

УДК 004.93'1+77.03.03

 10.25209/2079-3316-2025-16-4-23-50

## Использование доменной адаптации для задачи распознавания ключевых точек на теле человека

Андрей Сергеевич **Токарев**<sup>1✉</sup>, Илья Михайлович **Воронков**<sup>2</sup>

<sup>1,2</sup> Московский физико-технический институт, Москва, Россия

<sup>✉</sup>[tokarev.as@phystech.su](mailto:tokarev.as@phystech.su)

**Аннотация.** В статье проводится исследование алгоритмов доменной адаптации для задачи распознавания ключевых точек на теле человека с целью применения в спорте, когда требуется повысить точность распознавания и сократить трудоемкость ручной разметки данных. Результатом работы является алгоритм итеративной адаптации модели на собственных псевдометках. Экспериментально показано, что метод позволяет получить более эффективную итоговую нейросетевую модель в сравнении с обычным дообучением.

*(Здесь только русскоязычная часть оригинальной двуязычной статьи)*

**Ключевые слова и фразы:** Ключевые точки, оценка позы человека, доменная адаптация без учителя

Для цитирования: Токарев А. С., Воронков И. М. *Использование доменной адаптации для задачи распознавания ключевых точек на теле человека* // Программные системы: теория и приложения. 2025. Т. 16. № 4(67). С. 23–50. [https://psta.psir.ru/read/psta2025\\_4\\_23-50.pdf](https://psta.psir.ru/read/psta2025_4_23-50.pdf)

## **Введение**

В последние годы среди стремительно развивающихся направлений машинного обучения и компьютерного зрения выделяется активно применяемая в спорте оценка позы человека (Human Pose Estimation), основанная на распознавании ключевых точек на теле человека (Keypoint Detection). На ее основе создана система полуавтоматического определения офсайда, которая применяется на международных футбольных соревнованиях с 2022 года.

Большое количество видов спорта позволяет в полной мере раскрыть потенциал технологии оценки позы человека для решения сложных динамических задач на основе видеоаналитики, а также в ходе соревнований в режиме онлайн. Для каждой отдельной постановки задачи имеет большое значение точность оценки позы человека, а также необходимо учитывать специфику видов спорта. Эффективность технологии немыслима без большого количества качественно размеченных наборов данных для последующего обучения нейросетевых моделей.

Среди обилия видеозаписей спортивных мероприятий крайне мало размеченных людьми, так как это требует достаточно много времени. Методы доменной адаптации без учителя позволяют обойтись без разметки данных, при этом улучшать качество работы модели. Эти методы позволяют преодолеть разрыв между исходными и целевыми данными путем выравнивания распределений данных, адаптации признаков и применения обученных моделей к новому домену с минимальной дополнительной обработкой. Таким образом, модели становятся более гибкими и способны эффективно работать в различных условиях, не требуя значительного объема новых данных или переработки архитектуры. В конечном итоге использование методов доменной адаптации не только ускоряет процесс внедрения, но и снижает затраты на разработку, поскольку уменьшает необходимость в полном цикле обучения новой модели.

## **1. Смежные исследования**

### **1.1. Оценка позы человека**

Для моделей оценки позы необходимо определить топологию точек, которые и будут эту позу составлять. Исторически поза определялась

датасетом, на котором модель обучается, но с 2015 года наиболее распространенной стала 17 точечная топология, предложенная исследователями из Microsoft [1]. Все более новые описания «скелетов» являются ее расширениями и уточняют положение конечностей.

На сегодняшний день в мире распознавания ключевых точек на теле человека существуют два подхода к решению задач: нисходящий (top-down) и восходящий (bottom-up). Для первого подхода понадобится дополнительный детектор, который локализует человека на изображении. Потом на локализованной области распознаются необходимые ключевые точки и собирается скелет. Это обеспечивает высокую точность предсказания, но делает архитектуру уязвимой к ошибкам на уровне обнаружения человека. Второй подход, наоборот, распознает сразу все ключевые точки на изображении, а уже после этого решает задачу корректного распределения точек по «скелетам». Модели, использующие данный подход, требуют на вход изображение высокого разрешения, что увеличивает потребление ресурсов для их обучения.

