et al.]. A Multi-View Deep Learning Approach for Human Pose Estimation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018. Vol. 40 (7). P. 1627–1635. doi: 10.1109/TPAMI.2017.2747980
5. Felzenszwalb P., Girshick R., McAllester D., Ramanan D. Articulated Human Detection with Flexible Mixtures of Parts // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010. Vol. 32 (10). P. 1972–1981. doi: 10.1109/TPAMI.2009.161
6. Belagiannis V., Zisserman A. Human Pose Estimation using Deep Consensus Voting // IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 7291–7299. doi: 10.1109/CVPR.2017.773
7. Shotton J., Fitzgibbon A., Cook M. [et al.]. Efficient Human Pose Estimation from Single Depth Images // IEEE Conference on Computer Vision and Pattern Recognition. 2013. P. 244–251. doi: 10.1109/CVPR.2013.37
8. Nie X., Liang X., Chen Y. [et al.]. Multi-Stage Human Pose Estimation with Spatial Context // European Conference on Computer Vision. 2018. P. 3–19. doi: 10.1007/978-3-030-01234-2\_1
9. Wei S., Ramakrishna V., Kanade T., Sheikh Y. Multi-Stage Human Pose Estimation via Part Detection and Association // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39 (7). P. 1346–1359. doi: 10.1109/TPAMI.2016.2603409
10. Sun K., Xiao B., Liu D. [et al.]. Multi-Stage Human Pose Estimation with Pose Refinement // IEEE Conference on Computer Vision and Pattern Recognition. 2018. P. 7103–7112. doi: 10.1109/CVPR.2018.00745
11. Ye Y., Rehg J. M. Temporally coherent 3D tracking of human body using structured light // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2010. P. 1–8. doi: 10.1109/CVPR.2010.5540143
12. Tekin B., Rozantsev A., Lepetit V., Fua P. Iterative Least Squares for 3D Human Pose Estimation from Single Depth Images // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016. Vol. 38 (9). P. 1813–1820. doi: 10.1109/TPAMI.2015.2509999
13. Chen C., Wang Y., Wang Z. [et al.]. Iterative Least-Squares Optimization for Multiple Human Pose Estimation // IEEE Transactions on Image Processing. 2018. Vol. 27 (8). P. 3941–3954. doi: 10.1109/TIP.2018.2838740
14. Zhang X., Liang X., Liu S. [et al.]. Iterative Least-Squares Pose Estimation with Deep Neural Network Features // IEEE Transactions on Image Processing. 2019. Vol. 28 (1). P. 289–299. doi: 10.1109/TIP.2018.2868741
15. 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. doi: 10.1109/CVPR.2016.90
16. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation // European Conference on Computer Vision. 2016. doi: 10.1007/978-3-319-46475-6\_5
17. Sun X., Zhang S., Xue Y., Zhang H. Learning Pose Specific Representations by Predicting Different Views // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018. doi: 10.1109/TPAMI.2017.2778404

18. Zhang Y., Li Y., Zhang Y. [et al.]. ResNet-Based Human Pose Estimation with Multi-Scale Feature Fusion // *IEEE Access*. 2020. Vol. 8. P. 201684–201693. doi: 10.1109/ACCESS.2020.3031665
19. Li Y., Zhang Y., Zhang Y. [et al.]. ResNet-Based Human Pose Estimation with Multi-View Consistency // *IEEE Transactions on Image Processing*. 2021. Vol. 30. doi: 10.1109/TIP.2021.3051745
20. Zhang Y., Li Y., Zhang Y. [et al.]. ResNet-Based Human Pose Estimation with Self-Attention and Spatial Transformer Network // *IEEE Transactions on Multimedia*. 2021. Vol. 23. doi: 10.1109/TMM.2021.3072389
21. Yang W., Li S., Ouyang W., Wang H. Learning feature pyramids for human pose estimation // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. P. 931–940.
22. Girdhar R., Fouhey D. F. Detect-and-select: Efficient object detection in videos using keypoint proposals // *Proceedings of the European Conference on Computer Vision*. 2018. P. 20–36.
23. Kendall A., Grimes M., Cipolla R. PoseNet: A convolutional network for real-time 6-DOF camera relocalization // *Proceedings of the IEEE International Conference on Computer Vision*. 2015. P. 2938–2946.
24. Kendall A., Cipolla R. Geometric loss functions for camera pose regression with deep learning // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. P. 2492–2500.
25. Walch F., Hazirbas C., Leal-Taixé L. [et al.]. Image-based localization using LSTMs for structured feature correlation // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. P. 6289–6298.
26. Wei S.-E., Ramakrishna V., Kanade T., Sheikh Y. Convolutional pose machines // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. P. 4724–4732.
27. Cao Z., Simon T., Wei S.-E., Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. P. 1302–1310.
28. Insafutdinov E., Pishchulin L., Andres B. [et al.]. Articulated multi-person tracking in the wild // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. P. 6457–6465.
29. Pishchulin L., Insafutdinov E., Schiele B. DeepCut: Fast training of deep neural networks using large datasets // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017. P. 1–9.
30. Yang W., Izadi Sh., Shotton J. Articulated Human Pose Estimation Using Deep Hierarchical Networks // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014. P. 729–736. doi: 10.1109/CVPR.2014.98
31. Chu X., Yang W., Ouyang W. [et al.]. Deep Learning for Human Pose Estimation: A Survey // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. Vol. 42, № 8. P. 1863–1895. doi: 10.1109/TPAMI.2019.2919907
32. Pishchulin L., Insafutdinov E., Schiele B. DeepCut: Fast training of deep neural networks using large datasets // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017. P. 1–9.
33. Insafutdinov E., Pishchulin L., Andres B. [et al.]. DeeperCut: A deeper, stronger, and faster multi-person pose estimation model // *European Conference on Computer Vision*. 2016. P. 34–50.
34. Pishchulin L., Insafutdinov E., Tang S. [et al.]. Poselet conditioned pictorial structures // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013. P. 588–595. doi: 10.1109/CVPR.2013.82

