

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ВСЕРОССИЙСКАЯ ГРУППА ТЕОРИИ ИНФОРМАЦИИ ИЕЕЕ
АКАДЕМИЯ ИНФОРМАТИЗАЦИИ ОБРАЗОВАНИЯ
ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ООО «ОТКРЫТЫЕ РЕШЕНИЯ»
ОБЩЕСТВО «ЗНАНИЕ» РОССИИ
ПРИВОЛЖСКИЙ ДОМ ЗНАНИЙ

*XXII Международная
научно-техническая конференция*

**ПРОБЛЕМЫ ИНФОРМАТИКИ
В ОБРАЗОВАНИИ, УПРАВЛЕНИИ,
ЭКОНОМИКЕ И ТЕХНИКЕ**

Сборник статей

Декабрь 2022 г.

Пенза

УДК 004
ББК 32.81я43+74.263.2+65.050.2я43
П781

П781 **ПРОБЛЕМЫ ИНФОРМАТИКИ В ОБРАЗОВАНИИ,
УПРАВЛЕНИИ, ЭКОНОМИКЕ И ТЕХНИКЕ :**
сборник статей XXII Международной научно-технической
конференции. – Пенза: Приволжский Дом знаний, 2022. – 356 с.

ISBN 978-5-8356-1800-2
ISSN 2311-0406

Под редакцией В.И. Горбаченко, доктора технических наук,
профессора;
В.В. Дрождина, кандидата технических наук,
профессора

Информация об опубликованных статьях предоставлена в систему Рос-
сийского индекса научного цитирования (РИНЦ) по договору
№ 573-03/2014К от 18.03.2014.

ISBN 978-5-8356-1800-2
ISSN 2311-0406

© Пензенский государственный
университет, 2022
© АННМО «Приволжский Дом знаний», 2022

*XXII International
scientific and technical conference*

**PROBLEMS OF INFORMATICS
IN EDUCATION, MANAGEMENT,
ECONOMICS AND TECHNICS**

December, 2022

Penza

ДЕТЕКТИРОВАНИЕ И ТРЕКИНГ ИГРОКОВ В СИСТЕМАХ СПОРТИВНОЙ ВИДЕОАНАЛИТИКИ

В.В. Хрящев, Д.В. Матвеев, Л.И. Ивановский, М.Н. Голубев

PLAYERS DETECTION AND TRACKING IN SPORTS VIDEO ANALYTICS SYSTEMS

V.V. Khryashchev, D.V. Matveev, L.I. Ivanovsky, M.N. Golubev

Аннотация. Представлены результаты исследования алгоритма JDE, предназначенного для решения задачи обнаружения и отслеживания людей на видеоданных в режиме реального времени. Численные эксперименты по оценке качества работы разработанной модели были проведены на изображениях, извлеченных из видеофайлов базы данных MOT-16 Challenge. Критерием оценки качества работы алгоритма служила метрика MOTA. Качество работы модели JDE проверялось также и на собранной базе видеозаписей баскетбольных матчей, снятых на купольную камеру GV-EVD3100. В результате обучения и тестирования модели JDE для набора видеофайлов с конкурса MOT-16 Challenge значение метрики MOTA составило 62,6%.

Ключевые слова: спортивная видеоаналитика, детектирование объектов, трекинг, глубокое обучение.

Abstract. The results of a study of the JDE algorithm designed to solve the problem of detecting and tracking people on video data in real time are presented. Numerical experiments to assess the quality of the developed model were carried out on images extracted from the video files of the MOT-16 Challenge database. The criterion for assessing the quality of the algorithm was the MOTA metric. The quality of the JDE model was also tested on the collected base of video recordings of basketball games filmed with the GV-EVD3100 dome camera. As a result of training and testing the JDE model for a set of video files from the MOT-16 Challenge, the value of the MOTA metric was 62.6%.

Key words: sport video analytics, object detection, tracking, deep learning.

На сегодняшний день, в области компьютерного зрения одним из ключевых направлений исследований являются алгоритмы обнаружения и отслеживания объектов на видеоданных. Под детектированием объектов понимается определение их местоположений на изображении или видеокадре. Основным отличием алгоритма отслеживания от модели обнаружения

объектов является тот факт, что после обработки первого видеокadra, на котором были обнаружены целевые объекты, на последующих видеокadraх поиск прямоугольных рамок, окружающих целевые объекты, осуществляется в окрестности их местоположений [1]. И хотя человек без особого труда способен отслеживать спортсменов во время трансляций матчей, для компьютера детектирование и трекинг являются трудновыполнимыми задачами.