Из-за ограниченности в ресурсах в исследовании рассматривались модели top-down подхода, такие как алгоритм Simple Coordinate Classification (SimCC) [2] с нейронной сетью ResNet, в качестве магистральной части, и модель RTMPose [3]. Обе архитектуры описаны в проекте MMPose от OpenMMLab [4], что позволило универсализировать код для процесса обучения и адаптации разных моделей.

## 1.2. Доменная адаптация

Доменная адаптация является частным случаем задачи переноса обучения (transfer learning): когда постановка задачи остается постоянной, а изменяется только набор данных (домен). При наличии только сырых данных без аннотаций принято использовать доменную адаптацию без учителя. Она позволяет модели, обученной на большом и универсальном домене, сохранить высокую точность предсказанных результатов и производительность на неизвестном ранее датасете, избегая затрат на дополнительное обучение. Таким образом создается возможность быстрой настройки нейронной сети для работы над узкоспециализированной задачей, сохраняя при этом колоссальные ресурсы на анализ и разметку новых данных.

Существующие методы адаптации моделей оценки позы человека на неразмеченном домене используют дополнительные нейронные сети, такие как состязательный перцептор в Regressive Domain Adaptation

[5] или модели переноса стиля и алгоритм среднего учителя в Unified Framework [6]. Это повышает вычислительную сложность метода, а также требует манипуляций со внутренними слоями архитектуры, что усложняет программную реализацию.

Также есть более простой, но эффективный метод Progressive Unsupervised Learning [7], который работает без добавления дополнительных нейронных сетей. Основная идея PUL состоит в итеративном дообучении модели на лучших из своих предсказаний и позволяет повысить точность работы. В таком случае предсказания модели называются псевдо-разметкой или набором псевдометок, а отбор лучших осуществляется выбранной функцией фильтрации, а задача сводится к выбору такой функции, чтобы алгоритм не оказался вычислительно сложным, и нейросеть демонстрировала улучшение результатов работы.

Ниже мы показываем, что для задачи распознавания ключевых точек на теле человека существует функция фильтрации, которая позволяет использовать метод PUL для адаптации нейросетевых моделей к новым доменам. Функция позволяет отсеивать шумные предсказания с помощью показателя уверенности модели в своих предсказаниях и регулируемого порогового значения. Все предсказания, для которых средняя уверенность по скелету переходит порог, попадают в набор псевдометок, на которых в текущей итерации модель дообучается.

## 2. Эксперимент

Практическая часть исследования включала несколько этапов, таких как подготовка данных, обучение моделей и последующее сравнение их результатов в различных состояниях. На первом этапе осуществлялась подготовка данных, которая включала сбор, обработку и частичную разметку. Далее на основе подготовленных данных проводились адаптация и обучение моделей. После завершения процесса обучения проведены анализ и сравнение полученных результатов, чтобы определить эффективность метода доменной адаптации относительно обычного обучения модели.

### 2.1. Система полуавтоматической разметки данных

Для аннотирования набора видеозаписей из целевого домена данных была создана система полуавтоматической разметки «pose-markup» (<https://github.com/andrwtokar/pose-markup>). Пример работы программы можно увидеть на рисунке 1.



Рисунок 1. Пример работы системы полуавтоматической разметки данных

Основой системы стала идея автоматической разметки с помощью модели распознавания ключевых точек на теле человека BlazePose [8] с последующей корректировкой результата экспертом. Для первой части использовалось публичное API на языке Python, предоставленное проектом MediaPipe. Правка результатов предсказаний производилась с помощью инструментов обработки изображений, предоставляемых библиотекой OpenCV.

В качестве базовой топологии была выбрана 17-точечная топология COCO [1]. Для каждого кадра аннотация записывается в формате JSON с возможностью последующей агрегации результатов нескольких кадров в один JSON файл.