35. Pishchulin L., Insafutdinov E., Tang S. [et al.]. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 4929–4937.
36. Toshev A., Szegedy C. Deeppose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014. P. 1653–1660. doi: 10.1109/CVPR.2014.213
37. Carreira J., Zisserman A. Human pose estimation with iterative error feedback // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 4733–4742. doi: 10.1109/CVPR.2016.515
38. Wei S. E., Ramakrishna V. Convolutional pose machines // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 4724–4732. doi: 10.1109/CVPR.2016.511
39. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks // Proceedings of the Neural Information Processing Systems (NIPS). 2015. P. 91–99. doi: 10.1007/s11263-015-0816-x
40. Lin T. Y., Maire M., Belongie S. [et al.]. Microsoft COCO: Common objects in context // European conference on computer vision. Springer, Cham., 2014. P. 740–755.
41. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 779–788.
42. Chen T., Li M., Li Y. [et al.]. Microsoft COCO: Common Objects in Context // Proceedings of the European Conference on Computer Vision (ECCV). 2014. P. 740–755. URL: <https://www.microsoft.com/en-us/research/publication/microsoft-coco-common-objects-in-context/> (дата обращения: 15.03.2023).
43. Carreira J., Zisserman A. Human pose estimation with iterative error feedback // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 4733–4742.
44. Sun K., Xiao B., Liu D., Wang J. Deep high-resolution representation learning for human pose estimation // Proceedings of the IEEE conference on computer vision and pattern recognition. 2019. P. 5693–5703.
45. Newell A., Yang K., Deng J. Stacked hourglass networks for human pose estimation // Proceedings of the European Conference on Computer Vision (ECCV). 2016. P. 483–499. doi: 10.1007/978-3-319-46487-9\_28
46. Chen Y., Wang Z., Peng Y. [et al.]. Cascaded pyramid network for multi-person pose estimation // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. P. 7103–7112.
47. Ionescu C., Papava D., Olaru V., Sminchisescu C. Human3.6m: Large scale datasets and predictive methods for 3D human sensing in natural environments // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014. Vol. 36 (7). P. 1325–1339.
48. Theobalt C., Hasler N., Stoll C., Rosenhahn B. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010. P. 1065–1080. doi: 10.1109/TPAMI.2009.155
49. Tekin B., Rozantsev A., Lepetit V., Fua P. Direct prediction of 3D body poses from motion compensated sequences // In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. P. 991–1000.
50. Johnson S., Everingham M. Clustered pose and nonlinear appearance models for human pose estimation // Proceedings of the British Machine Vision Conference (BMVC). 2010. doi: 10.5244/C.24.37