В данной работе представлены результаты исследования работы сверточной нейронной сети для решения задачи обнаружения и отслеживания игроков на трансляциях баскетбольных матчей. Задача обнаружения и отслеживания людей, в частности спортсменов, на видеоданных решалась с помощью алгоритма JDE (JointDetectionandEmbedding) [2]. В качестве базовой архитектуры для него использовалась одна из разновидностей пирамидальной сети FPN (Feature Pyramid Network) – DarkNet-53 [3]. Сети FPN используют пирамидальную архитектуру для последовательного выделения признаков разного масштаба [4], что позволяет улучшить качество обнаружения объектов, размер которых варьируется на видеокadraх.

При прямом проходе входного изображения по пирамидальной сети из нее извлекаются карты признаков в трех масштабах: $1/32$, $1/16$, $1/8$ соответственно. Далее признаки с наименьшим масштабом после применения операции повышения дискретизации объединяются с признаками большего масштаба с помощью пропускного соединения, то же самое происходит и с картами признаков большего масштаба. Далее к объединенным картам признаков всех трех масштабов добавляются небольшие сверточные сети. В конце осуществляется этап прогнозирования, при выполнении которого выделяется три набора данных:

1. Результаты классификации объектов, находящихся в окантовочных прямоугольных рамках.
2. Коэффициенты регрессии окантовочных прямоугольных рамок для определения на изображении их точного местоположения и формы.
3. Карта встраивания объектов, для ассоциации выделенных прямоугольных рамок с траекториями движения.

На рис. 1 изображена схема архитектуры алгоритма JDE.

Для обучения алгоритма JDE была собрана коллекция из четырех общедоступных наборов данных по обнаружению и отслеживанию пешеходов: CalTech Pedestrian [5], CUHK-SYSU [6], PRW [7] и MOT-16 Challenge [8]. Во избежание переобучения алгоритма, к обучающим данным модели применялись методы аугментации: случайное вращение и масштабирование изображения, а также случайное изменение яркости, контрастности и насыщенности кадра.

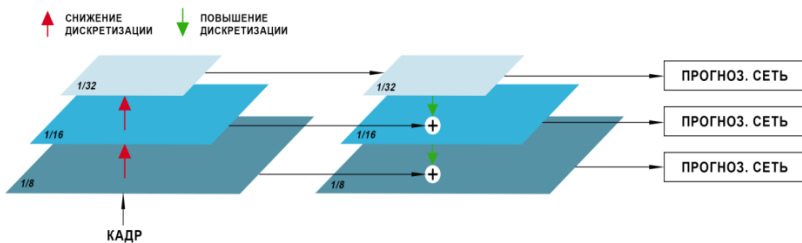


Рис. 1. Архитектура модели JDE

Тестирование алгоритма JDE осуществлялось на наборе валидационных видеофайлов с конкурса MOT-16 Challenge. Помимо этого, для проверки качества работы модели, тестирование алгоритма осуществлялось также на видеоданных со спортивных мероприятий. Для этого с помощью купольной камеры GV-EVD3100 была проведена съемка 12 отрывков с реальных игр баскетбольных команд в спортивном зале ЯрГУ им. П.Г. Демидова.

В качестве алгоритма численной оптимизации разработанной модели искусственного интеллекта использовался стохастический градиентный спуск с начальной скоростью обучения 0.01. Для более быстрой сходимости модели, на 15-ой и 23-ей эпохе обучения, скорость оптимизатора уменьшалась в 10 раз. Алгоритм JDE заканчивал свое обучение после совершения 30 эпох обучения. Критерием оценки качества алгоритма JDE служил коэффициент MOTA (Multiple Object Tracking Accuracy) [9]. Ключевой особенностью этого показателя является то, что он напрямую соотносится с тем, как человеческий глаз отслеживает целевые объекты на видеоданных.