Анализ разметки данных, выполненной системой, показал насколько это сокращает количество затрачиваемых ресурсов. Было замечено, что в среднем на кадре изменялось 6.5 точек из 17. Причем чаще всего приходилось править координаты точек руки, расположенной дальше от камеры, и головы. Более детальное распределение произведенных изменений можно увидеть на рисунке 2.

Время автоматической разметки видеозаписи с 420 кадрами заняло порядка 2 минут. Корректировка точек занимала в среднем 3 минуты на один кадр. Суммарно на разметку 420 кадров целевых данных было потрачено 24 часа работы. При разметке 2500 кадров это заняло бы более 120 часов. Хотя система и автоматизировала часть работы, мы все равно получаем большие затраты ресурсов на разметку данных.

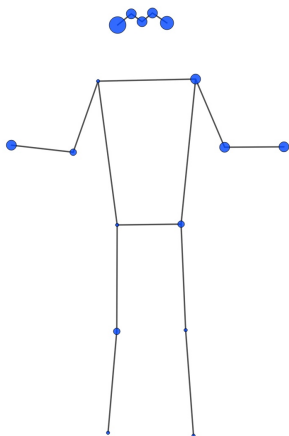


РИСУНОК 2. Схематическое представление изменений в точках, внесенных экспертом

## 2.2. Данные

Для проведения исследования было выбрано два набора данных: универсальный, на котором базово будет обучена модель, и целевой, к которому мы будем адаптироваться.

В качестве первого из них был выбран датасет COCO [1]. Из-за его большого объема и ограниченности в ресурсах для обучения было принято решение уменьшить объем. Из более чем 250 тысяч поз были выбраны только те, в которых представлены все 17 ключевых точек, согласующихся с топологией COCO. В итоге получен универсальный набор, состоящий из 8 тысяч аннотированных поз, представленных на 6 тысячах изображений.

Второй набор данных был создан вручную в небольшом объеме. Для этого было снято два видео одиночных тренировок спортсменов боксеров по 2500 кадров каждое. Первое использовалось для итеративного метода адаптации модели. Из второго видео были выбраны 420 кадров, которые в дальнейшем были размечены с помощью созданной для этого системы pose-markup и использовались для дообучения модели и снятия метрик. Пример полученных данных можно видеть на рисунке 3.



Рисунок 3. Примеры изображений целевого домена

### 2.3. Метрики

В качестве базовой метрики в статьях про оценку позы используют метрику average precision (AP) на основе метрики object keypoint similarity (OKS). Однако для нее необходимо знать стандартное отклонение ключевых точек для выбранного датасета, что требует большого объема выборки и ресурсов для сбора статистических показателей.

Из множества похожих метрик, которые оценивают ошибку между предсказанной и реальной точкой, выделяется метрика percentage of correct keypoints (РСК). Преимущество данной метрики в том, что она может быть применима как в случае двумерных координат, так и в случае трехмерных. Ошибка представляется как евклидово расстояние между точками и сравнивается с пороговой величиной. Математическая формула выглядит следующим образом:

$$(1) \quad PCK = \frac{\sum_{i=1}^n \text{bool}(d_i < \text{threshold} * \text{body\_height})}{n},$$

где

$d_i$  — расстояние между предсказанной и правильной точкой,

$\text{threshold}$  — порог, задаваемый исследователем,

$\text{body\_height}$  — высота прямоугольника, внутри которого находится человек,

$\text{bool}(\ast)$  — логическое условие, возвращает 1, если оно верно и 0 в ином случае,

$n$  — размер выборки.

В рамках эксперимента использовалась метрика РСК с пороговыми значениями равными 0.05, 0.25 и 0.5.

## 2.4. Полученные результаты

В рамках эксперимента производилось сравнение трех состояний алгоритма:

- (1) Базовый случай. Когда модель обучена только на универсальном датасете.
- (2) Адаптированное состояние. К базовой модели применили 10 итераций алгоритма доменной адаптации. Схема адаптации моделей представлена на рисунке 4.