### References

1. Andriluka M., Pishchulin L., Gehler P., Schiele B. 2d human pose estimation: New benchmark and state of the art analysis. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014;June.
2. Zhang X., Zhou F., Lin Y. et al. A Hierarchical Deep Learning Architecture for Human Pose Estimation. *Computer Vision and Image Understanding*. 2017;161:1–10. doi: 10.1016/j.cviu.2017.05.011
3. Almeida T., Silva J., Oliveira L. et al. Real-time Human Pose Estimation using Deep Learning in Video Streams for Smart Environments. *Sensors*. 2019;19(7):1576. doi: 10.3390/s19071576
4. Sun Y., Wang L., Yang Y. et al. A Multi-View Deep Learning Approach for Human Pose Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018;40(7):1627–1635. doi: 10.1109/TPAMI.2017.2747980
5. Felzenszwalb P., Girshick R., McAllester D., Ramanan D. Articulated Human Detection with Flexible Mixtures of Parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010;32(10):1972–1981. doi: 10.1109/TPAMI.2009.161
6. Belagiannis V., Zisserman A. Human Pose Estimation using Deep Consensus Voting. *IEEE Conference on Computer Vision and Pattern Recognition*. 2017:7291–7299. doi: 10.1109/CVPR.2017.773
7. Shotton J., Fitzgibbon A., Cook M. et al. Efficient Human Pose Estimation from Single Depth Images. *IEEE Conference on Computer Vision and Pattern Recognition*. 2013:244–251. doi: 10.1109/CVPR.2013.37
8. Nie X., Liang X., Chen Y. et al. Multi-Stage Human Pose Estimation with Spatial Context. *European Conference on Computer Vision*. 2018:3–19. doi: 10.1007/978-3-030-01234-2\_1
9. Wei S., Ramakrishna V., Kanade T., Sheikh Y. Multi-Stage Human Pose Estimation via Part Detection and Association. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;39(7):1346–1359. doi: 10.1109/TPAMI.2016.2603409
10. Sun K., Xiao B., Liu D. et al. Multi-Stage Human Pose Estimation with Pose Refinement. *IEEE Conference on Computer Vision and Pattern Recognition*. 2018:7103–7112. doi: 10.1109/CVPR.2018.00745
11. Ye Y., Rehg J.M. Temporally coherent 3D tracking of human body using structured light. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2010:1–8. doi: 10.1109/CVPR.2010.5540143
12. Tekin B., Rozantsev A., Lepetit V., Fua P. Iterative Least Squares for 3D Human Pose Estimation from Single Depth Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016;38(9):1813–1820. doi: 10.1109/TPAMI.2015.2509999
13. Chen C., Wang Y., Wang Z. et al. Iterative Least-Squares Optimization for Multiple Human Pose Estimation. *IEEE Transactions on Image Processing*. 2018;27(8):3941–3954. doi: 10.1109/TIP.2018.2838740
14. Zhang X., Liang X., Liu S. et al. Iterative Least-Squares Pose Estimation with Deep Neural Network Features. *IEEE Transactions on Image Processing*. 2019;28(1):289–299. doi: 10.1109/TIP.2018.2868741
15. 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. doi: 10.1109/CVPR.2016.90
16. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation. *European Conference on Computer Vision*. 2016. doi: 10.1007/978-3-319-46475-6\_5
17. Sun X., Zhang S., Xue Y., Zhang H. Learning Pose Specific Representations by Predicting Different Views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018. doi: 10.1109/TPAMI.2017.2778404

18. Zhang Y., Li Y., Zhang Y. et al. ResNet-Based Human Pose Estimation with Multi-Scale Feature Fusion. *IEEE Access*. 2020;8:201684–201693. doi: 10.1109/ACCESS.2020.3031665
19. Li Y., Zhang Y., Zhang Y. et al. ResNet-Based Human Pose Estimation with Multi-View Consistency. *IEEE Transactions on Image Processing*. 2021;30. doi: 10.1109/TIP.2021.3051745
20. Zhang Y., Li Y., Zhang Y. et al. ResNet-Based Human Pose Estimation with Self-Attention and Spatial Transformer Network. *IEEE Transactions on Multimedia*. 2021;23. doi: 10.1109/TMM.2021.3072389
21. Yang W., Li S., Ouyang W., Wang H. Learning feature pyramids for human pose estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:931–940.
22. Girdhar R., Fouhey D.F. Detect-and-select: Efficient object detection in videos using keypoint proposals. *Proceedings of the European Conference on Computer Vision*. 2018:20–36.
23. Kendall A., Grimes M., Cipolla R. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. *Proceedings of the IEEE International Conference on Computer Vision*. 2015:2938–2946.
24. Kendall A., Cipolla R. Geometric loss functions for camera pose regression with deep learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:2492–2500.
25. Walch F., Hazirbas C., Leal-Taixé L. et al. Image-based localization using LSTMs for structured feature correlation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:6289–6298.
26. Wei S.-E., Ramakrishna V., Kanade T., Sheikh Y. Convolutional pose machines. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:4724–4732.
27. Cao Z., Simon T., Wei S.-E., Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:1302–1310.
28. Insafutdinov E., Pishchulin L., Andres B. et al. Articulated multi-person tracking in the wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:6457–6465.
29. Pishchulin L., Insafutdinov E., Schiele B. DeepCut: Fast training of deep neural networks using large datasets. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017:1–9.
30. Yang W., Izadi Sh., Shotton J. Articulated Human Pose Estimation Using Deep Hierarchical Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014:729–736. doi: 10.1109/CVPR.2014.98
31. Chu X., Yang W., Ouyang W. et al. Deep Learning for Human Pose Estimation: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020;42(8):1863–1895. doi: 10.1109/TPAMI.2019.2919907
32. Pishchulin L., Insafutdinov E., Schiele B. DeepCut: Fast training of deep neural networks using large datasets. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017:1–9.
33. Insafutdinov E., Pishchulin L., Andres B. et al. DeeperCut: A deeper, stronger, and faster multi-person pose estimation model. *European Conference on Computer Vision*. 2016:34–50.
34. Pishchulin L., Insafutdinov E., Tang S. et al. Poselet conditioned pictorial structures. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013:588–595. doi: 10.1109/CVPR.2013.82