В результате обучения модели и ее последующего тестирования алгоритм JDE показал результаты, сопоставимые с качеством работы современных моделей обнаружения и отслеживания объектов. Так, согласно результатам, приведенным в статьях [10, 11], алгоритм DeepSort_2 с прогнозирующей сетью WRN (Wide Residual Connection) (DeepSort_2 + WRN), а также модель CNNMTT с прогнозирующей пятислойной сверточной нейронной сетью (CNNMTT + 5 layer CNN), показали значения метрики MOTA, равные 61.4% и 65.2% соответственно. Для тестовой выборки базы данных MOT-16 Challenge, качество обнаружения и отслеживания объектов с помощью алгоритма JDE составило 62.6%. В то же самое время, алгоритмы DeepSort_2 + WRN и CNNMTT + 5 layer CNN, продемонстрировали существенно более медленную скорость работы: 8 и 6 FPS против 18 FPS – скорости, полученной с помощью модели JDE.

Результаты работы алгоритма JDE для видеозаписей спортивных мероприятий, снятых с диагонального ракурса купольной камеры GV-EVD3100 показаны на рис. 2.



Рис. 2. Результат работы алгоритма JDE для кадров видеозаписей матчей, снятых на купольную камеру

Таким образом, в работе представлены результаты исследования алгоритма JDE, предназначенного для решения задачи обнаружения и отслеживания людей на видеоданных в режиме реального времени. Обучение и тестирование разработанной модели осуществлялось на графическом процессоре суперкомпьютера NVIDIA DGX-1. Численные эксперименты по оценке качества работы разработанной модели были проведены на изображениях, извлеченных из видеофайлов базы данных MOT-16 Challenge. Критерием оценки качества работы алгоритма служила метрика MOTA. Качество работы модели JDE проверялось также и на собранной базе видеозаписей баскетбольных матчей, снятых на купольную камеру GV-EVD3100, расположенную в спортивном зале ЯрГУ им. П.Г. Демидова. В результате обучения и тестирования модели JDE для набора видеофайлов с конкурса MOT-16 Challenge значение метрики MOTA составило 62,6%.

В будущем планируется улучшить качество работы алгоритма JDE путем его обучения на видеокдрах, полученных при съемке состязаний, а также с помощью более точной настройки гиперпараметров модели.

Работа выполнена в рамках НИР СП-02/2021 в рамках реализации Программы развития ЯрГУ на период 2021-2030 годов.

Библиографический список

1. Минаев Е.Ю., Кутикова В.В., Никоноров А.В. Трекинг объектов в видеопотоке на основе сверточных нейронных сетей и фрактального анализа // Сборник трудов IV международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ-2018). - Самара: Новая техника, 2018. - С. 2792-2798.

2. Wang Z., Zheng L., Liu Y., Wang S. Towards Real-Time Multi-Object Tracking. Веб: <https://arxiv.org/pdf/1909.12605.pdf>.
3. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. Веб: <https://arxiv.org/pdf/1804.02767v1.pdf>.
4. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature Pyramid Networks for Object Detection. URL: <https://arxiv.org/pdf/1612.03144.pdf>.
5. Dollar P., Wojek C., Schiele B., Perona P. Pedestrian detection: A benchmark // 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 304-311.
6. Xiao T., Li S., Wang B., Lin L., Wang X. Joint Detection and Identification Feature Learning for PersonSearch // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 2017, pp. 3376–3385.
7. Zheng L., Zhang H., Sun S., Chandraker M., Yang Y., Tian Q. Person Re-identification in the Wild. Веб: <https://arxiv.org/pdf/1604.02531.pdf>.
8. Milan A., Leal-Taixe L., Reid I., Roth S., Schindler K. MOT16: A Benchmark for Multi-Object Tracking. URL: <https://arxiv.org/pdf/1603.00831.pdf>.
9. Щелкунов А. Е., Ковалев В. В., Морев К. И., Сидько И. В. Метрики оценки алгоритмов автоматического сопровождения // Таганрог. Известия ЮФУ. Технические науки. 2020. №1. С. 233-245.
10. Zagoruyko S., Komodakis N. Wide Residual Networks. URL: <https://arxiv.org/pdf/1605.07146.pdf>.
11. Mahmoudi N., Ahadi S.M., Rahmati M. Multi-target tracking using CNN-based features: CNNMTT // Multimedia Tools and Applications, 2019, v. 78, No. 6, pp. 7077-7096.

Хряшев В.В.

Матвеев Д.В.

Ивановский Л.И.

Голубев М.Н.

Ярославский государственный
университет им. П.Г. Демидова,
г. Ярославль, Россия

Khryashchev V.V.

Matveev D.V.

Ivanovsky L.I.

Golubev M.N.

Yaroslavl State University
named after P.G. Demidov,
Yaroslavl, Russia