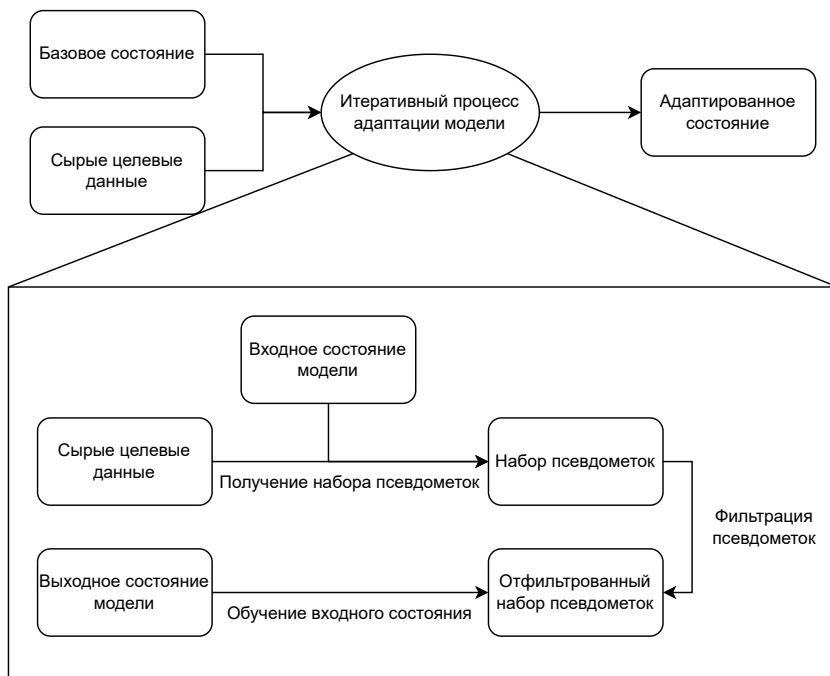


Рисунок 4. Схема адаптации моделей

- (3) Дообученное состояние. Базовую модель дополнительно обучили на размеченных данных из целевого домена. Схема дополнительного обучения моделей представлена на рисунке 5.

Описанная выше функция фильтрации псевдометок подразумевает получение значения уверенности в предсказанной позе. Это значение получается путем усреднения значения уверенности по всем 17 точкам топологии.

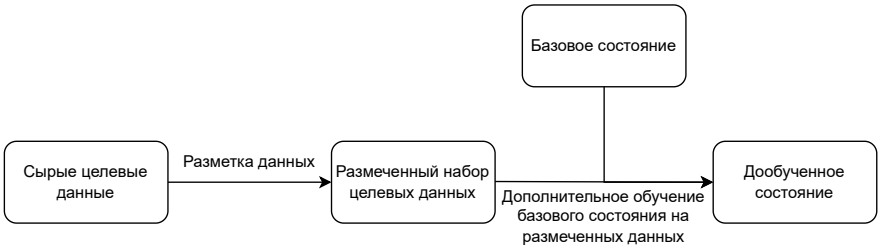


Рисунок 5. Схема дополнительного обучения моделей

Данный способ на первых итерациях отсеивал более 75 процентов данных, используемых для адаптации. На 10 итерациях количество шумных поз было близким к нулю, что говорит об улучшении качества предсказания модели с увеличением количества итераций.

Остановка итеративного процесса производилась после фиксированного числа из 10 шагов. Если бы шагов было выбрано больше, то останавливать итеративный процесс можно либо по достижении заранее выбранной точности, либо при отсутствии изменений в отфильтрованном наборе псевдометок.

Обучение нейронных сетей производилось на платформе Google Colaboratory с использованием графического ускорителя Tesla T4 с 15 Гб памяти. Оценка метрики производилась локально на ноутбуке с графическим ускорителем NVidia MX250 с 2 Гб памяти.

В таблице представлены результаты, полученные для алгоритма SimCC с порогом уверенности равным 0.6. По результатам в таблице 1

Таблица 1. Сравнительная статистика нескольких состояний для алгоритма SimCC

|                | PCK@0.05 | PCK@0.25 | PCK@0.5 |
|----------------|----------|----------|---------|
| Baseline       | 0.024    | 0.429    | 0.758   |
| Adapted model  | 0.037    | 0.445    | 0.792   |
| Finetune model | 0.0062   | 0.411    | 0.76    |

видно улучшение метрики PCK не только относительно базового состояния, но и относительно дополнительно обученного. При этом стоит заметить, что суммарно на адаптацию было потрачено не более 80 минут, что значительно меньше, чем на разметку тех же данных.