35. Pishchulin L., Insafutdinov E., Tang S. et al. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015:4929–4937.
36. Toshev A., Szedgy C. Deeppose: Human pose estimation via deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014:1653–1660. doi: 10.1109/CVPR.2014.213
37. Carreira J., Zisserman A. Human pose estimation with iterative error feedback. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:4733–4742. doi: 10.1109/CVPR.2016.515
38. Wei S.E., Ramakrishna V. Convolutional pose machines. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:4724–4732. doi: 10.1109/CVPR.2016.511
39. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the Neural Information Processing Systems (NIPS)*. 2015:91–99. doi: 10.1007/s11263-015-0816-x
40. Lin T.Y., Maire M., Belongie S. et al. Microsoft COCO: Common objects in context. *European conference on computer vision*. Springer, Cham., 2014:740–755.
41. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:779–788.
42. Chen T., Li M., Li Y. et al. Microsoft COCO: Common Objects in Context. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2014:740–755. Available at: <https://www.microsoft.com/en-us/research/publication/microsoft-coco-common-objects-in-context/> (accessed 15.03.2023).
43. Carreira J., Zisserman A. Human pose estimation with iterative error feedback. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:4733–4742.
44. Sun K., Xiao B., Liu D., Wang J. Deep high-resolution representation learning for human pose estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2019:5693–5703.
45. Newell A., Yang K., Deng J. Stacked hourglass networks for human pose estimation. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2016:483–499. doi: 10.1007/978-3-319-46487-9\_28
46. Chen Y., Wang Z., Peng Y. et al. Cascaded pyramid network for multi-person pose estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018:7103–7112.
47. Ionescu C., Papava D., Olaru V., Sminchisescu C. Human3.6m: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2014;36(7):1325–1339.
48. Theobalt C., Hasler N., Stoll C., Rosenhahn B. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010:1065–1080. doi: 10.1109/TPAMI.2009.155
49. Tekin B., Rozantsev A., Lepetit V., Fua P. Direct prediction of 3D body poses from motion compensated sequences. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018:991–1000.
50. Johnson S., Everingham M. Clustered pose and nonlinear appearance models for human pose estimation. *Proceedings of the British Machine Vision Conference (BMVC)*. 2010. doi: 10.5244/C.24.37



***Информация об авторах / Information about the authors***

**Артём Евгеньевич Павликов**

ассистент кафедры математической  
кибернетики и информационных  
технологий,  
Московский технический университет  
связи и информатики  
(Россия, г. Москва, ул. Авиамоторная, 8)  
E-mail: a.e.pavlikov@mtuci.ru

**Artem E. Pavlikov**

Assistant of the sub-department  
of mathematical cybernetics  
and information technologies,  
Moscow Technical University  
of Communications and Informatics  
(8 Aviamotornaya street, Moscow, Russia)

**Михаил Геннадьевич Городничев**

кандидат технических наук, доцент,  
декан факультета информационных  
технологий,  
Московский технический университет  
связи и информатики  
(Россия, г. Москва, ул. Авиамоторная, 8)  
E-mail: m.g.gorodnichev@mtuci.ru

**Mikhail G. Gorodnichev**

Candidate of technical sciences, associate  
professor, dean of the faculty  
of information technology,  
Moscow Technical University  
of Communications and Informatics  
(8 Aviamotornaya street, Moscow, Russia)

**Авторы заявляют об отсутствии конфликта интересов /  
The authors declare no conflicts of interests.**

**Поступила в редакцию/Received 10.04.2023**

**Поступила после рецензирования/Revised 24.05.2023**

**Принята к публикации/Accepted 04.07.2023**