Базовая версия RTMPose показала результаты значительно лучше предыдущей модели. Из таблицы 2 видно, что среди всего эксперимента,

Таблица 2. Сравнительная статистика нескольких состояний для модели RTMPose

|                | PCK@0.05 | PCK@0.25 | PCK@0.5 |
|----------------|----------|----------|---------|
| Baseline       | 0.282    | 0.882    | 0.976   |
| Adapted model  | 0.145    | 0.652    | 0.709   |
| Finetune model | 0.076    | 0.709    | 0.922   |

Baseline дает наилучшие результаты.

Чтобы объяснить причины ошибок, построим гистограмму распределения верных предсказаний по точкам топологии. Рисунки 6 и 7 показывают, что количество правильных координат точек верхних конечностей падает практически до нуля и вносит большую ошибку в значение метрики. Также падает качество предсказания для точек лодыжек. Описанные точки у боксеров чаще других находятся в движении и являются размытыми на некоторых кадрах видеозаписи.

В текущей версии функции фильтрации эти точки могли попасть в обучающую выборку на выбранной итерации из-за механизма фильтрации и усреднения уверенности по всем точкам для получения уверенности в позе.

Чтобы сравнить работу алгоритма без учета размытия, исключим данные 6 точек из результатов. Как видно из таблицы 3, адаптированная

Таблица 3. Результаты предсказания модели RTMPose без учета размытых ключевых точек

|               | PCK@0.25 | PCK@0.5 |
|---------------|----------|---------|
| Baseline      | 0.944    | 0.988   |
| Adapted model | 0.964    | 1.0     |

модель показывает более высокое качество предсказаний относительно базового случая. В дальнейшем, чтобы побороть ошибки, вносимые размытыми частями изображения, необходимо модифицировать функцию фильтрации.

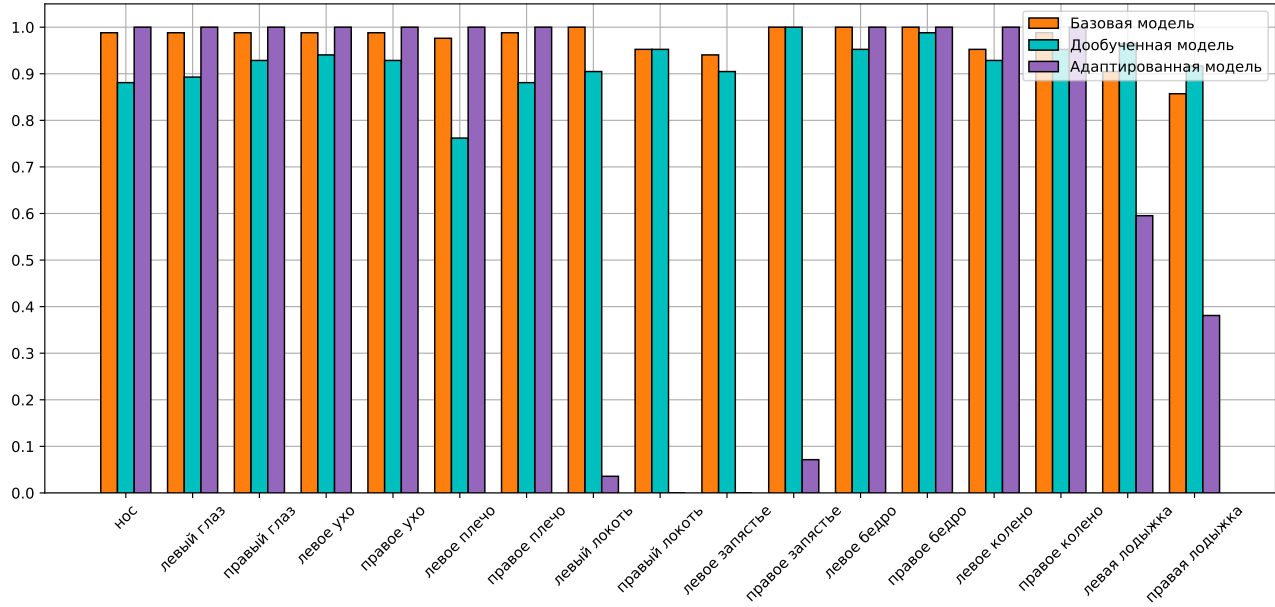


Рисунок 6. Распределение корректно предсказанных точек по всей топологии для модели RTMPose, аналог PCK@0.5

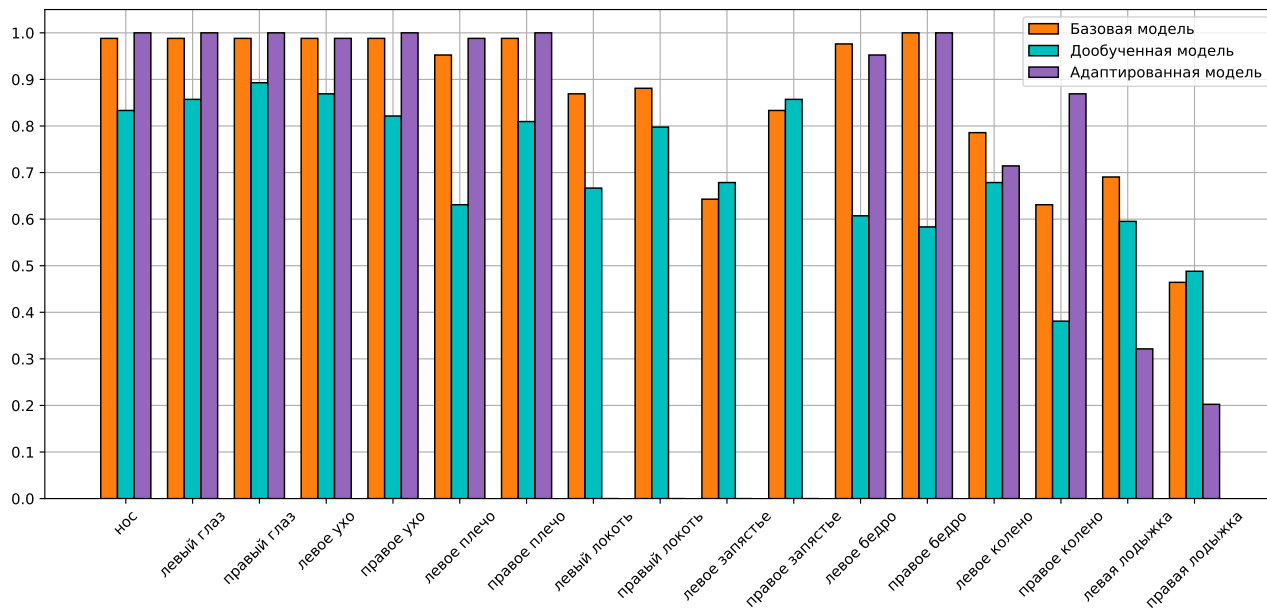


Рисунок 7. Распределение корректно предсказанных точек по всей топологии для модели RTMPose, аналог PCK@0.25

## Заключение

В результате проведенного исследования была показана эффективность метода доменной адаптации без учителя Progressive Unsupervised Learning, основанного на генерации и фильтрации псевдоразмеченных данных, для моделей оценки позы. Полученные результаты можно применять на практике в спортивных системах реального времени. За счет выбора функции фильтрации данный метод не только экономит много ресурсов, ранее затрачиваемых на разметку, но и показывает улучшение качества предсказания в сравнении с другими подходами по обучению аналогичной модели.

Предложенный подход использования доменной адаптации существенно сокращает время на разработку решения задач под конкретные цели. Поэтому возможно продолжать исследования более сложных функций фильтрации, которые повысят эффективность модели.

## Список использованных источников

- [1] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. Lawrence Zitnick *Microsoft COCO: Common objects in context*, Computer Vision — ECCV 2014 (Zurich, Switzerland, September 6–12, 2014), Lecture Notes in Computer Science.— vol. **8693**, Cham: Springer.— 2014.— ISBN 978-3-319-10602-1.— Pp. 740–755.  <sup>↑38, 40, 41</sup>
- [2] Y. Li, S. Yang, P. Liu, S. Zhang, Y. Wang, Z. Wang, W. Yang, S.-T. Xia *SimCC: A simple coordinate classification perspective for human pose estimation*, Computer Vision — ECCV 2022 (Tel Aviv, Israel, October 23–27, 2022), Lecture Notes in Computer Science.— vol. **13666**, Cham: Springer.— 2022.— ISBN 978-3-031-20068-7.— Pp. 89–106.  <sup>↑38</sup>
- [3] T. Jiang, P. Lu, L. Zhang, N. Ma, R. Han, C. Lyu, Y. Li, K. Chen *RTMPose: Real-time multi-person pose estimation based on MMPose*.— 2023.— 11 pp. [arXiv:2303.07399](https://arxiv.org/abs/2303.07399) [cs.CV] <sup>↑38</sup>
- [4] MMPose Contributors *OpenMMLab pose estimation toolbox and benchmark*, <https://github.com/open-mmlab/mmpose>.— 2020. <sup>↑38</sup>
- [5] J. Jiang, Y. Ji, X. Wang, Y. Liu, J. Wang, M. Long *Regressive domain adaptation for unsupervised keypoint detection // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR 2021 (Nashville, TN, USA, June 20–25, 2021).— IEEE.— 2021.— ISBN 978-1-6654-4509-2.— Pp. 6776–6785.  <sup>↑39</sup>
- [6] D. Kim, K. Wang, K. Saenko, M. Betke, S. Sclaroff *A unified framework for domain adaptive pose estimation*, Computer Vision — ECCV 2022 (Tel Aviv, Israel, October 23–27, 2022), Lecture Notes in Computer Science.— vol. **13693**, Cham: Springer.— 2022.— ISBN 978-3-031-19827-4.— Pp. 603–620.  <sup>↑39</sup>

- [7] Q. Wu, J. Wan, A. B. Chan *Progressive unsupervised learning for visual object tracking* // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021 (Nashville, TN, USA, June 20–25, 2021).– IEEE.– 2021.– ISBN 978-1-6654-4509-2.– Pp. 2993–3002. doi ↑39
- [8] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, M. Grundmann *BlazePose: On-device real-time body pose tracking.*– 2020.– 4 pp. arXiv:2006.10204 [cs.CV] ↑40

Поступила в редакцию 20.12.2024;  
одобрена после рецензирования 16.01.2025;  
принята к публикации 07.04.2025;  
опубликована онлайн 30.08.2025.

Рекомендовал к публикации

к.т.н. Е. П. Куршев

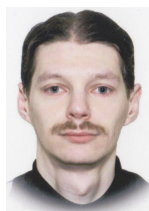
### Информация об авторах:



Андрей Сергеевич Токарев

Магистр МФТИ. Научные интересы: применение нейросетей для анализа изображений и видео, технологии обработки данных.

e-mail: tokarev.as@phystech.su



Илья Михайлович Воронков

старший преподаватель кафедры интеллектуальных информационных систем и технологий МФТИ. Научные интересы: разработка и применение нейросетей в практических задачах, анализ и обработка изображений, видео, текстовых данных, параметических данных, в том числе для применения в задачах робототехники, в задачах кибербезопасности.

e-mail: voronkov.im@phystech.su

Вклад авторов: А. С. Токарев – 75% (идея, программное обеспечение, валидация, формальный анализ, проведение экспериментов, сбор материала, курирование данных, написание черновой версии, доработка и редактирование, визуализация); И. М. Воронков – 25% (идея, методология, доработка и редактирование, наставничество).

Декларация об отсутствии личной заинтересованности: благополучие авторов не зависит от результатов исследования.